



UNIVERSIDADE FEDERAL DE SANTA CATARINA
CAMPUS REITOR JOÃO DAVID FERREIRA LIMA
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA COMPUTAÇÃO

Luís Eduardo Ramos de Carvalho

Desenvolvimento de métodos computacionais em volumes: segmentação de estruturas orgânicas em rochas

Florianópolis
2020

Luís Eduardo Ramos de Carvalho

Desenvolvimento de métodos computacionais em volumes: segmentação de estruturas orgânicas em rochas

Tese submetida ao Programa de Pós-Graduação em Ciência da Computação da Universidade Federal de Santa Catarina para a obtenção do título de doutor em Ciência da Computação.
Orientador: Prof. Aldo von Wangenheim, Dr. rer.nat.
Coorientador: Prof. Celso Peres Fernandes, Dr.

Florianópolis
2020

Ficha de identificação da obra elaborada pelo autor,
através do Programa de Geração Automática da Biblioteca Universitária da UFSC.

Ramos de Carvalho, Luís Eduardo

Desenvolvimento de métodos computacionais em volumes:
segmentação de estruturas orgânicas em rochas / Luís Eduardo
Ramos de Carvalho ; orientador, Aldo von Wangenheim,
coorientador, Celso Peres Fernandes, 2020.

274 p.

Tese (doutorado) - Universidade Federal de Santa
Catarina, Centro Tecnológico, Programa de Pós-Graduação em
Ciência da Computação, Florianópolis, 2020.

Inclui referências.

1. Ciência da Computação. 2. Identificação e Extração de
Microfósseis. 3. Visão computacional. 4. Redes Neurais
Convolucionais. I. von Wangenheim, Aldo. II. Peres
Fernandes, Celso. III. Universidade Federal de Santa
Catarina. Programa de Pós-Graduação em Ciência da Computação.
IV. Título.

Luís Eduardo Ramos de Carvalho

Desenvolvimento de métodos computacionais em volumes: segmentação de estruturas orgânicas em rochas

O presente trabalho em nível de doutorado foi avaliado e aprovado por banca examinadora composta pelos seguintes membros:

Prof. Gerson Fauth, Dr.
Universidade do Vale do Rio dos Sinos

Prof. Rafael de Santiago, Dr.
Universidade Federal de Santa Catarina

Prof. Eros Comunello, Dr.
Universidade do Vale do Itajaí

Certificamos que esta é a **versão original e final** do trabalho de conclusão que foi julgado adequado para obtenção do título de doutor em Ciência da Computação.

Prof.^a Vania Bogorny, Dra.
Coordenadora do Programa

Prof. Aldo von Wangenheim, Dr. rer.nat.
Orientador

Prof. Celso Peres Fernandes, Dr.
Coorientador

Florianópolis, 19 de Fevereiro de 2020.

Este trabalho é dedicado a minha família e amigos em especial aos meus pais, que tanto fizeram por mim, a meu irmão que sempre me apoiou e me auxiliou e a minha noiva que me deu forças para continuar e nunca desistir.

AGRADECIMENTOS

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior- Brasil (CAPES) - Código de Financiamento 001. Destacam-se aqui outros apoiadores os quais auxiliaram no desenvolvimento deste trabalho: Laboratório de Meios Porosos e Propriedades Termofísicas (LMPT) - UFSC; Instituto Tecnológico de Micropaleontologia - itt Fossil- em especial aos paleontologistas Lucas Oliveira, Gabriel Klaser Hennemann e Andressa Nauter por todo trabalho que tiveram na identificação dos microfósseis nas amostras;

RESUMO

A aplicabilidade da análise computacional de imagens paleontológicas vai desde o estudo da evolução de animais, plantas e microrganismos até a criação de simulações sobre o habitat dos seres de determinada época, podendo ser aplicada nos mais diversos nichos, como por exemplo, na parte de exploração de petróleo. Nesta área existem diversos fatores a serem analisados a fim de minimizar os gastos atrelados ao processo de extração de petróleo. Uma destas análises é a do meio a ser explorado. Esta análise é realizada de diversas formas: utilização de sondas, extração de amostras para avaliação de componentes petrofísicos, correlacionamento com dados de outros poços de perfuração, entre outras. Na parte de extração de amostras existem algumas áreas que fazem diferentes análises sobre as mesmas amostras de rocha e para tanto é utilizada a Tomografia Computadorizada (TC), a qual preserva a amostra e a disponibiliza para diversas análises. Com base nas imagens geradas pela TC várias análises e simulações podem ser realizadas e processos, atualmente realizados de forma manual e exaustiva, podem ser automatizados. Um destes processos é o de identificação e extração de microfósseis, e a automatização do mesmo é o objetivo geral deste trabalho. Para tanto é proposto um *pipeline* que inicia no processo de digitalização e finaliza no processo de extração virtual do microfóssil. Na parte de identificação e extração dos microfósseis, uma abordagem com Redes Neurais Convolucionais (RNC) foi empregada, a qual obteve um valor de 98% de *Intersection Over Union* (IOU), métrica atualmente sendo utilizada para validação de resultados de segmentação semântica. A estratégia de validação do *pipeline* proposto se deu através de análise realizada por um grupo de especialistas da área de micropaleontologia.

Palavras-chave: Identificação e extração de Microfósseis. Automatização. Redes Neurais Convolucionais.

ABSTRACT

The applicability of the computational analysis of paleontological images ranges from the study of the animals, plants and microorganisms evolution to the simulations on the habitat of the beings of a given epoch, which can be applied in several niches, e.g. oil exploration. In this area, there are several factors to be analyzed in order to minimize the related expenses to the oil extraction process. One of these analyzes is the environment to be explored. This analysis occurs in several ways: use of probes, extraction of samples for petrophysical components evaluation, the correlation with logs of other drilling wells and so on. During the samples extraction phase, some areas make different analyses on the same sample and for that the Computed Tomography (CT) is used, which preserves the sample and makes it available for several analyzes. Based on the images generated by the CT, several analyzes and simulations can be performed and processes that are currently performed manually in a exhaustive manner can be automated. One of these processes is the microfossils identification and extraction, which, its automation is the general objective of this work. To do so, a pipeline is proposed that starts in the scanning process and ends with the identification and extraction processes. In the identification and extraction step, an approach with Convolutional Neural Networks (CNN) was used, obtaining a 98% Intersection Over Union (IOU) value. The proposed pipeline validation strategy was carried out by specialists in the micropaleontology field.

Keywords: Microfossils Identification and extraction. Automation. Convolutional Neural Networks.

LISTA DE FIGURAS

Figura 1 – Exemplo da substituição do processo manual por um processo digital através de um programa específico de análise de microfósseis. . . .	18
Figura 2 – Esquemática, de forma completa, da proposição de Popper. . . .	21
Figura 3 – Gráfico com o número de trabalhos avaliados em cada grupo. . . .	25
Figura 4 – <i>Pipeline</i> do primeiro formato.	26
Figura 5 – <i>Pipeline</i> do segundo formato.	27
Figura 6 – Volume renderizado para visualização	32
Figura 7 – <i>Pipeline</i> inicialmente proposto.	34
Figura 8 – Demonstração de parte do procedimento para digitalização. (A) Mostra evidenciando a parte utilizada para a digitalização. (B) Parte extraída da amostra apresentada em escala. (C) Exemplo de uma fatia resultante da digitalização da parte extraída da amostra.	35
Figura 9 – Exemplo de aplicação de filtro mediana.	36
Figura 10 – Exemplo de aplicação de filtro gaussiano.	36
Figura 11 – Exemplo de uma operação de convolução.	38
Figura 12 – Exemplo de uma rede convolucional extraindo características de uma imagem para combinar em conceitos mais elaborados.	38
Figura 13 – <i>Workflow</i> dos processos executados nesse trabalho.	39
Figura 14 – Mapa da Bacia de Sergipe com o local exato de extração de amostras de rocha marcado com uma cruz vermelha. A amostra foi coletada a uma profundidade de aproximadamente 2.500 metros.	40
Figura 15 – Amostra de rocha analisada (A) e uma de suas seções 2D de microtomografia (B).	41
Figura 16 – Microfósseis sendo analisados nos estudos das amostras em questão. Foraminíferos planctônicos: 1) <i>Globogerinoides ruber</i> ; 2a-b) <i>Candeina nitida</i> ; 3) <i>Orbulina universa</i> ; 4) <i>Globigerinoides trilobus saculifera</i> ; 5) <i>Globigerinoides trilobus</i> ; 6a-b) <i>Globorotalia truncatulinoides</i> . Foraminíferos bentônicos: 1) <i>Bulinina</i> ; 2) <i>Bolivinita</i> ; 3a-c) <i>Cibicides</i> ; 4) <i>Laticarinina</i> ; 5) <i>Uvigerina</i> ; 6) <i>Sphaeroidina</i> ; 7) <i>Siphonaperta</i> ; 8) <i>Quinqueloculina</i>	43
Figura 17 – Modelo digital de alguns dos microfósseis isolados fisicamente e que estão presentes na amostra de rocha sendo analisada.	44
Figura 18 – Fatia de uma amostra e seu respectivo GT, utilizado como padrão ouro na identificação da localização dos microfósseis.	45
Figura 19 – Aplicação de uma técnica de segmentação por contornos ativos ao volume sendo analisado.	47

Figura 20 – Aplicação de uma sequência de métodos ao volume sendo analisado: classificação e binarização.	48
Figura 21 – Aplicação de uma técnica de detecção de borda ao volume sendo analisado.	48
Figura 22 – Aplicação de uma técnica de filtro de tamanho das componentes (> 50000) sobre rótulos realizado sobre o resultado encontrado no sequenciamento representado pela figura 25.	49
Figura 23 – Aplicação de uma técnica de filtro de tamanho das componentes (entre 200 e 9000) sobre rótulos realizado sobre o resultado encontrado no sequenciamento representado pela figura 25.	49
Figura 24 – Aplicação de um conjuntos de técnicas: Classificação, detector de borda, binarização e o operador lógico <i>And</i> entre os resultados das imagens de borda e do classificador.	50
Figura 25 – Aplicação de um conjuntos de técnicas: Ajuste de brilho e contraste, binarização, detector de borda e rotulamento por componentes conexas.	51
Figura 26 – Exemplo da arquitetura UNET.	52
Figura 27 – Resultado da segmentação de microfósseis obtido com a abordagem de 4 classes. (A) Imagem digitalizada original. (B) GT gerado manualmente pelos paleontólogos. (C) UNET + ResNet 34.	53
Figura 28 – Abordagem de treinamento de resolução de imagem progressiva.	55
Figura 29 – O GT (A) e o resultado obtido da segmentação de microfósseis (B) com 4 classes, ajuste dos hiper-parâmetros e melhor exploração de <i>data augmentation</i> e <i>batch size</i>	56
Figura 30 – (A) Fatia 2D da amostra original digitalizada. (B) GT referente a fatia 2D. (C) Resultado da predição UNET + ResNet18 + otimização dos hiper-parâmetros. (D) Resultado da predição UNET + ResNet101 + otimização dos hiper-parâmetros. (E) Resultado da predição UNET + ResNet34 + otimização dos hiper-parâmetros. (F) UNET + ResNet50 + otimização dos hiper-parâmetros.	57
Figura 31 – Resultado da aplicação da máscara de segmentação obtida sobre a imagem digitalizada. (A) imagem da fatia 2D, com contraste ajustado, após a aplicação da máscara resultante. Destacado por um retângulo vermelho encontra-se um microfóssil específico. (B) Microfóssil destacado extraído e ampliado para visualização. (C) Microfóssil isolado fisicamente, digitalizado com o Versa XRM-500 MicroCT. (D) Aquisição multidimensional do microfóssil cibicidoides com o estereoscópio Zeiss Discovery V20.	57

- Figura 32 – Resultado da aplicação da máscara de segmentação obtida sobre a imagem digitalizada. (A) imagem da fatia 2D, com contraste ajustado, após a aplicação da máscara resultante. Em vermelho encontra-se um microfóssil específico destacado. (B) Microfóssil destacado extraído e ampliado para visualização. (C) Microfóssil isolado fisicamente, digitalizado com o Versa XRM-500 MicroCT. (D) Aquisição multidimensional do microfóssil cibicoides com o estereoscópio Zeiss Discovery V20. 58
- Figura 33 – Resultado da aplicação da máscara de segmentação obtida sobre a imagem digitalizada. (A) imagem da fatia 2D, com contraste ajustado, após a aplicação da máscara resultante. Em vermelho encontra-se um microfóssil específico destacado. (B) Microfóssil destacado extraído e ampliado para visualização. (C) Microfóssil isolado fisicamente, digitalizado com o Versa XRM-500 MicroCT. (D) Aquisição multidimensional do microfóssil cibicoides com o estereoscópio Zeiss Discovery V20. 58

LISTA DE TABELAS

Tabela 0 – Tabela resumando o resultado da bateria de testes para análise dos parâmetros do método de contornos ativos.	50
Tabela 1 – Desempenho da segmentação em termos de valor de IOU. Cada método foi avaliado em um conjunto de 1000 imagens contendo marcações dos dados de microfósseis.	56

LISTA DE ABREVIATURAS E SIGLAS

CC	Camada Convolucional
DL	<i>Deep learning</i>
GT	Ground Truth
IOU	Intersect Over Union
IRM	Imagem por Ressonância Magnética
Micro-CT	Microtomografia Computadorizada
RNC	Redes Neurais Convolucionais
SS	Segmentação Semântica
TC	Tomografia Computadorizada
VC	Visão Computacional

SUMÁRIO

1	INTRODUÇÃO	15
1.1	APRESENTAÇÃO DO PROBLEMA DE PESQUISA	16
1.2	OBJETIVOS	17
1.2.1	Objetivos Específicos	17
1.3	CENÁRIO DE APLICAÇÃO	17
1.4	JUSTIFICATIVA	18
1.5	ESTRUTURA DA TESE	19
2	METODOLOGIA	20
2.1	ESTUDO DE VALIDAÇÃO	20
3	REVISÃO DA LITERATURA	22
3.1	1ª ETAPA: TRABALHOS COM MÉTODOS COMPUTACIONAIS DE ANÁLISE DE IMAGEM TOMOGRÁFICAS DE FÓSSEIS	22
3.1.1	Discussão sobre os trabalhos analisados	23
3.2	2ª ETAPA: ANÁLISE DE ALGORITMOS DE SEGMENTAÇÃO 3D EM GERAL	23
3.2.1	Discussão sobre os trabalhos analisados	24
3.3	3ª ETAPA: ANÁLISE DE MÉTODOS DE RECONHECIMENTO/CLASSIFICAÇÃO DE OBJETOS 3D	24
3.3.1	Discussão sobre os trabalhos analisados	26
3.4	4ª ETAPA: ANÁLISE DE MÉTODOS AUTOMÁTICOS APLICADOS A IDENTIFICAÇÃO DE MICRO-ORGANISMOS	27
3.4.1	Discussão sobre os trabalhos analisados	28
4	FUNDAMENTAÇÃO TEÓRICA	29
4.1	PROCESSOS DE DIGITALIZAÇÃO	29
4.1.1	Tomografia destrutiva	29
4.1.2	Tomografia não-destrutiva	30
4.2	VISUALIZAÇÃO DIGITAL	31
4.3	PROCESSAMENTO DIGITAL 3D	33
4.3.1	Digitalização	34
4.3.2	Filtragem	36
4.3.3	Identificação e extração de possíveis candidatos	37
5	DESENVOLVIMENTO	39
5.1	PRÉ-PROCESSAMENTO	41
5.2	CONJUNTOS DE DADOS/BASES DE DADOS DISPONÍVEIS	42
5.3	FORMA DE VALIDAÇÃO E DEMONSTRAÇÃO DOS RESULTADOS	45
5.4	IDENTIFICAÇÃO E EXTRAÇÃO DE POSSÍVEIS CANDIDATOS	46
5.4.1	Visão Computacional Clássica	46

5.4.2	Visão computacional moderna	51
6	CONCLUSÃO	59
6.1	CONTRIBUIÇÕES	60
6.1.1	Publicações	61
6.2	TRABALHOS FUTUROS	61
	REFERÊNCIAS	62
	ANEXO A – REVISÃO DA LITERATURA SOBRE MÉTODOS DE VISÃO COMPUTACIONAL PARA ANÁLISE DE IMAGEM DE FÓSSIS	67
	ANEXO B – RELATÓRIO TÉCNICO - REVISÃO DA LITERATURA PARA ALGORITMOS DE SEGMENTAÇÃO 3D APLICADOS A IMAGENS TOMOGRÁFICAS	76
	ANEXO C – RELATÓRIO TÉCNICO - REVISÃO DA LITERATURA PARA RECONHECIMENTO/CLASSIFICAÇÃO DE OBJETOS 3D	108
	ANEXO D – BIOARXIV ARTICLE- AUTOMATED MICROFOSSIL IDENTIFICATION AND SEGMENTATION USING A DEEP LEARNING APPROACH	257

1 INTRODUÇÃO

A área da paleontologia é um campo de muito estudo que tem como base um passado cheio de informações e um futuro promissor. Com base na análise do passado foram feitas as primeiras hipóteses de como surgiu a vida, como se deu a evolução humana e as principais transformações do mundo. A palavra paleontologia vem do grego *palaiós* que quer dizer antigo associado a *óntos* - ser e *lógos* - estudo, que por definição é uma ciência que estuda a vida do passado e seu desenvolvimento ao longo dos anos (SOUZA CARVALHO, 2004). A área da paleontologia pode ser dividida, tradicionalmente, em algumas subdisciplinas:

- Micropaleontologia: Estudo dos fósseis microscópicos, indiferente do grupo ao qual pertencem;
- Paleobotânica: Estudo dos fósseis de plantas, tradicionalmente inclui o estudo de fósseis de algas e fungos em adição plantas terrestres;
- Palinologia: Estudo de polens e esporos, vivos e fósseis, produzidos por plantas terrestres e protistas;
- Paleontologia Invertebrada: Estudo de fósseis de animais invertebrados, tais como moluscos, equinodermas e entre outros;
- Paleontologia Vertebrada: Estudo de fósseis de vertebrados, partindo de peixes primitivos até mamíferos;
- Paleoantropologia: Estudo de humanos pré-históricos e proto-humano fósseis;
- Tafonomia: Estudo do processo de decaimento, preservação e a formação dos fósseis em geral;
- Icnologia: Estudo de rastros, trilhas e pegadas de fósseis;
- Paleoecologia: Estudo da ecologia e do clima do passado, revelado pela análise de fósseis e por outros métodos;

A paleontologia é uma ciência antiga, existem relatos da época anterior à cristo, e seu entrelaçamento com a área da computação se deu com maior afinco a partir da década de 1990. Em um *review* elaborado por Tipper (TIPPER, 1991), onde uma análise dos estudos paleontológicos utilizando aplicações computacionais no fim da década de 1980 é apresentada, o autor fez uma pesquisa nas principais revistas de paleontologia da época. Em sua pesquisa percebeu que em algumas destas revistas existia, de forma escassa, um artigo que descrevia alguma aplicação de métodos computacionais, em outras existiam poucos artigos e em algumas uma quantidade

elevada de artigos. Outros jornais e livros, tanto gerais como explicitamente da área da paleontologia, apresentavam o mesmo comportamento: a inserção de métodos computacionais na paleontologia parece, na melhor das hipóteses, desigual. Com a industrialização dos tomógrafos médicos houve um incentivo na pesquisa utilizando imagens tomográficas aplicadas a paleontologia, o que impulsionou o desenvolvimento de softwares voltados à análise das mesmas. Alguns exemplos são: DRISHTI, SPIERS, SEG3D, IMAGEJ, MIMICS, VGSTUDIO MAX, AVIZO, AMIRA, Geomagic, Rhinoceros, Imaris, ITK-SNAP e TurtleSEG.

Com a popularização dos tomógrafos e o desenvolvimento de ferramentas de análise, a paleontologia obteve uma ferramenta indispensável para suas análises, seja utilizando métodos de análise de imagens ou utilizando simulações de propriedades físicas, por exemplo. A aplicabilidade da então intitulada paleontologia virtual, vai desde as análises da evolução de animais, plantas e microrganismos até a criação de simulações sobre o habitat dos seres de determinada época, podendo ser aplicada nos mais diversos nichos, como por exemplo, na parte de exploração de petróleo.

A aplicação do estudo dos microfósseis na área petrolífera teve seu primeiro passo em 1890 na Polônia, apresentado por Josef Gryzbowski (SINGH, 2008), mas foi na década de 1920 nos Estados Unidos, com a generalização do uso dos microfósseis para determinar a idade dos testemunhos retirados das sondas de perfuração da indústria petrolífera, que se obteve um passo maior no desenvolvimento da micropaleontologia aplicada a esta área (MOLINA; ED., 2004).

Na área petrolífera, existem diversos fatores a serem analisados a fim de minimizar os gastos atrelados ao processo de extração de petróleo. Uma destas análises é a do meio a ser explorado. Esta análise ocorre de diversas formas: utilização de sondas, extração de amostras para avaliação de componentes petrofísicos, correlacionamento com dados de outros poços de perfuração, entre outras. Na parte de extração de amostras, existem algumas áreas que fazem diferentes análises sobre a mesma amostra. Por exemplo, na área da geologia o interesse está voltado para o estudo da rocha (formação, estrutura e porosidade, por exemplo), na parte da paleontologia o foco está na identificação e classificação dos microfósseis presentes na amostra.

1.1 APRESENTAÇÃO DO PROBLEMA DE PESQUISA

O problema de focos distintos das áreas, além da disponibilidade da amostra, está no fato que algumas destas análises acabam destruindo as amostras. Por exemplo, há um processo na área da paleontologia que, com a finalidade de isolar o fóssil, utiliza um material químico que preserva o fóssil mas destrói a rocha, tornando a amostra inutilizável para uma avaliação da área da geologia. Para contornar estes problemas utiliza-se a digitalização das amostras entretanto até mesmo na área de digitalização existem métodos destrutivos (como por exemplo, cortar a amostra em vá-

rias fatias para fazer o processo de digitalização) bem como métodos não-destrutivos (Tomografia Computadorizada (TC)). Com o foco na preservação da amostra e na disponibilização das mesmas utiliza-se TC como forma de digitalização. Com base nas imagens geradas pelo TC, várias análises e simulações podem ser realizadas e processos atualmente conduzidos de forma manual e exaustiva podem ser automatizados. Um destes processos o qual é atualmente realizado de forma manual e exaustiva é o de **identificação, extração, caracterização e classificação de microfósseis**.

1.2 OBJETIVOS

Tendo como base o problema de pesquisa apresentado, o objetivo geral deste trabalho é a elaboração de métodos computacionais em volumes tomográficos visando a identificação e extração de microfósseis em rochas carbonáticas de forma automática ou semiautomática.

1.2.1 Objetivos Específicos

- Desenvolvimento de métodos computacionais para a identificação e extração de microfósseis de forma automática ou semiautomática, visando o isolamento dos microfósseis presentes na rocha;
- Estudo de validação do encadeamento de métodos utilizados, com base em padrões ouro (*Ground Truths*) gerados manualmente por especialistas da área;
- Disponibilização de todos os métodos, dados e resultados visando formar uma *baseline* para o problema de extração ao mesmo tempo que serve de alicerce para problemas de cunho tafonômico e taxonômicos.

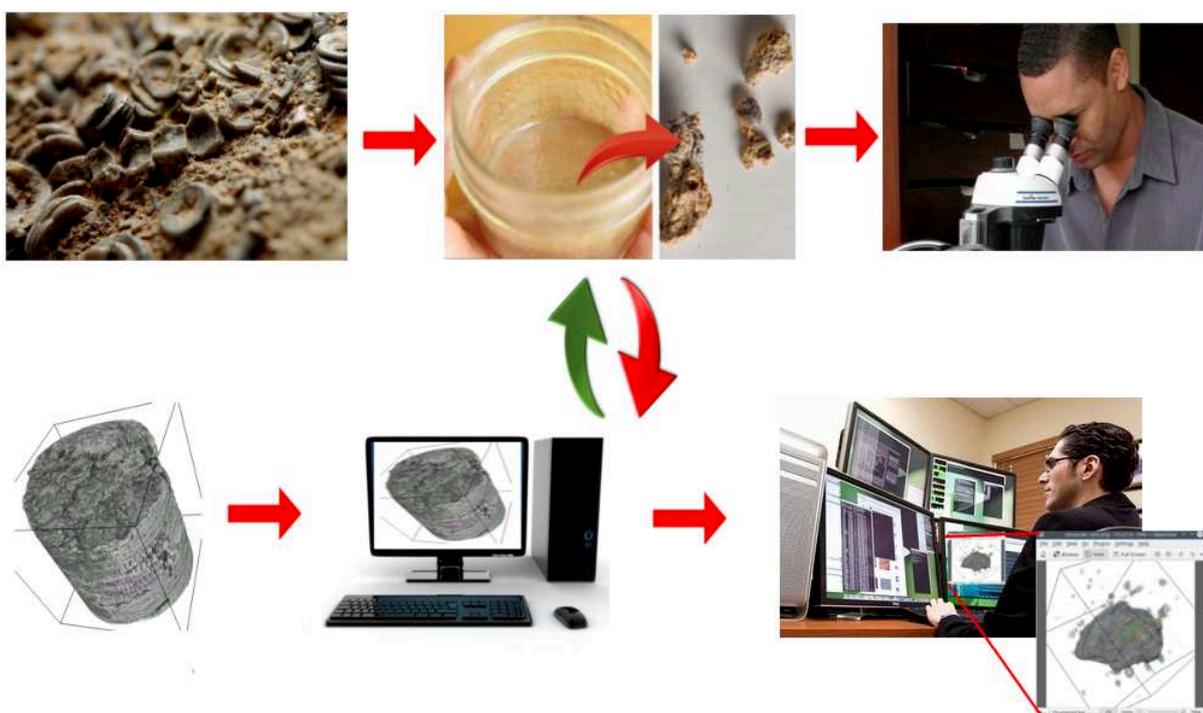
1.3 CENÁRIO DE APLICAÇÃO

Nesta seção apresenta-se um cenário de aplicação o qual relata um dos possíveis cenários de aplicação ao se resolver o problema de pesquisa previamente apresentado. **Dia a dia do paleontólogo**

Um paleontólogo recebe uma amostra de rocha contendo microfósseis para análise. O tempo necessário para todo o processo de isolamento dos microfósseis (realizado manualmente utilizando produtos químicos que diluem a rocha preservando os microfósseis) é alto se comparado com o processo de digitalização - somado ao fato de que a rocha acaba sendo destruída. Após esse processo de isolamento, o paleontólogo analisa o resultado do isolamento físico e classifica os microfósseis com base em uma caracterização específica de forma manual para cada espécime presente na rocha. Por fim, ele gerará um relatório com os resultados de análise dos

microfósseis. Agora, hipoteticamente, a empresa para qual este paleontólogo trabalha adquiriu um software de análise de imagens tomográficas, específico para a parte de microfósseis em rochas. Esse software realiza a identificação, extração e caracterização dos microfósseis de forma automática e com mínima ou nenhuma supervisão. Desta forma, o paleontólogo pode carregar sua amostra tomográfica, selecionar um sequenciamento de métodos específico e deixar o procedimento executando durante a noite, cabendo ao mesmo somente a avaliação dos resultados desta análise que antes exigia um esforço predominantemente manual. Na Figura 1, mostra-se a substituição do processo manual por um processo totalmente digital realizado no programa planejado para análise de rochas.

Figura 1 – Exemplo da substituição do processo manual por um processo digital através de um programa específico de análise de microfósseis.



1.4 JUSTIFICATIVA

Este estudo surgiu da necessidade de um conjunto de métodos para extração, identificação, caracterização e classificação de microfósseis de forma automática ou semiautomática a fim de facilitar o processo que atualmente é realizado de forma custosa e manual por uma paleontólogo; e da carência de métodos para análise de microfósseis, dado que antes as análises eram realizadas por fatias, ou seja, de forma bidimensional. A aplicação prática que iniciou este estudo é a utilização deste *pipeline* na classificação de microfósseis provenientes de amostras de rochas carbonáticas

retiradas de lugares de exploração de petróleo. A identificação de microfósseis é estratégica para a prospecção de petróleo em função da utilização em bioestratigrafia. A bioestratigrafia utiliza microfósseis de diferentes grupos para a realização da caracterização temporal de estratos rochosos sedimentares, fundamental para a indústria do petrolífera.

1.5 ESTRUTURA DA TESE

Essa tese está dividida em 6 capítulos. Inicialmente há o capítulo de introdução onde é feita a apresentação do tema sendo trabalhado, os objetivos do estudo, a justificativa do estudo, a apresentação dos riscos vinculados ao estudo e a estrutura da tese. O segundo capítulo apresenta a metodologia aplicada neste trabalho, descrevendo uma visão geral da metodologia e especificando as etapas realizadas. O terceiro capítulo apresenta todas as revisões sistemáticas realizadas no decorrer do projeto, fornecendo todos os passos do funcionamento da parte sistemática da revisão e uma análise geral dos trabalhos encontrados que estão em acordo com os critérios estabelecidos. O capítulo quatro detalha toda a parte de desenvolvimento e aprimoramento dos métodos de segmentação sendo utilizados. O capítulo cinco faz uma análise dos resultados obtidos por meio de experimentos realizados. O último capítulo apresenta as conclusões do estudo realizado nessa tese e as expectativas de trabalhos futuros.

2 METODOLOGIA

Este trabalho segue o método científico hipotético-dedutivo proposto pelo filósofo Karl Popper (POPPER, 2002), no qual se busca a eliminação dos erros de uma hipótese. Realiza-se isto a partir da ideia de testar a falsidade de uma proposição, ou seja, a partir de uma hipótese, estabelece-se que situação ou resultado experimental nega essa hipótese e tenta-se realizar experimentos para negá-la. Na Figura 2, mostra-se uma esquematização de forma completa para a proposição de Popper.

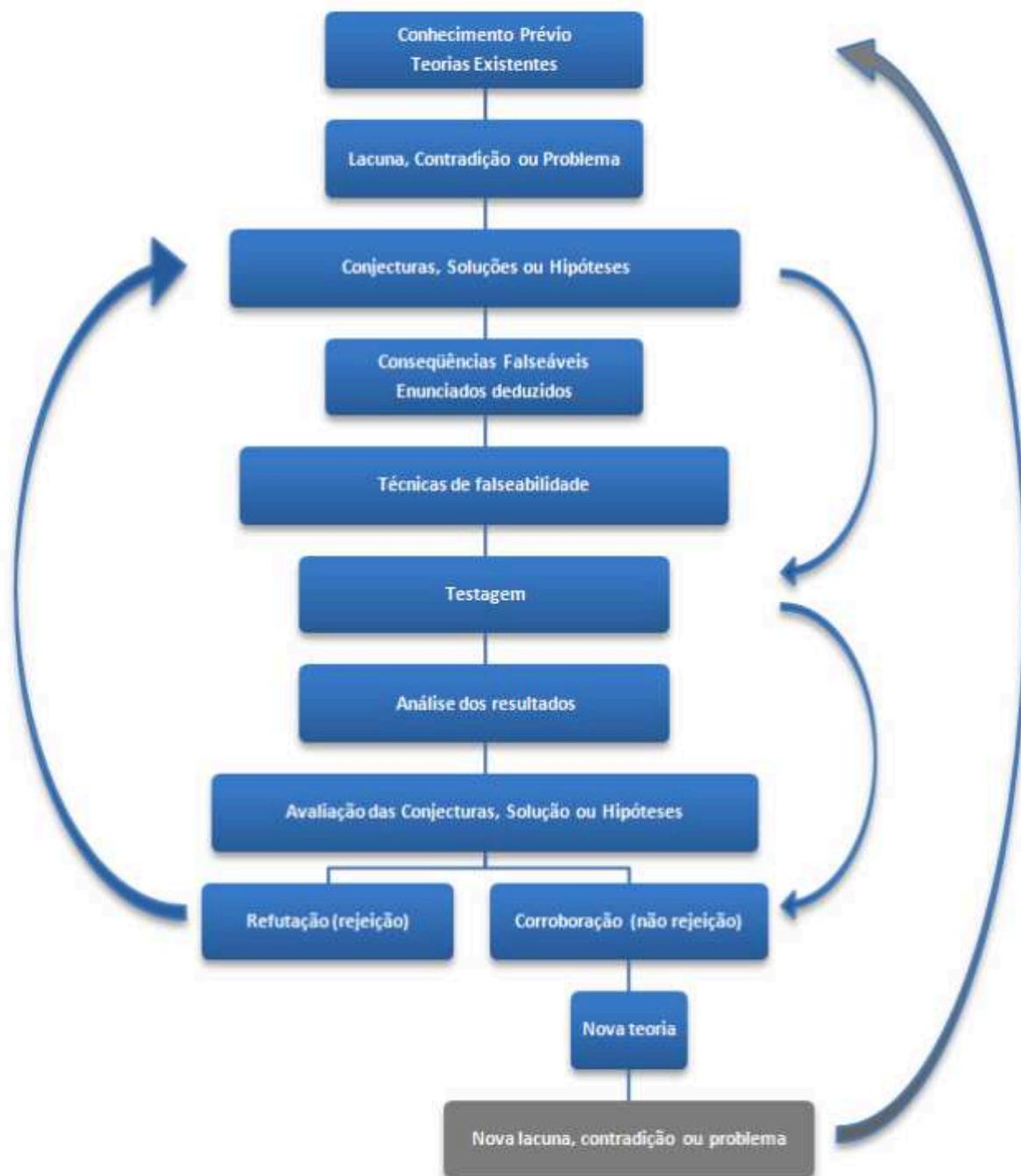
Com base nesse método, elaborou-se uma pergunta de pesquisa e realizou-se uma revisão sistemática da literatura, como descrito por Kitchenham (KITCHENHAM, 2004). Depois, com base neste levantamento, elaborou-se uma hipótese de solução, seguindo as diretrizes dos trabalhos analisados voltados para a área da paleontologia. Por fim, com base no método científico, realizou-se a consolidação da solução e a avaliação da mesma com base em experimentos e em um estudo de validação.

2.1 ESTUDO DE VALIDAÇÃO

Para o estudo de validação do *pipeline* proposto será utilizado um grupo de especialistas da área de micropaleontologia. Esse grupo de especialistas tem como propósito fornecer confiabilidade aos resultados das técnicas utilizadas no processo proposto e, conseqüentemente, à solução proposta. O procedimento de validação seguirá as seguintes etapas:

- Digitalização das amostras e dos exemplares de espécimes isolados;
- Execução de experimentos utilizando configurações do *pipeline* proposto;
- Análise dos resultados obtidos, quanto a localização dos espécimes, de forma automática;
- Comparação entre os resultados obtidos e a análise do grupo de especialistas sobre as amostras em questão.

Figura 2 – Esquematização, de forma completa, da proposição de Popper.



Fonte: Imagem retirada de (PENSAMENTO, Acessado em: 17/06/2016)

3 REVISÃO DA LITERATURA

Com a finalidade de revisar os estudos sobre fósseis e as técnicas sendo aplicadas para a análise dos mesmos, uma revisão sistemática da literatura foi realizada seguindo os procedimentos descritos por Kitchenham (KITCHENHAM, 2004). Com base em Kitchenham, definiu-se a pergunta de pesquisa inicial: **O processo de identificação e classificação de fósseis em rochas, que atualmente é feito de forma manual, pode ser realizado, de forma satisfatória, automaticamente ou semi-automaticamente?**. Esta pergunta de pesquisa acabou por ser desmembrada em quatro tópicos de pesquisa, dado que se procurou: 1) analisar o que existe sobre a parte de métodos computacionais aplicados a imagens tomográficas de fósseis; 2) avaliar quais os métodos de segmentação, aplicados a imagens tomográficas, poderiam possivelmente ser utilizados na parte de segmentação de microfósseis; 3) analisar e compreender os métodos utilizados para o reconhecimento/classificação de objetos 3D em um contexto geral; 4) analisar métodos aplicados à micro-organismos em contexto geral.

Estes quatro tópicos de pesquisa, por sua vez, geraram três etapas de revisões sistemáticas e uma revisão simplificada da literatura. A primeira etapa teve como foco os trabalhos com métodos computacionais de análise de imagem de fósseis. A segunda etapa teve como principal objetivo analisar algoritmos de segmentação 3D em geral, e a terceira e última etapa teve como objetivo a análise de métodos de reconhecimento/classificação de objetos 3D. Para todas as três etapas, três bases de artigos científicos foram utilizadas (ScienceDirect, ACM e IEEE) e, os artigos disponíveis para leitura na íntegra e publicados entre 2006 e junho de 2016, para as duas primeiras etapas, e entre 2006 e dezembro de 2016 para a terceira etapa, foram analisados. As três etapas estão descritas detalhadamente nas seções seguintes. A revisão simplificada da literatura teve como foco fazer uma análise rápida de métodos automáticos aplicados à identificação de micro-organismos de forma geral e também se encontra descrita nas seções seguintes.

3.1 1ª ETAPA: TRABALHOS COM MÉTODOS COMPUTACIONAIS DE ANÁLISE DE IMAGEM TOMOGRÁFICAS DE FÓSSEIS

Para a primeira etapa de revisão sistemática, com o tópico de pesquisa sobre o que existe sobre **métodos computacionais aplicados a imagens tomográficas de fósseis**, as seguintes chaves de busca, para cada base, foram empregadas:

ScienceDirect: *pub-date > 2005 and pub-date < 2017 and TITLE-ABSTRACT-KEY(fossil) and TITLE-ABSTRACT-KEY(tomography).*

ACM: *(+"fossil"tomography) e Published years: 2006 - 2016).*

IEEE: *((fossil) AND tomography) e Year: 2006 - 2016.*

Obteve-se um total de 250 artigos, dos quais foram analisados os *abstracts*. Na leitura dos *abstracts*, utilizaram-se os seguintes **critérios de exclusão**, a fim de selecionar os trabalhos a serem analisados na íntegra:

- Artigos escritos em outras línguas diferentes da inglesa;
- Artigos repetidos e que não fazem referência ao estudo de fósseis com base na análise de imagem.
- Artigos que não possuem uma descrição de etapas computacionais realizadas (aplicação de métodos de análise de imagem, filtros, segmentações, entre outros).
- O trabalho relatado, na parte computacional, se resume apenas ao uso de um software para visualização e medição do fóssil.

Com a aplicação do critério de exclusão, obteve-se um total de 18 artigos, os quais foram lidos na íntegra, sumarizados e analisados. Visando o espaço utilizado para tamanho detalhamento, toda a descrição e análise está anexada no Anexo A, transcrevendo aqui apenas a discussão sobre os trabalhos analisados.

3.1.1 Discussão sobre os trabalhos analisados

De forma geral todos os trabalhos selecionados utilizam técnicas de processamento de imagens para a análise das imagens de fósseis, entretanto grande parte destes depende de uma exaustiva iteração manual por parte do operador do método. Visto que o estudo proposto está direcionado à geração de um encadeamento de métodos, os quais necessitem o mínimo possível da iteração do operador, para fins de identificação e classificação de fósseis em imagens tomográficas, os trabalhos analisados servem para fornecer uma visão geral dos métodos geralmente aplicados para os estudos voltados para esta área.

3.2 2ª ETAPA: ANÁLISE DE ALGORITMOS DE SEGMENTAÇÃO 3D EM GERAL

Após uma análise inicial dos métodos geralmente empregados nos estudos de imagens de fósseis, passa-se para a segunda etapa de revisão sistemática, com o tópico de pesquisa sobre o que existe de **métodos de segmentação 3D, aplicados a imagens tomográficas**, e que eventualmente poderiam ser utilizados na parte de segmentação de microfósseis. As seguintes chaves de busca, para cada base, foram empregadas:

ScienceDirect: *pub-date > 2005 and pub-date < 2016 (july) and TITLE-ABSTRACT-KEY(3D segmentation) and TITLE-ABSTRACT-KEY(computer tomography).*

ACM: ("3D segmentation"+computer +tomography) and Published years: 2006 - 2016 (july).

IEEE: ("3D segmentation") AND computer tomography) and Year: 2006-2016(july).

Obteve-se um total de 114 artigos, dos quais foram analisados os *abstracts*. Na leitura dos *abstracts*, utilizaram-se os seguintes **critérios de exclusão**, a fim de selecionar os trabalhos a serem analisados na íntegra:

- Artigos escritos em outras línguas diferentes da inglesa;
- Artigos repetidos e que não fazem referência ao estudo de métodos de segmentação 3D;
- O artigo não possui uma descrição de etapas computacionais realizadas (pré-processamento, pré-segmentações, entre outros);

Com a aplicação do critério de exclusão, obteve-se um total de 67 artigos, os quais foram lidos na íntegra, sumarizados, analisados e categorizados de acordo com o conceito principal de construção do método de segmentação em questão. Visando o espaço utilizado para tamanho detalhamento, toda a descrição e análise está anexada no Anexo B, transcrevendo aqui apenas a discussão sobre os trabalhos analisados.

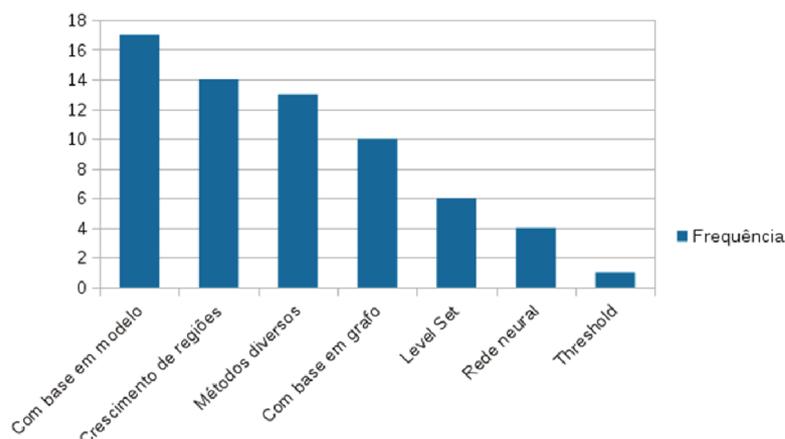
3.2.1 Discussão sobre os trabalhos analisados

Com base nos métodos analisados pode-se demonstrar, inicialmente, uma tendência no uso de métodos de segmentação 3D com base em modelos e crescimento de regiões, ilustrado pelo gráfico presente na figura 3. Esta tendência é um bom indicativo de quais os tipos de métodos aplicados para segmentação de imagens tomográficas (no período compreendido entre 2006-2016) possuem uma possível aplicabilidade nas imagens tomográficas de microfósseis. Entretanto, apesar desta tendência, métodos com pouca frequência, como por exemplo, *threshold* e baseados em rede neural, não podem ser descartados dado as características atreladas a estes dois tipos de métodos. Os métodos de *threshold*, por exemplo, possuem uma simplicidade na sua implementação e na computação dos resultados, o que facilita sua aplicação em imagens tomográficas com alta resolução.

3.3 3ª ETAPA: ANÁLISE DE MÉTODOS DE RECONHECIMENTO/CLASSIFICAÇÃO DE OBJETOS 3D

Para a terceira etapa de revisão sistemática, com o tópico de pesquisa sobre o que existe de **métodos utilizados para o reconhecimento/classificação de objetos 3D em um contexto geral**, as seguintes chaves de busca, para cada base, foram empregadas:

Figura 3 – Gráfico com o número de trabalhos avaliados em cada grupo.



ScienceDirect: $2017 < \text{pub-date and pub-date} > 2005$ and TITLE-ABSTR-KEY("3D object classification") or TITLE-ABSTR-KEY("3D object recognition")

ACM: "3D object recognition3D object classification"published after 2006 until 2016

IEEE: ("3D object recognition") OR "3D object classification") and refined by Year: 2006-2016

Obteve-se um total de 446 artigos dos quais foram analisados os *abstracts*. Nesta análise, utilizaram-se os seguintes **critérios de exclusão** a fim de selecionar os trabalhos a serem analisados:

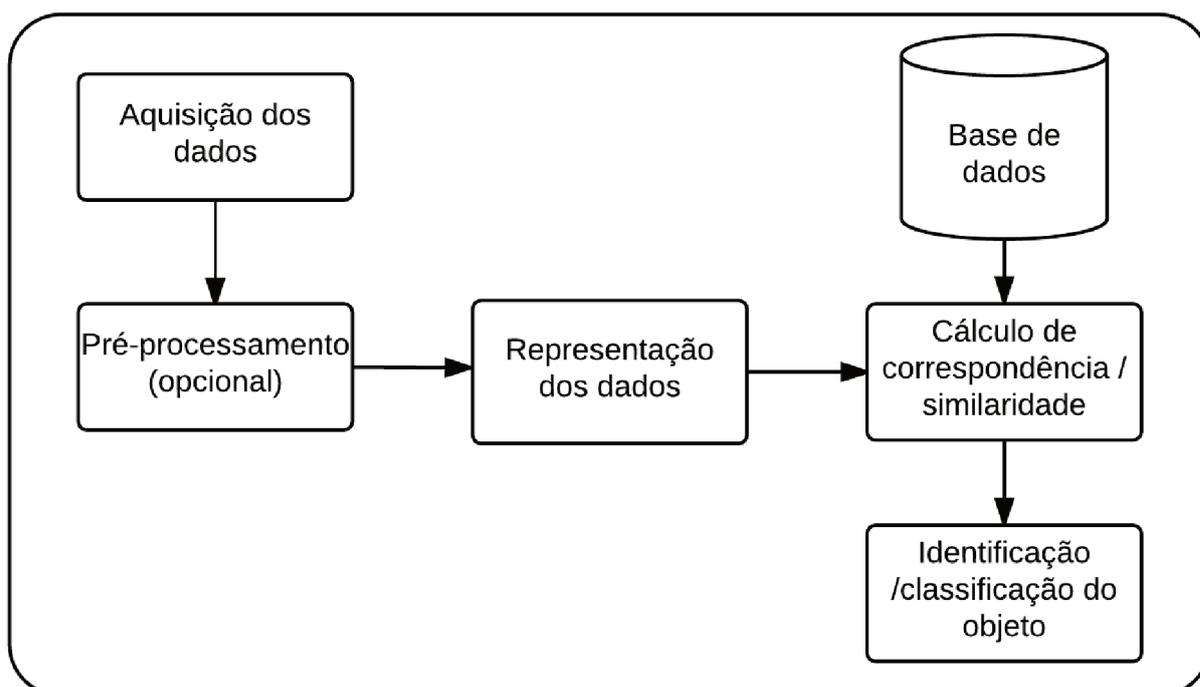
- Artigos escritos em outras línguas diferentes da inglesa;
- Artigos repetidos e que não descrevessem nos *abstracts* uma técnica para reconhecimento/classificação de objetos 3D.
- Artigos que apenas fazem uso de softwares para a parte de reconhecimento/classificação de objetos e/ou não possuem foco na técnica de reconhecimento/classificação de objetos 3D apenas citando o seu uso.

Com a aplicação do critério de exclusão, obteve-se um total de 277 artigos, os quais foram lidos na íntegra. Destes 277, 10 trabalhos são referentes a livros, apresentação de bases de dados 3D e revisões de métodos e tipos de classificação e, portanto, foram separados do processo de categorização pelo tipo de representação de objetos 3D sendo utilizado. Visando o espaço utilizado para tamanho detalhamento, toda a descrição e análise está anexada no Anexo C, transcrevendo aqui apenas a discussão sobre os trabalhos analisados.

3.3.1 Discussão sobre os trabalhos analisados

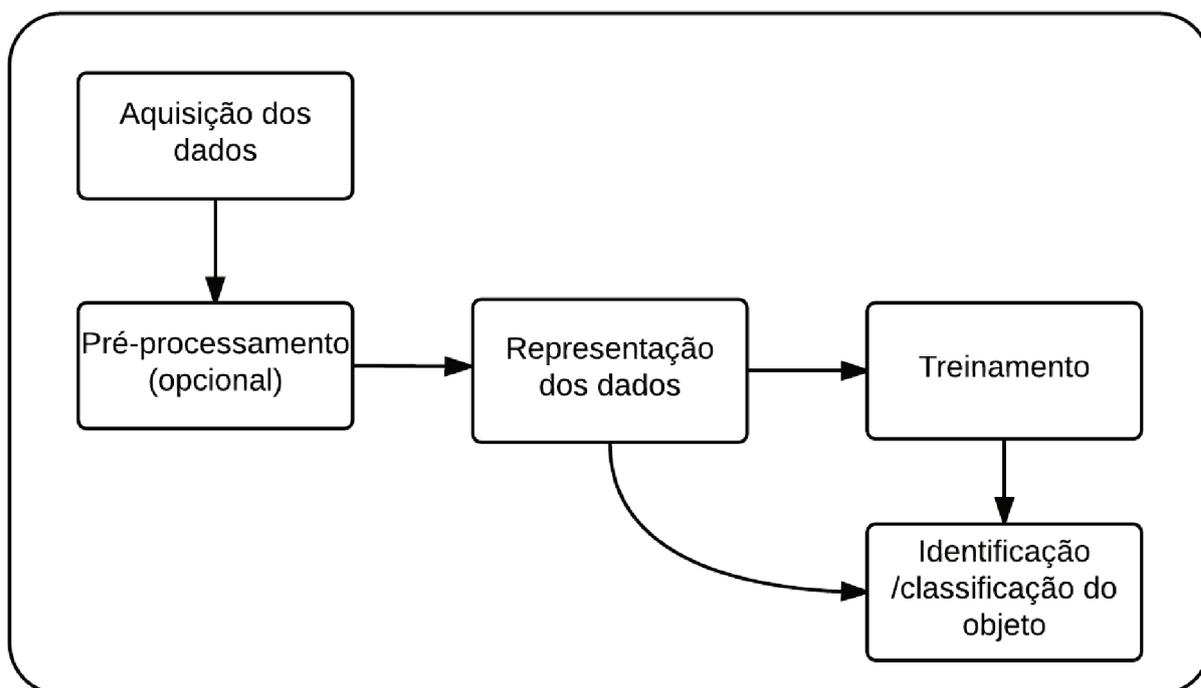
Com base em todos os trabalhos analisados pode-se sumarizar dois formatos básicos dos *pipelines* para reconhecimento/classificação de objetos 3D, presentes na grande maioria dos trabalhos. Tanto a primeira forma, ilustrada pela figura 4, quanto a segunda forma, ilustrada pela figura 5, possuem etapas similares e diferem na parte de como realizam o reconhecimento/classificação do objeto. Na primeira forma, a representação escolhida para os dados de entrada é correspondida por meio de um método de correspondência, ou de cálculo de similaridade, entre a representação do dado de entrada e as representações armazenadas em uma base de dados. As representações armazenadas na base de dados foram previamente obtidas em uma fase *offline* e o processo de cálculo de similaridade/correspondência elege, com base no maior valor de similaridade ou correspondência, a classificação da representação de entrada ou a representação correspondente a ela na base de dados.

Figura 4 – *Pipeline* do primeiro formato.



Na segunda forma, a representação escolhida serve de entrada para o treinamento de um classificador, realizado *offline* ou em alguns casos com treinamento *online*, utilizando-o posteriormente para classificação/reconhecimento de outros dados de entrada. Cada etapa dos dois *pipelines* gerais será melhor exemplificada nas seções seguintes.

Figura 5 – Pipeline do segundo formato.



3.4 4ª ETAPA: ANÁLISE DE MÉTODOS AUTOMÁTICOS APLICADOS A IDENTIFICAÇÃO DE MICRO-ORGANISMOS

Na quarta etapa da revisão da literatura, foram analisados métodos para identificação automática de micro-organismos em geral. Nesta revisão, foram encontrados alguns trabalhos de pesquisa realizados nas últimas décadas (LIU; THONNAT; BERTHOD, 1994), (CULVERHOUSE *et al.*, 1996), (BEAUFORT; DOLLFUS, 2004). O objetivo desses estudos era reduzir o tempo e o custo do processo de identificação e melhorar a reprodutibilidade da classificação. Vários métodos para classificação automática de microfósseis marinhos foram analisados: os recursos de imagem (morfologia, textura e intensidade) fornecidos pelos métodos de processamento de imagem foram combinados com a rede neural artificial (HIBBETT, 2009) (SCHULZE *et al.*, 2013), estatística (CULVERHOUSE *et al.*, 1996) ou regra baseado em classificadores (YU *et al.*, 1996). Outros trabalhos empregaram diretamente a imagem inteira (FRANCUS, 2007) ou usaram uma combinação de imagem e morfologia (BARBARIN, 2014) com Redes Neurais Convolucionais (RNC). Algumas abordagens mais recentes são apresentadas nos seguintes trabalhos: em (CHARLES, 2011), Charles propõe segmentar as partículas de primeiro plano na imagem e identificar aquelas que contêm um único palinomorfo elíptico completo. Para fazer isso, ele emprega classificadores treinados para distinguir entre regiões que contêm um único palinomorfo e um que contêm outros materiais; em (BUENO *et al.*, 2017), Bueno *et al.* lidam com a identificação automá-

tica de táxons com base em métodos de aprendizado de máquina, com o objetivo de classificar automaticamente diatomáceas; em (MARCHANT *et al.*, 2019), os autores aplicaram RNCs para classificação de foraminíferos; in (MITRA *et al.*, 2019), Mitra *et al.* emprega uma RNC para identificar seis espécies de foraminíferos planctáticos existentes e distinguir estas de outros táxons; in (HSIANG *et al.*, 2019), Hsiang *et al.* construíram um conjunto de imagens de foraminíferos planctônicos, usando várias informações de especialistas e, em seguida, empregaram uma classificação de imagem baseada em RNC, comparando seus resultados da abordagem de classificação com a classificação realizada por humanos.

3.4.1 Discussão sobre os trabalhos analisados

No geral, os trabalhos encontrados realizam a classificação de micro-organismos de forma satisfatória. Entretanto, a maioria deles realiza uma análise classificatória sem preservar o local de deposição dos microfósseis. Este é um aspecto que neste trabalho é abordado, executando a (Segmentação Semântica (SS)) na amostra de volume completa, preservando assim o local de deposição e as informações dos microfósseis.

4 FUNDAMENTAÇÃO TEÓRICA

Este capítulo tem como propósito fornecer uma base sobre o processo de digitalização (alguns tipos existentes) e como os conjuntos de dados gerados são posteriormente processados. Para fundamentar este embasamento utilizou-se como ponto de partida a primeira etapa da revisão da literatura e o livro *techniques for virtual palaeontology* (SUTTON; RAHMAN; GARWOOD, 2014), os quais forneceram uma visão dos processos de digitalização e das técnicas de pré e pós-processamento empregadas.

4.1 PROCESSOS DE DIGITALIZAÇÃO

Fósseis muito delicados podem ser destruídos ou perderem sua estrutura ao serem utilizadas técnicas de isolamento manual. Com o processo de digitalização evitam-se possíveis danos estruturais e o uso de equipamentos caros para lidar com o fóssil, havendo apenas a necessidade de um computador para trabalhar com os dados digitalizados. Esse processo possibilita dissecar e seccionar o fóssil sem danificar a sua estrutura original. Por fim, ao se digitalizar permite-se também a disseminação do fóssil, facilitando análises e publicações colaborativas.

Entre as técnicas de digitalização existem técnicas destrutivas e não-destrutivas. Nas destrutivas existem métodos bastante consolidadas de *serial grinding*, *sawing*, *slicing* e entre outras, categorizadas como tomografias de origem física-ótica. Os métodos não-destrutivos são diversos e incluem muitas variantes, tais como Tomografia Computadorizada (TC), Tomografia utilizando nêutrons como fonte, (Imagem por Ressonância Magnética (IRM)) e Tomografia Óptica. Existem técnicas que tem como base a superfície, nestas a geometria de uma superfície externa é digitalizada de alguma forma (fotogrametria, escaneamento a laser e digitalização mecânica).

Historicamente, a TC tornou-se mais utilizada na área da paleontologia a partir de 1990, com a maior disponibilidade de tomógrafos e após a evolução destes aparelhos, utilizados principalmente na área médica. Entretanto, o estudo forte de invertebrados e pequenos fósseis começou apenas no século XXI, com o advento da (Microtomografia Computadorizada (Micro-CT)). A Micro-CT possibilitou trabalhar em escalas menores, tipicamente em resoluções de alguns microns, impulsionando muitos estudos. Por fim, com o advento das tomografias de raio-x com feixes de luz, foram proporcionadas instalações para o estudo de material paleontológico com grandes resoluções e alta confiabilidade (SUTTON; RAHMAN; GARWOOD, 2014).

4.1.1 Tomografia destrutiva

Tomografia destrutiva envolve a exposição física das superfícies. Todos os métodos de tomografia destrutiva consomem uma grande quantidade de tempo e danificam

ou destroem o espécime, entretanto podem produzir reconstruções de alta fidelidade. Métodos óticos-físicos envolvem exposição da superfície (através de técnicas de *grinding*, *slicing* e *sawing*) e imageamento ótico (fotografia ou tracejamento, algumas vezes através de cascas de acetato), cada uma com suas peculiaridades. O processo de tomografia destrutiva pode ser defendido como um processo que converte o espécime físico em digital, mesmo sem levar em conta a destruição do espécime, o processo como todo é necessariamente imperfeito. Limitações práticas na resolução do imageamento e espaço entre tomografias irá resultar em perda de dados, e o imageamento, indiferente da técnica utilizada, não conseguirá capturar todas as informações contidas em uma superfície exposta. Além disso, são técnicas que exigem muito cuidado e um intenso trabalho. Apesar de todos os aspectos negativos, a tomografia destrutiva permanece sendo a melhor alternativa para o estudo de alguns fósseis, dado que alguns tipos de materiais não são facilmente tratáveis através de técnicas não-destrutivas devido à insuficiente atenuação do contraste dos raios-X, para tomografia computadorizada, e do nêutron, para tomografia à base de nêutron, por exemplo (SUTTON; RAHMAN; GARWOOD, 2014).

Existem diversas técnicas de tomografia destrutiva. Destas, vale salientar a técnica de tracejamento, a qual historicamente era a forma utilizada na maioria dos conjuntos de dados de tomografias publicados. Nesta técnica, traços das estruturas de interesse a serem reconstruídas são produzidos e esses traços são utilizados ao invés das imagens de tomografia em qualquer reconstrução. As desvantagens deste método, além de ser uma técnica destrutiva, são: a exigência da inserção da subjetividade no processo de reconstrução e a grande quantidade de tempo exigido pelo processo.

4.1.2 Tomografia não-destrutiva

Tomografia não-destrutiva envolve um grupo de técnicas de escaneamento que podem criar um conjunto de dados tomográficos de um fóssil sem causar nenhum dano ao mesmo. Esta diversificada coleção de técnicas possui um alcance variado de princípios básicos. Por exemplo, a Tomografia Computadorizada (TC) constrói tomografias a partir de radiografias de uma amostra rodando através do mapeamento da atenuação do raio-X em três dimensões, sendo esta a técnica mais disseminada no estudo tomográfico dos fósseis. A tomografia com base em nêutrons é similar ao TC, mas, neste caso, são os nêutrons livres que fornecem a radiação. A imagem de ressonância magnética (IRM) utiliza campos magnéticos fortes os quais podem mapear o núcleo de elementos leves tal como o hidrogênio dentro de uma amostra; esta técnica é melhor empregada para o imageamento de tecidos biológicos suaves. Por fim, técnicas de tomografia óptica usam a luz visível para adquirir tomogramas, mas requerem uma matriz ou fósseis macerados translúcidos.

A Tomografia computadorizada é uma tecnologia de escaneamento que utiliza

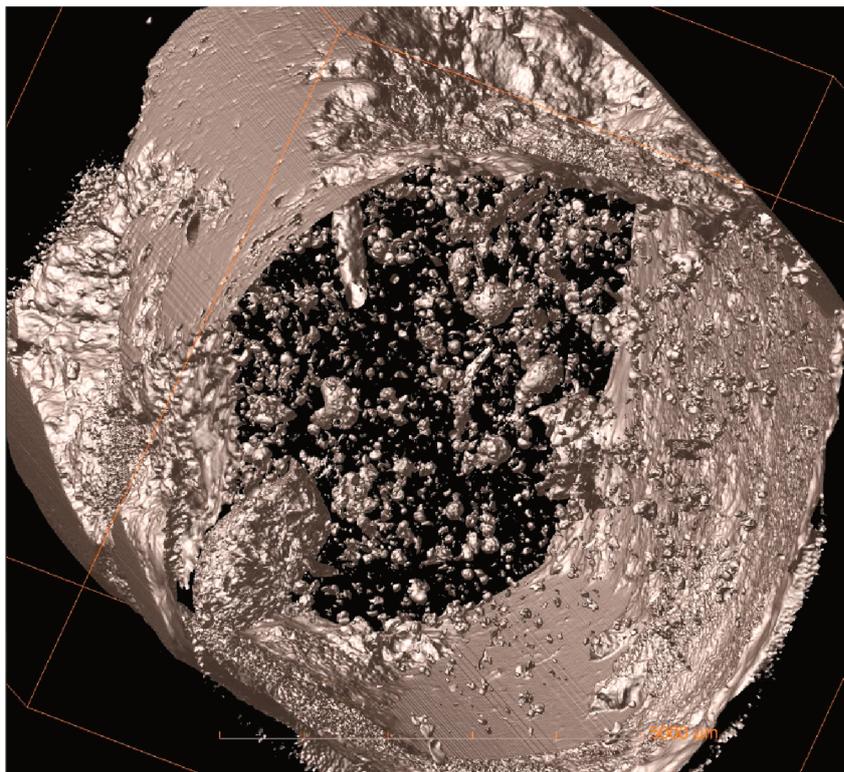
raios-X para criar um conjunto de dados tomográficos. Existem diferentes formas de escaneamento com TC que diferem em suas resoluções e aplicações. Alguns são de escala laboratorial (médica, micro e nano tomografia), enquanto outras empregam uma grande infra-estrutura (síncrotron). Entretanto, todas as variantes compartilham um princípio comum: uma fonte de raio-X e detectores são utilizados para adquirir as radiografias ou projeções de um objeto em múltiplos ângulos. Disto, uma sequência de tomografias paralelas e igualmente espaçadas é criada, a qual mapeia a atenuação do raio-X dentro de uma amostra. A atenuação do raio-X é dada em função de várias propriedades dos materiais, principalmente a composição/concentração do elemento e a densidade. A atenuação é a perda de intensidade na radiação-X à medida que ela passa através da amostra, resultado da absorção ou espalhamento desta radiação. Este processo de digitalização, por sua vez, gera alguns artefatos provenientes da parte física envolvida no processo, como por exemplo *beam hardening*, *ring*, *streak*, *noise*, *field of view*, *cone beam* e *partial volume averaging*. Para amenizar esse problemas são utilizadas algumas técnicas de aplicação de filtro, o aumento no número de aquisições para uma mesma posição, ajuste dos parâmetros de escaneamento e calibração do equipamento de digitalização.

4.2 VISUALIZAÇÃO DIGITAL

Atualmente existem muitos fluxos de trabalho para converter dados em modelos virtuais ou físicos. Destes, para a parte de dados tomográficos, existem uma variedade grande de técnicas com diferentes graus de complexidades. Os conjunto de dados tomográficos registrados são essenciais para a maioria das técnicas de visualização. Uma alternativa subsequente ao registro é o *vector surfacing* (SUTTON; RAHMAN; GARWOOD, 2014) que envolve o tracejado das linhas de contorno e sua posterior conversão em uma malha de triângulo, entretanto, atualmente, é raramente utilizada. Os trabalhos atuais dão maior preferência para a reconstrução de volumes, onde os *pixels* dos tomogramas são tratados como *voxels*, de forma que os volumes podem ser virtualmente preparados a fim de melhorar a precisão. A Figura 6, mostra um exemplo de renderização do volume empregado nos experimentos. Podem também ser renderizados diretamente, ou através de *isosurfacing* para gerar uma malha de triângulos. Esta malha de triângulos pode ser pós-processada e subsequentemente visualizada utilizando hardware gráfico, *Ray tracing* ou impressão 3D.

Quando se fala em conjuntos de dados tomográficos registrados isto diz respeito a um conjunto de tomogramas corretamente alinhados um com os outros. Muitas técnicas de tomografia fazem esta etapa de registro automaticamente, entretanto, uma etapa separada de registro é normalmente requerida para dados gerados por tomografia física-óptica. Existem algumas premissas assumidas pelas metodologias de reconstrução e sua violação pode degradar a metodologia de reconstrução sendo

Figura 6 – Volume renderizado para visualização .



utilizada. As premissas assumidas são: equivalência de *pixels* entre tomogramas; tomogramas paralelos; espaço entre tomogramas conhecido; independência entre fatias. O registro é utilizado a fim de manter a primeira premissa, corrigindo problemas de rotação, deslocamento, e, algumas vezes, problemas de escala, garantindo o correto alinhamento entre tomogramas.

Após o registro, duas técnicas bem diferentes podem ser usadas para reconstrução. A primeira é a *vector surfacing*, a qual envolve a identificação, manual ou semi-automática de estruturas de interesse em cada tomograma na forma de *vector-graphic objects*, normalmente laços fechados definidos como curvas *splines*. Estas *splines* de duas dimensões são empilhadas em um espaço virtual tridimensional e utilizadas para gerar uma superfície tridimensional matematicamente definida, tipicamente como uma malha de triângulos. A técnica de *vector surfacing* possui algumas vantagens sobre as técnicas de reconstrução baseadas em volumes tais como: um melhor resultado quando o espaçamento entre tomogramas é relativamente grande e uma adaptação relativamente simples para o conjunto de dados. Enquanto não se pode recriar a informação perdida entre tomogramas amplamente espaçados, interpolações suaves podem ser geradas. Esta técnica produz malhas de triângulos, as quais são mais compactas, produzindo assim menos triângulos que as técnicas de *isosurfacing*. Por fim, esta técnica não é sensível a variação de luminosidade ou outra condição óptica entre tomogramas, enquanto o tracejamento pode ser realizado, a abordagem é viável.

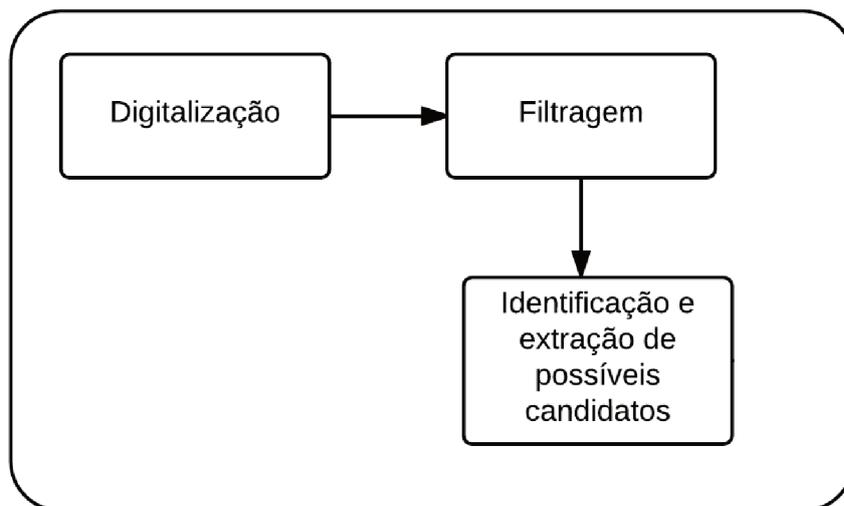
Esta técnica, entretanto, necessita da interação do usuário sendo, portanto, custosa e inserindo subjetividade ao processo. Logo, é recomendada apenas para conjuntos de dados físico-óticos, onde o espaçamento entre tomogramas é grande e os objetos são relativamente simples. Nenhuma metodologia de reconstrução irá reconstruir adequadamente espécimes onde o espaçamento entre tomogramas é muito grande, tanto para um humano quanto para um algoritmo para determinar a conectividade entre objetos.

A segunda técnica é a reconstrução baseada em volume, uma alternativa à técnica de *vector surfacing*, e a mais utilizada nos estudos paleontológicos neste século. Em vez de modelar o fóssil com *vector-graphic objects*, uma abordagem baseada em *pixel (raster-based)* é utilizada. O conjunto de dados é tratado como um volume composto de *voxels* cúbicos ou sub-cúbicos. Da mesma forma que uma imagem bidimensional é composta de uma grade retangular de *pixels* de mesmo tamanho, um volume é composto de uma grade tridimensional de *voxels* de mesmo tamanho. Um conjunto de dados tomográficos registrados com espaçamento entre tomogramas regular, não necessita de conversão para um volume, precisa apenas da mudança conceitual de considerar um *pixel* como *voxel*. Um volume pode ser visualizado em uma das duas formas: através da renderização direta do volume ou pela extração de *isosurfaces*, ambas podem ser realizadas automaticamente e de forma rápida. A grande vantagem da reconstrução com base em volume é sua velocidade e a simplicidade da perspectiva do usuário, a visualização pode ser feita rapidamente e com o mínimo de intervenção, dado que não há uma etapa de interpretação. A reconstrução com base em volume é tão visualmente satisfatória quanto a utilizando *vector surfacing*, logo reconstrução com base em volume, seja por renderização direta ou por *isosurfacing* (utilizando o algoritmo *marching cube* (LORENSEN; CLINE, 1987), por exemplo) é o método escolhido pela grande maioria dos conjuntos de dados tomográficos, paleontológicos ou outros.

4.3 PROCESSAMENTO DIGITAL 3D

Com o modelo digital em mãos, os próximos passos englobam a aplicação de técnicas de processamento digital 3D. Tais técnicas passam por métodos de filtragem, segmentação, detecção de características da imagem, como por exemplo bordas e padrões, e a classificação/reconhecimento e rastreamento de objetos de interesse. Com o propósito de correlacionar esses métodos com a área de aplicação sendo estudada, propõe-se um *pipeline* inicial básico, representado pela figura 7. Sob este *pipeline* básico, será fornecida uma explicação sobre os métodos base utilizados, mostrando exemplos voltados para a área da paleontologia.

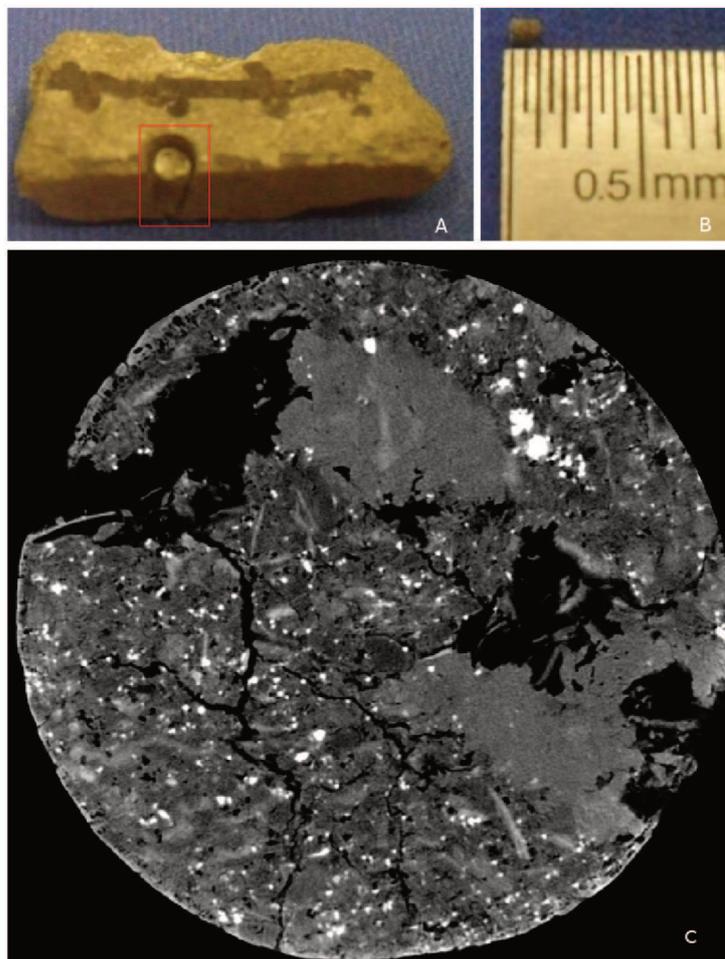
Figura 7 – Pipeline inicialmente proposto.



4.3.1 Digitalização

A etapa da digitalização tem como propósito transformar o modelo físico em um modelo virtual, de modo não-destrutivo, com a finalidade de executar procedimentos sobre o elemento de estudo preservando a integridade do mesmo. Na Figura 8-A, apresenta-se uma amostra com destaque para a parte utilizada na digitalização, e em destaque na figura 8-B, a parte da mesma a qual foi extraída para ser digitalizada utilizando o Micro-CT. Na figura 8-C mostra-se uma fatia resultante da digitalização desta amostra. Os dados do equipamento utilizado são: Versa XRM-500, melhor resolução $0.7 \mu m$, voltagem 30-160 kV, potência 2-10 W, câmeras CCD 2048x2048 *pixel*, lentes 0.4X, 4X, 10X, 20X e 40X, capacidade de massa máxima 15 kg e tamanho limite de amostra (diâmetro/altura) 80/300 mm. Os dados da aquisição foram: resolução $1.08 \mu m$ (tamanho do *pixel* da imagem), tamanho da imagem 956x1004x983, sem filtragem para correção de *beam hardening*, lente óptica 10x, 30kV/2W, passo angular $0,255^\circ$ e tempo de exposição 11 segundos.

Figura 8 – Demonstração de parte do procedimento para digitalização. (A) Amostra evidenciando a parte utilizada para a digitalização. (B) Parte extraída da amostra apresentada em escala. (C) Exemplo de uma fatia resultante da digitalização da parte extraída da amostra.



4.3.2 Filtragem

O processo de digitalização é conhecido por gerar alguns artefatos tais como: *beam hardening*, *ring*, *streak*, ruído, entre outros. Para tanto existe a necessidade da aplicação de filtros para corrigi-los. Os filtros comumente utilizados para remoção de ruído são mediano, gaussiano e Kuwahara. Para a remoção de *beam hardening*, por ser um artefato gerado constantemente nos tomógrafos, existe a opção de aplicar filtro para sua remoção diretamente no software de digitalização.

Desta forma, o processo de filtragem se torna necessário para corrigir os artefatos gerados pelo processo de digitalização, justificando assim sua presença no *pipeline* proposto. Nas figuras 9 e 10 são mostrados exemplos das aplicações dos filtros mediana e gaussiano, respectivamente, em uma amostra de rocha.

Figura 9 – Exemplo de aplicação de filtro mediana.

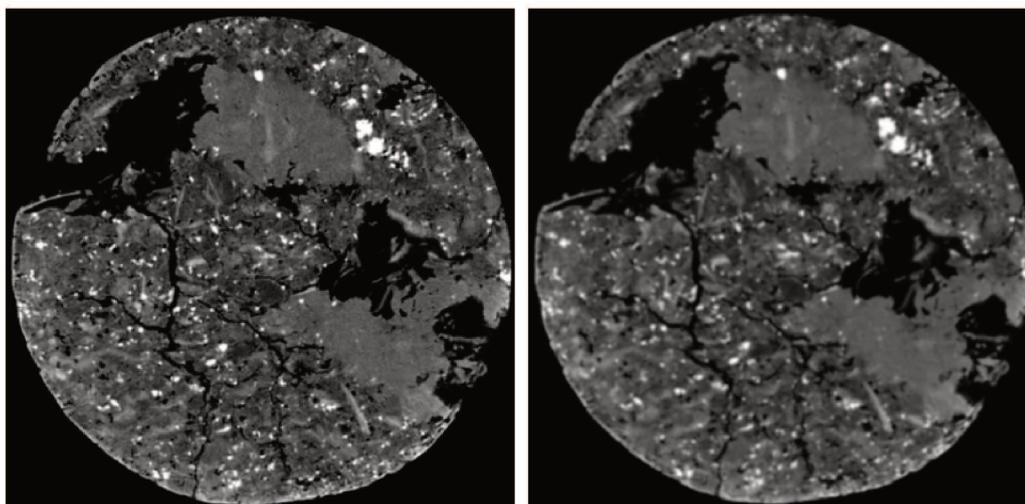
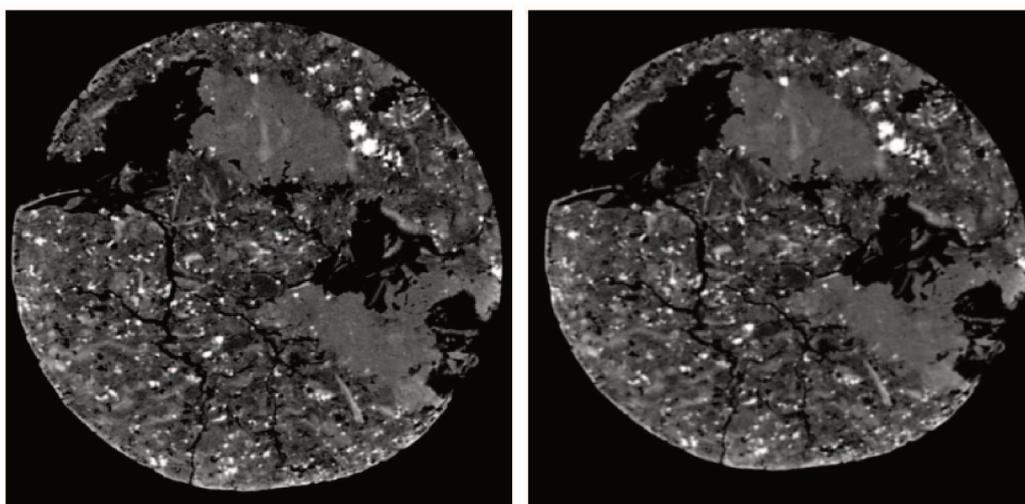


Figura 10 – Exemplo de aplicação de filtro gaussiano.



4.3.3 Identificação e extração de possíveis candidatos

O passo mais importante do *pipeline* é a identificação e extração de possíveis candidatos. Esse procedimento atualmente é realizado manualmente por isolamento físico, utilizando produtos químicos para preservar o fóssil e destruir a rocha, ou de forma digital através da investigação completa e manual da rocha digitalizada em busca de cada microfóssil na mesma. Para esta etapa duas vertentes da (Visão Computacional (VC)) foram empregadas.

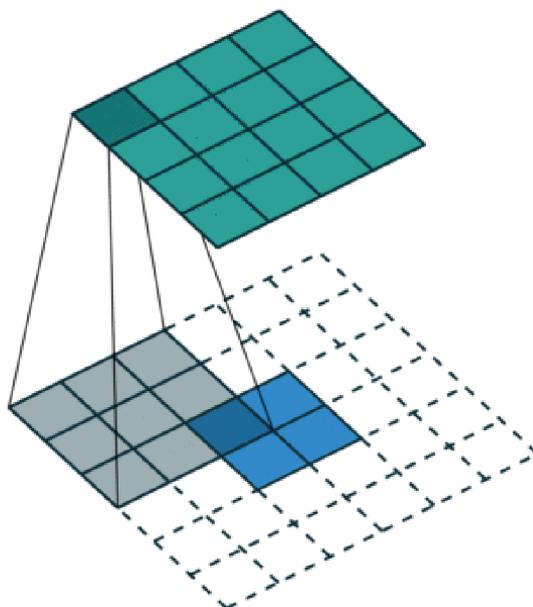
A primeira vertente é a VC clássica, onde se encontram os métodos mais tradicionais da VC. Entre estes podem-se destacar os métodos de segmentação de imagem 2D, *thresholding* simples e utilizando o método OTSU (OTSU, 1979), e 3D, contornos ativos (KASS; WITKIN; TERZOPOULOS, 1988) e Watershed (SIJBERS *et al.*, 1997), métodos de registro (BROWN, 1992) e rastreamento de objetos (WU; LIM; YANG, 2013).

A segunda vertente são os métodos englobados na visão computacional moderna. Destes, o atual maior representante são os métodos com base em *Deep learning* (DL) (O'MAHONY *et al.*, 2019). Nos últimos anos o uso de inteligência artificial, mais especificamente DL, teve um intenso crescimento. Para análise de imagens, por exemplo, o uso de DL para segmentação, filtragem, restauração de imagens, classificação e identificação de objetos tornou-se objeto de estudo de muitos pesquisadores.

O DL é uma forma de aprendizado de máquina que permite aos computadores aprender com a experiência e entender o mundo em termos de uma hierarquia de conceitos (GOODFELLOW; BENGIO; COURVILLE, 2016). O DL emprega Redes Neurais Convolucionais (RNC) muito profundas, com redes que às vezes consistem em mais de 100 camadas, em contraste com as Redes Neurais Artificiais (RNAs) empregadas entre as décadas de 1980 e 2000, que normalmente empregavam apenas três camadas. Um conceito-chave é a (Camada Convolutiva (CC)), uma estrutura de extração de recursos, apresentada pela primeira vez em (LECUN *et al.*, 1998), que permite o aprendizado hierárquico e a representação de conhecimentos complexos. A estrutura chave que permite esta representação de conhecimentos é a camada convolutiva, a qual pode ser resumida por um conjunto de operações de convolução, Figura 11 mostra o exemplo de uma operação de convolução. Este conjunto de operações de convolução aplicados na imagem extraem características as quais são combinadas nas camadas posteriores (sequência de camadas de *pooling* e convolucionais), formando conceitos mais complexos. A Figura 12, mostra um exemplo da extração de características de uma imagem de um gato, que posteriormente são combinadas para, por exemplo, identificar se existe um gato na foto.

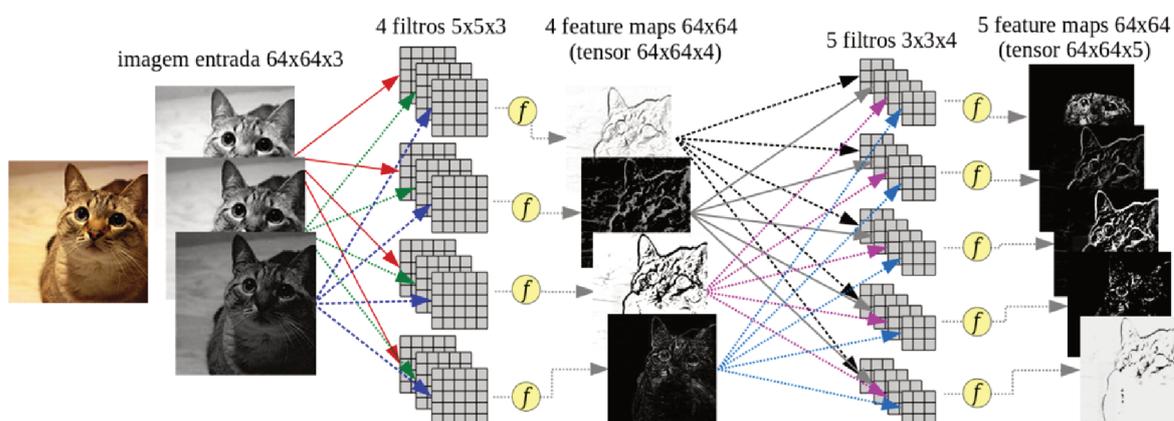
Como as RNC coletam conhecimento a partir de exemplos, não é necessário que um humano especifique formalmente todo o conhecimento que o computador precisa. A capacidade de representar uma hierarquia de conceitos em uma rede com

Figura 11 – Exemplo de uma operação de convolução.



Fonte: Imagem extraída de (DUMOULIN; VISIN, 2016)

Figura 12 – Exemplo de uma rede convolucional extraindo características de uma imagem para combinar em conceitos mais elaborados.



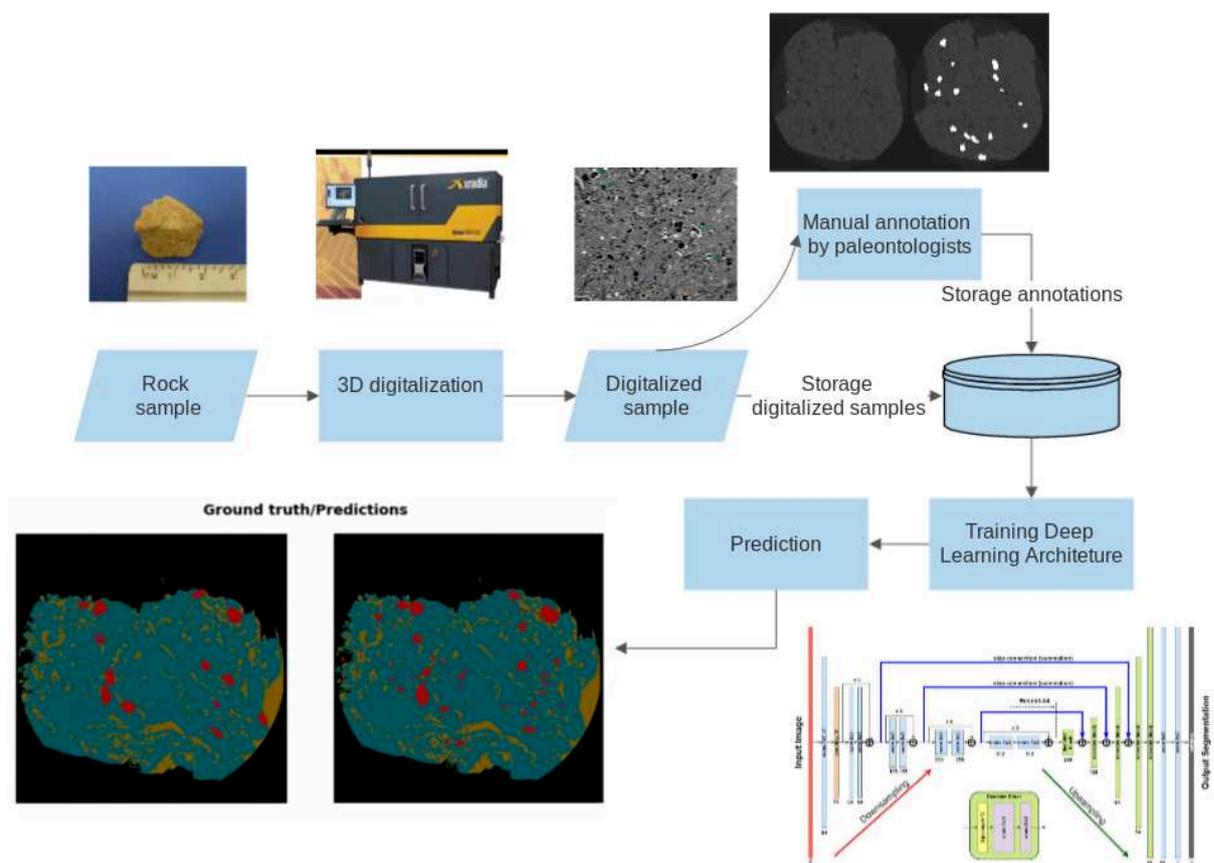
Fonte: Imagem extraída de (PONTI; COSTA, 2018)

dezenas de CCs permite que uma RNC profunda aprenda conceitos complicados construindo-os a partir de conceitos mais simples (GOODFELLOW; BENGIO; COURVILLE, 2016). O próximo capítulo demonstra como foram exploradas as duas vertentes e os resultados obtidos da aplicação das mesmas na área de micropaleontologia.

5 DESENVOLVIMENTO

Com base no *pipeline* inicial básico apresentado e nos estudos do atual estado da arte voltados para segmentação 3D e reconhecimento/classificação de objetos 3D, montou-se um *workflow* voltado para a identificação e extração de microfósseis. Tal descrição dos processos executados nesse trabalho passa pelos pipelines apresentados anteriormente nas Figuras 4 e 5 e pela análise do estado da arte descrita anteriormente. A Figura 13 apresenta todos os passos executados e resumidos em um *workflow*.

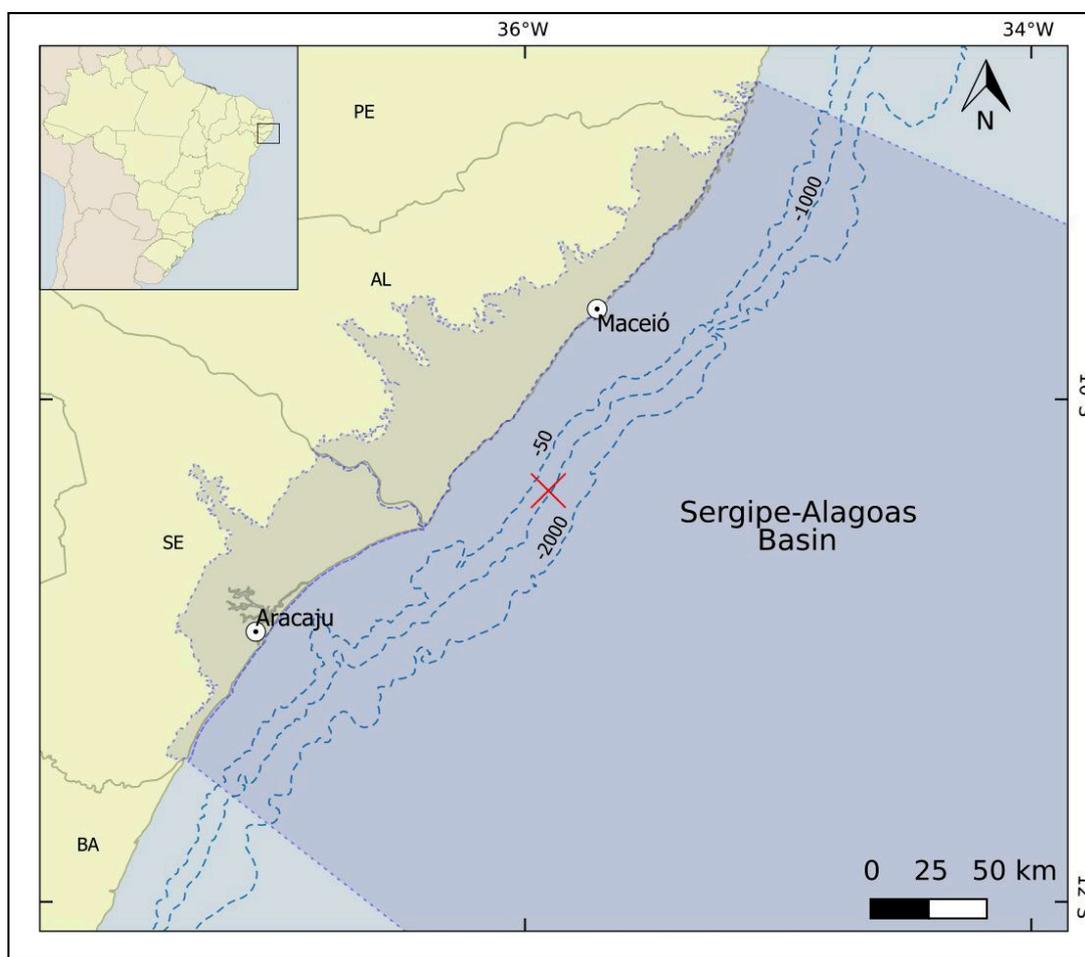
Figura 13 – *Workflow* dos processos executados nesse trabalho.



Para os experimentos, empregou-se uma amostra de rocha carbonática obtida de uma plataforma de perfuração coletada em sedimentos quaternários da bacia de Sergipe (Figura 14).

A parte de aquisição dos dados empregada neste trabalho continua sendo a mesma previamente descrita na seção de digitalização. A Figura 15 apresenta uma das amostras de rocha empregada e uma fatia 2D do resultado da digitalização. O conjunto de dados que contém as imagens segmentadas manualmente pelos espe-

Figura 14 – Mapa da Bacia de Sergipe com o local exato de extração de amostras de rocha marcado com uma cruz vermelha. A amostra foi coletada a uma profundidade de aproximadamente 2.500 metros.



cialistas e os dados do MicroCT estão disponíveis em: <http://www.lapix.ufsc.br/microfossil-segmentation>. A amostra sendo analisada contém uma grande variedade de espécimes e espécies. No caso foram considerados 15 espécies entre os mais de 4600 espécimes existentes na rocha.

Com intuito de analisar as principais características dos tipos de microfósseis sendo pesquisados empregou-se o mesmo processo de digitalização da amostra para a geração do modelo digital de cada uma destas espécies. A figura 16 mostra os exemplares de 15 espécies considerados no estudo das amostras de rochas sendo analisadas.

Esta figura foi gerada a partir do isolamento físico, realizado em laboratório seguindo precauções específicas de forma que não houvesse nenhuma mudança química e/ou física, de cada um dos espécimes seguido pela digitalização com um microscópio eletrônico de varredura. A seguir encontra-se descrito todo o processo para aquisição de tais imagens:

Figura 15 – Amostra de rocha analisada (A) e uma de suas seções 2D de microtomografia (B).



(i) O sedimento primeiramente foi imerso em água deionizada por aproximadamente 24 horas, objetivando a desagregação química; (ii) lavagem com água corrente, em peneira de $63\mu\text{m}$; (iii) secagem em estufa do material em 40°C , por aproximadamente 48 horas. Após a secagem das amostras, foi realizado a triagem e seleção dos principais microfósseis representativos na amostra em uma lupa. Neste trabalho os espécimes de microfósseis foram estampados com auxílio da aquisição multidimensional de um estereoscópio Zeiss Discovery V20 (modo Z-stak no software AxioVision 4.8).

Os modelos digitalizados das espécies presentes na rocha tem grande importância no momento de se analisar quais as características que auxiliam na extração dos mesmos da rocha. A figura 17 mostra a fatia mais central do modelo digitalizado para alguns dos microfósseis encontrados na amostra analisada.

5.1 PRÉ-PROCESSAMENTO

Esta etapa, presente em apenas alguns dos trabalhos analisados, envolve a possível aplicação de alguns dos seguintes métodos: Filtragem, Seleção de regiões de interesse e normalização dos dados. Neste trabalho, a filtragem teve sua utilização justificada pelo processo de digitalização, o qual produz artefatos inerentes ao processo de captura pela TC. Alguns exemplos destes artefatos são: *beam hardening*, ruídos e anéis (BOAS; FLEISCHMANN, 2012). Dado o conhecimento sobre tais artefatos, alguns métodos podem ser aplicados nesta etapa afim de minimizar a influência destes nas imagens reconstruídas (BOAS; FLEISCHMANN, 2012)

5.2 CONJUNTOS DE DADOS/BASES DE DADOS DISPONÍVEIS

Até o dado momento não foram encontradas bases de dados tomográficos voltadas para a área de microfósseis. Logo, todos os experimentos e resultados obtidos foram produzidos com base em amostras de rochas e espécimes cedidas em parceria com o *Instituto Tecnológico de Micropaleontologia - itt Fossil* - e digitalizados no *Laboratório de Meios Porosos e Propriedades Termofísicas LMPT - UFSC - EMC*. Esta base de dados encontra-se disponível em <http://www.lapix.ufsc.br/microfossil-segmentation>

Figura 16 – Microfósseis sendo analisados nos estudos das amostras em questão. Foraminíferos planctônicos: 1) *Globogerinoides ruber*; 2a-b) *Candeina nitida*; 3) *Orbulina universa*; 4) *Globigerinoides trilobus saculifera*; 5) *Globigerinoides trilobus*; 6a-b) *Globorotalia truncatulinoides*. Foraminíferos bentônicos: 1) *Bulinina*; 2) *Bolivinita*; 3a-c) *Cibicidoides*; 4) *Laticarinina*; 5) *Uvigerina*; 6) *Sphaeroidina*; 7) *Siphonaperta*; 8) *Quinqueloculina*

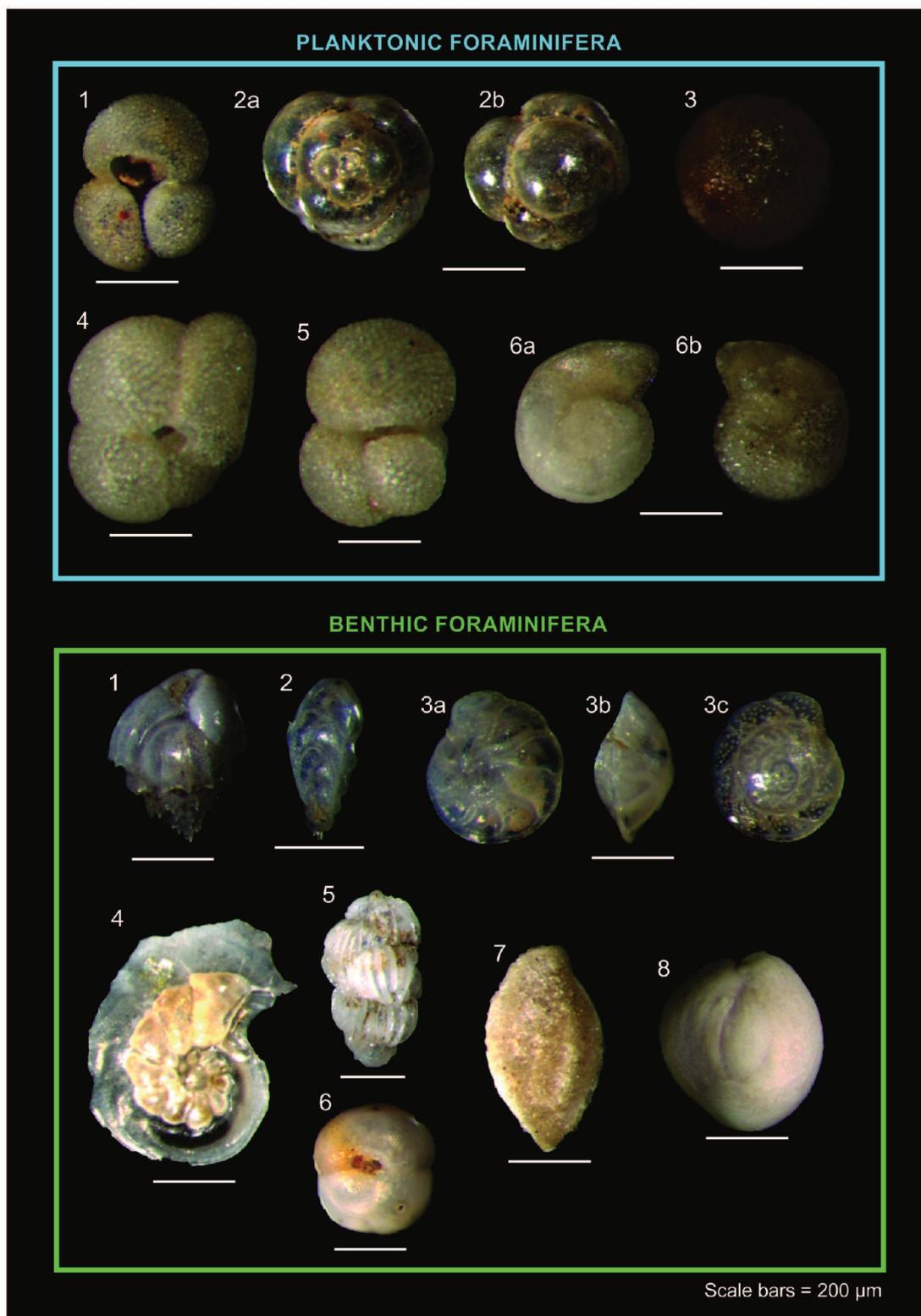
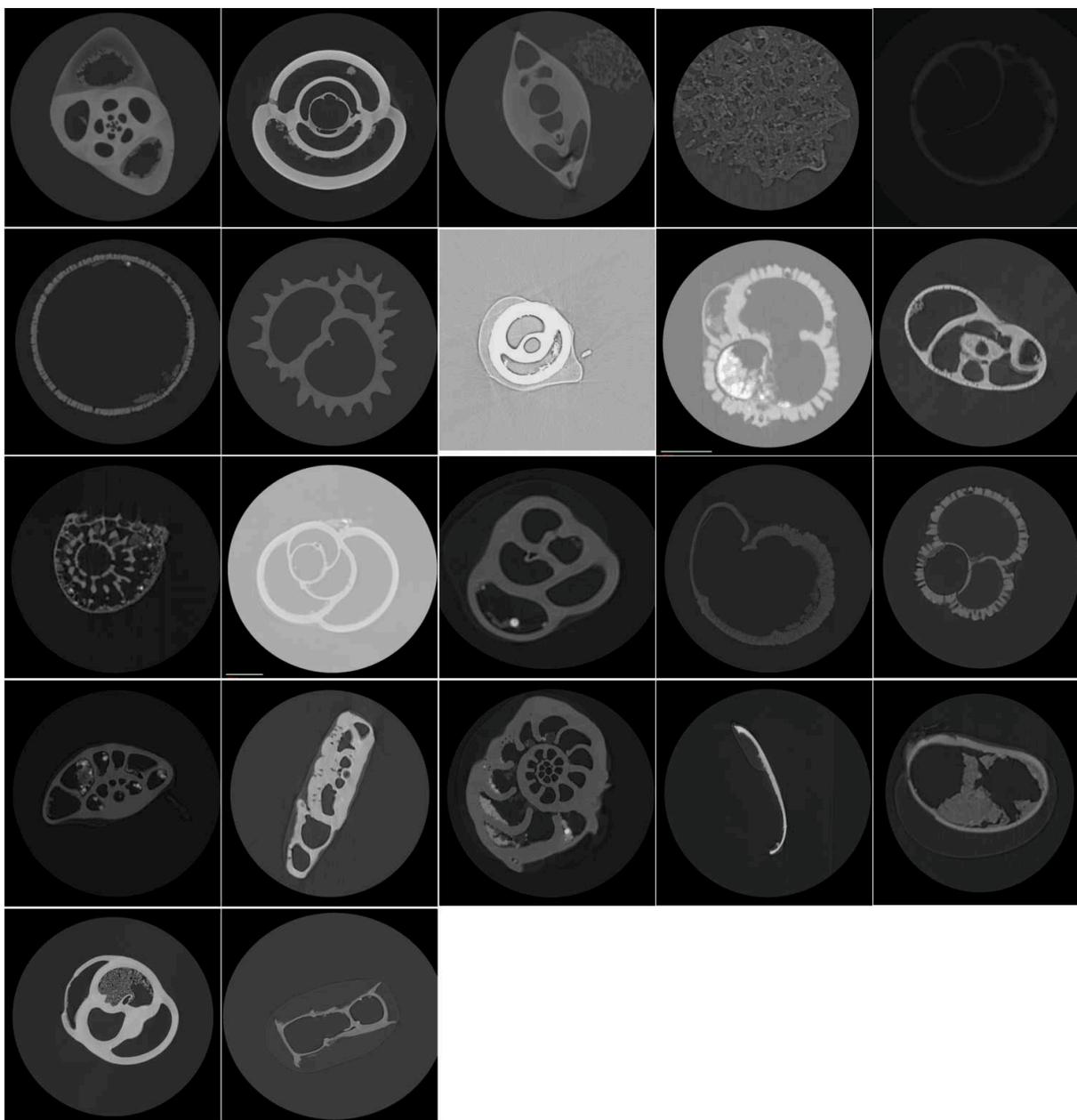


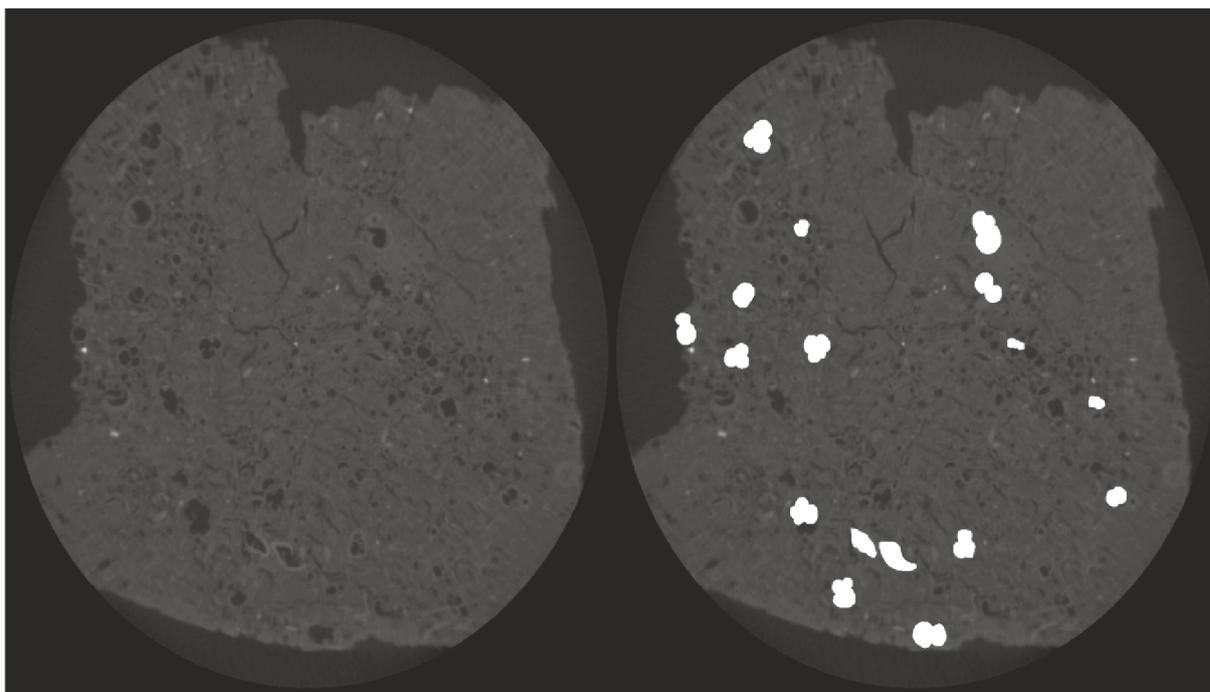
Figura 17 – Modelo digital de alguns dos microfósseis isolados fisicamente e que estão presentes na amostra de rocha sendo analisada.



5.3 FORMA DE VALIDAÇÃO E DEMONSTRAÇÃO DOS RESULTADOS

Na validação dos resultados, foi utilizado o conhecimento de três especialistas na área na de paleontologia, nos termos de determinação da localização, com base em (Ground Truth (GT)), vide figura 18. Para produzir estes padrões ouro para fins de comparação, foram necessários 3 meses. Tamanho duração pode ser justificada através dos seguintes fatores: escala dos espécimes sendo utilizada; disponibilidade do tomógrafo, o qual além de custoso em termos de valor para digitalização estava sujeito a filas e falhas; Processo de anotação executado pelos paleontólogos. Em razão do caráter 3D das amostras digitalizadas, o processo de identificação manual dos microfósseis é extremamente demorado, pois os especialistas necessitavam estimar a localização do microfóssil e averiguar nas fatias posteriores e anteriores se de fato aquela marcação era um microfóssil.

Figura 18 – Fatia de uma amostra e seu respectivo GT, utilizado como padrão ouro na identificação da localização dos microfósseis.



Com o intuito de avaliar um método para a área de segmentação de imagens e evitar a análise subjetiva utilizam-se mecanismos de validação. Um destes mecanismos é a validação baseada em GT. Esta forma de validação é uma técnica a qual compara o resultado de segmentação obtido pelo algoritmo sendo avaliado com as segmentações realizadas manualmente por especialistas, o GT, resultando em um valor ou índice de similaridade entre o GT e o resultado do algoritmo. A figura 18 mostra o exemplo de um GT extraído de uma imagem, o qual foi realizado manualmente por especialistas.

Existem diferentes métodos para validação por GT, alguns exemplos são: Rand (RAND, 1971), Fowlkes-Mallows (FOWLKES; MALLOWES, 1983), Jaccard (BEN-HUR; ELISSEEFF; GUYON, 2002), Dongen (DONGEN, 2000) e Dice (SØRENSEN, 1948). A escolha pelo uso do coeficiente Dice, mais especificamente uma variação do Dice, Intersect Over Union (IOU)(RAHMAN; WANG, 2016), tem por base o constante uso na literatura de tal métrica para avaliação de resultados de segmentação (REZATOFIGHI *et al.*, 2019) (NAGENDAR *et al.*, 2018).

5.4 IDENTIFICAÇÃO E EXTRAÇÃO DE POSSÍVEIS CANDIDATOS

Os experimentos realizados visaram a identificação e isolamento dos microfósseis nas amostras tomográficas utilizadas. Apenas esta etapa, se realizada com exatidão, pode auxiliar na análise tafonômica, a qual se refere ao estudo de organismos em decomposição ao longo do tempo e como se fossilizaram, caso isto tenha ocorrido (MARTIN, 1999). Para tanto, inicialmente delimitou-se a área sendo utilizada a fim de reduzir o tempo computacional exigido pelas técnicas e a quantidade de memória necessária. Este segundo fator é extremamente relevante, pois as amostras digitalizadas podem chegar a uma resolução de 1000 até 2000 ao cubo e, ao utilizar uma representação com 16 ou 8 bits por pixel, tem-se o uso de uma grande quantidade de memória para apenas carregar a imagem para o processamento (para 16 bits e uma resolução de 2000 ao cubo tem-se algo em torno dos 8 GB apenas para carregar a imagem).

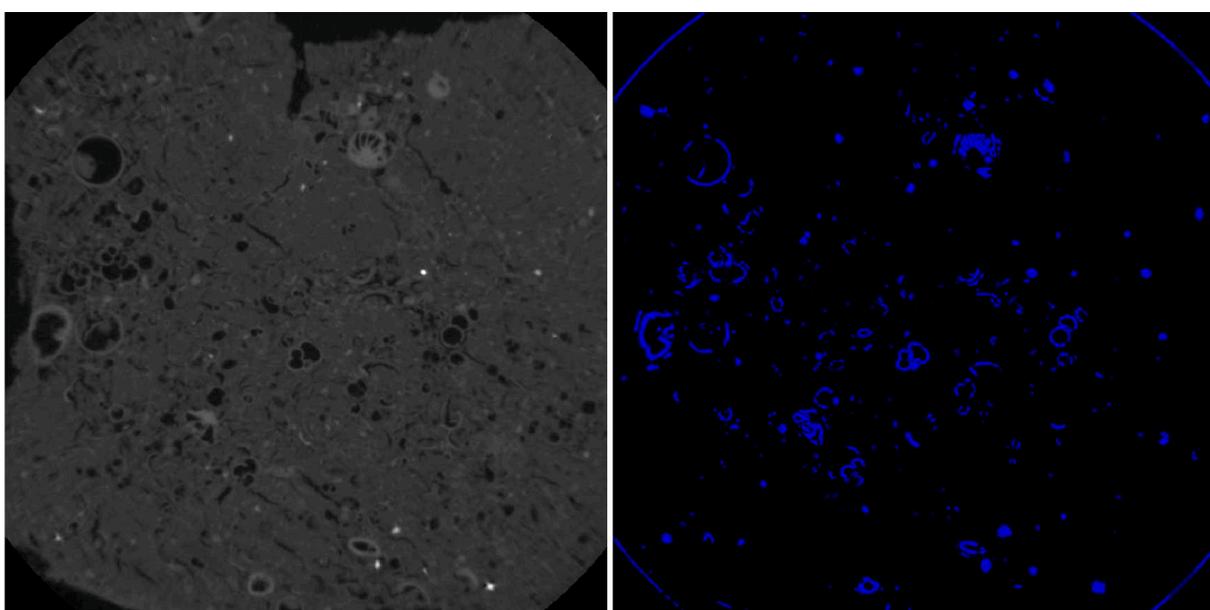
Após a delimitação da área de análise, passa-se para um processo de filtragem opcional para a remoção de artefatos provenientes do processo de digitalização, quando necessário. Subsequentemente se realiza a segmentação dos candidatos de interesse.

5.4.1 Visão Computacional Clássica

Na segmentação dos candidatos de interesse, inicialmente, vários experimentos foram realizados utilizando técnicas clássicas de Visão Computacional (VC), desde o simples método de segmentação por *thresholding* até métodos mais complexos como contornos ativos, apresentado na figura 19, e uma associação entre métodos de segmentação e classificação, apresentado na figura 20. Outros métodos de VC foram testados: detectores de borda, apresentado na figura 21, rotulamento por componentes conexos e filtros de tamanho aplicados sobre os mesmos, apresentados nas figuras 22 e 23. Um *pipeline* é demonstrado na figura 24, onde demonstra-se a aplicação de um conjunto de técnicas: Classificação, detector de borda, binarização e o operador lógico *And* entre os resultados das imagens de borda e do classificador. Outro sequenciamento de métodos com resultados promissores, utilizando técnicas

simples e automáticas é apresentado na figura 25. Neste *pipeline* foram utilizados os seguintes métodos: delimitação da região de interesse, parametrizado pelo usuário; aprimoramento das características de borda através de ajuste do contraste e do brilho da imagem; detecção de bordas pelo gradiente; *thresholding* utilizando o método OTSU; rotulamento por componentes conexas e por fim uma filtragem opcional com base no tamanho da componente. Aplicando o resultado deste processamento como uma máscara sobre a imagem original obteve-se um resultado o qual fornece as possíveis localizações dos microfósseis candidatos.

Figura 19 – Aplicação de uma técnica de segmentação por contornos ativos ao volume sendo analisado.



Como a parte de segmentação tem uma influência direta na qualidade do reconhecimento dos microfósseis, os experimentos com tais métodos foram extensivos com a finalidade de procurar o melhor algoritmo e os parâmetros mais adequados. Com esta política de testes, realizou-se uma bateria de testes com o método de contornos ativos implementado em 3D. A necessidade de tal bateria surgiu do questionamento sobre qual seriam os melhores parâmetros a serem utilizados de forma a se obter o melhor resultado possível. Os parâmetros do algoritmo de contornos ativos utilizado são: Número de passos, Sigma, Alpha, Suavização e Theta. Os testes buscaram avaliar qual a melhor combinação destes parâmetros a fim de obter o melhor resultado de segmentação. Para tanto, inicialmente, aplicou-se uma variação em cada um dos parâmetros, mantendo os demais fixos de forma a verificar a influência de tal variação. O próximo passo foi avaliar como a variação de um parâmetro afeta o outro. Estes dois passos associados geraram um número gigantesco de imagens e estas foram avaliadas de forma a delimitar a melhor faixa dos parâmetros para tal amostra. A Tabela 0

Figura 20 – Aplicação de uma sequência de métodos ao volume sendo analisado: classificação e binarização.

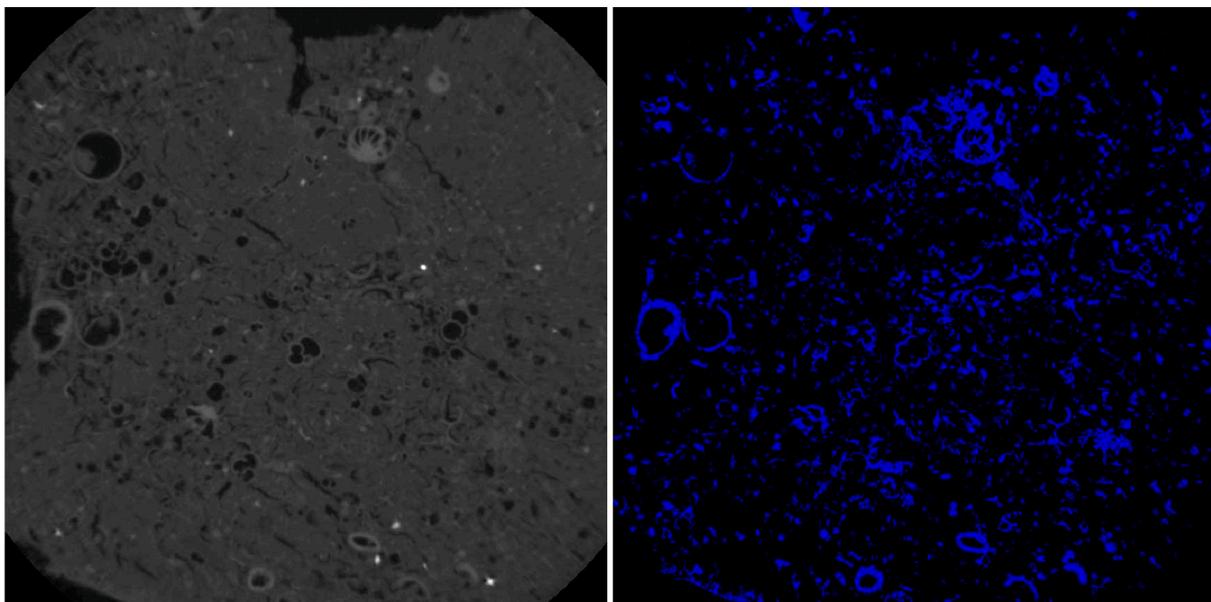
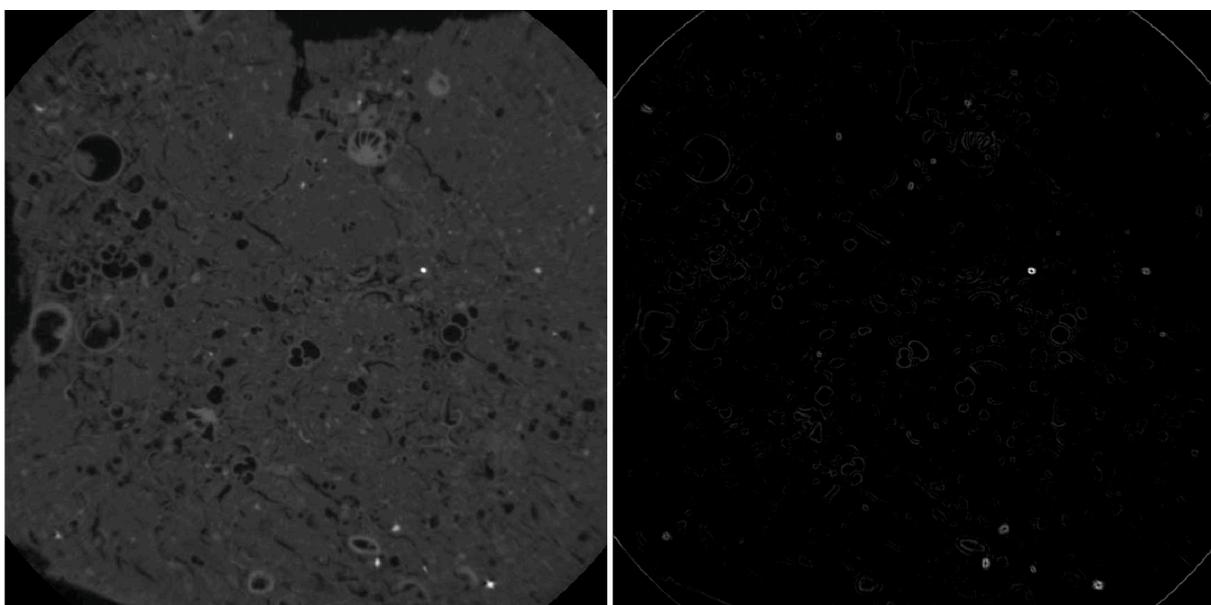


Figura 21 – Aplicação de uma técnica de detecção de borda ao volume sendo analisado.



mostra o resultado desta análise dos parâmetros.

A fim de tornar essa procura por melhores parâmetros mais automática, utilizou-se um algoritmo genético de forma a buscar uma otimização dos parâmetros de entrada do método de contornos ativos. No algoritmo genético utilizado, empregou-se o IOU como função de *fitness* a fim de avaliar os resultados obtidos com os parâmetros

Figura 22 – Aplicação de uma técnica de filtro de tamanho das componentes (> 50000) sobre rótulos realizado sobre o resultado encontrado no sequenciamento representado pela figura 25.

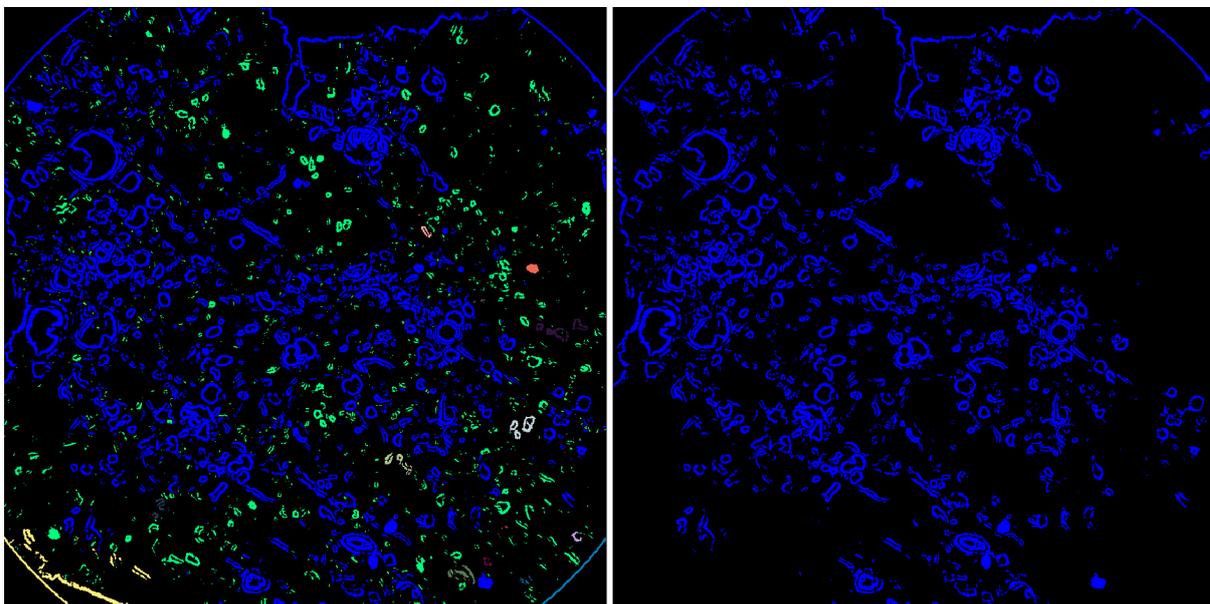
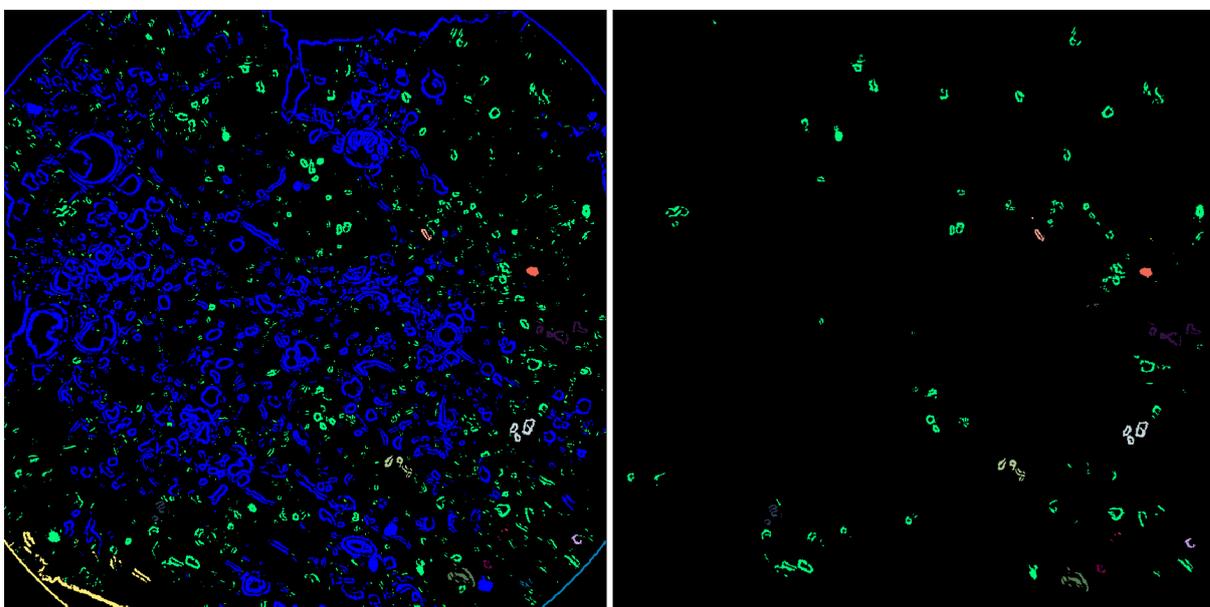


Figura 23 – Aplicação de uma técnica de filtro de tamanho das componentes (entre 200 e 9000) sobre rótulos realizado sobre o resultado encontrado no sequenciamento representado pela figura 25.



utilizados em comparação com o GT. Mesmo com excessivos testes, os resultados de segmentação não foram muito satisfatórios, gerando um valor próximo a 20% de IOU, e, dado que a segmentação é um passo de extrema importância, decidiu-se por

Figura 24 – Aplicação de um conjunto de técnicas: Classificação, detector de borda, binarização e o operador lógico *And* entre os resultados das imagens de borda e do classificador.

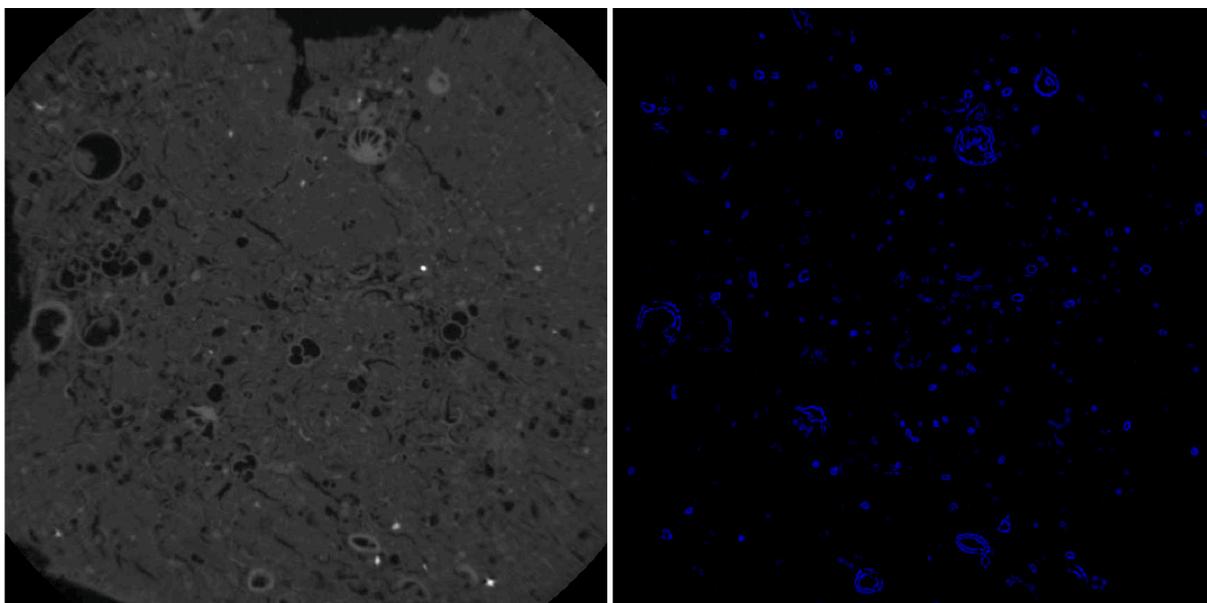
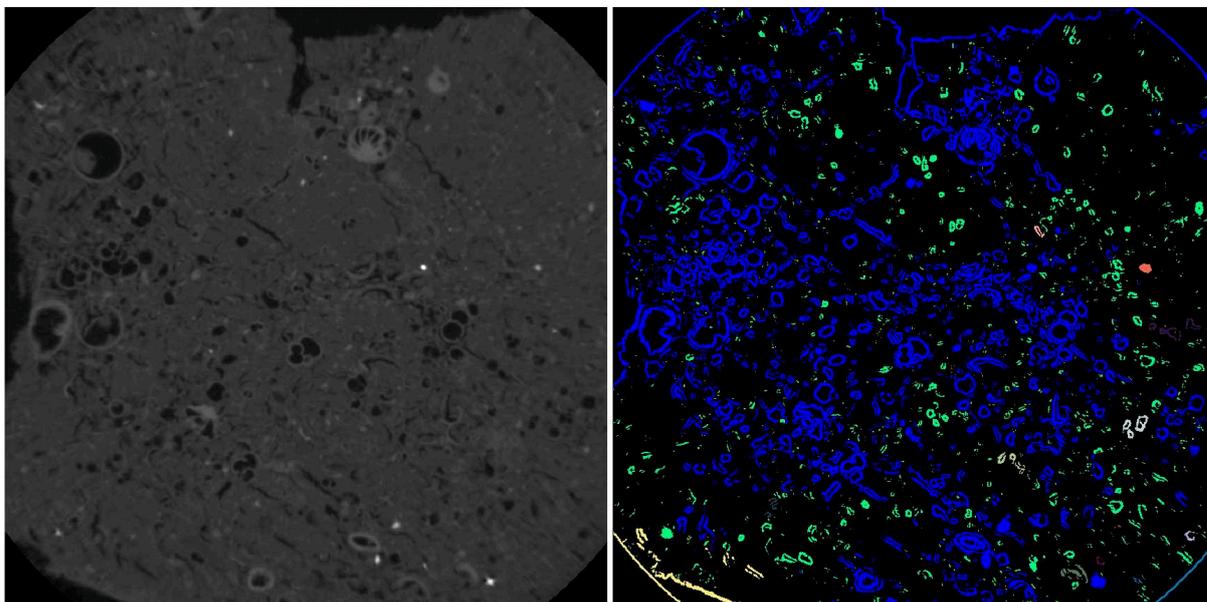


Tabela 0 – Tabela resumando o resultado da bateria de testes para análise dos parâmetros do método de contornos ativos.

Número de passos	Sigma	Alpha	Suavização	Theta
Para o caso analisado com os mesmos parâmetros o número de passos não interferiu. Sugerido: Manter em 100 o número de passos.	A medida que aumenta o sigma restringe o número de contornos. Sugerido: variar de 0 até 3.	A medida que aumenta o Alpha perdem-se os contornos, utilizando o resto dos parâmetros fixos. Entretanto pode interagir com o aumento do Sigma possibilitando o uso do Alpha em valores maiores. Sugerido: Variar entre 500 até 1000 conforme o parâmetro sigma utilizado	Em 1 as bordas já perdem um pouco de sua representatividade. De 0 até 0,5 sem grande alteração. Sugerido: variar de 0,1 em 0,1 de 0,5 até 1.	Ao fazer uma variação de 10 em 10 começando em 0 e indo até 100 verificou-se que nos valores iniciais (0 até 15) a quantidade de bordas é pequena. Por outro lado se o valor for acima de 40-45, dependendo dos demais parâmetros, os contornos são totalmente perdidos. Sugerido: Variar de 5 em 5 começando em 15 indo até o 40.

explorar outros métodos de segmentação.

Figura 25 – Aplicação de um conjunto de técnicas: Ajuste de brilho e contraste, binarização, detector de borda e rotulamento por componentes conexas.



5.4.2 Visão computacional moderna

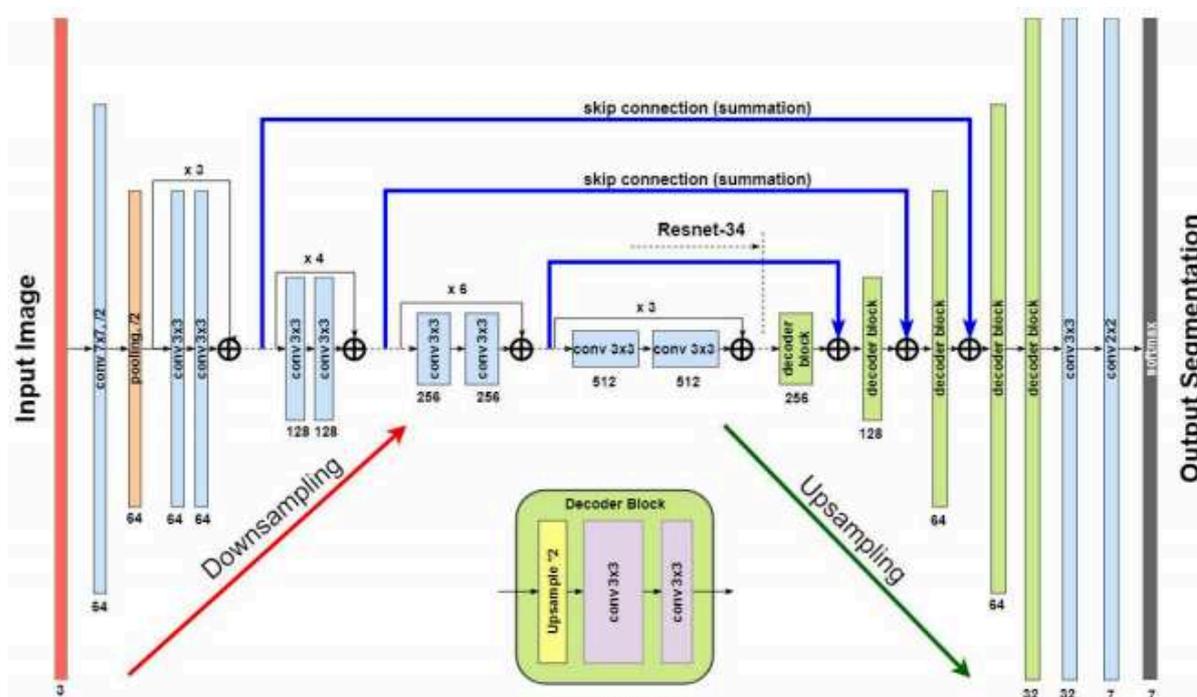
Devido aos resultados promissores encontrados na revisão do estado da arte que utilizaram DL como principal método, decidiu-se por empregar tal método para a parte de segmentação dos microfósseis. Previamente à aplicação de DL ao contexto de identificação de microfósseis, realizou-se uma revisão de métodos para identificação automática de micro-organismos. Nesta análise, foram encontrados alguns métodos de DL, os quais serviram como indicativo de uma possível solução para o problema apresentado neste trabalho.

Ao lidar com DL, mais especificamente aplicado a SS, pode-se simplificar o todo em 3 fatores principais que influenciam diretamente os resultados obtidos: o Modelo, a base de dados e a *loss function*. O modelo diz respeito a arquitetura da sua rede (profundidade e operações empregadas); A base de dados se refere a quantidade e a qualidade dos dados e, no caso de SS, a qualidade das marcações (GT); e a *loss function*, a qual é uma medida de quão bom um modelo de predição é em termos da habilidade de prever a saída esperada. Para o contexto de SS, um dos modelos mais empregado atualmente é a UNET (RONNEBERGER; FISCHER; BROX, 2015) e suas variações.

A arquitetura UNET foi apresentada em (RONNEBERGER; FISCHER; BROX, 2015), onde os autores mostram seu uso na segmentação de imagens médicas. As UNETs fornecem uma estrutura geral que pode ser parametrizada com um modelo específico de Redes Neurais Convolucionais (RNC) para classificação de imagens. A

UNET emprega duas instâncias levemente modificadas dessa RNC de classificação, uma codificadora e outra decodificadora uma para reconhecimento de imagem e outra, empregada no modo reverso, para a geração de máscara de segmento (BADRINARAYANAN; KENDALL; CIPOLLA, 2017). Desta forma, a UNET utiliza o codificador para mapear entradas brutas e representar características e o decodificador para tomar essas representações de características como entrada, processá-las para tomar sua decisão e produzir uma saída. Como o UNET produz uma SS de ponta, decidiu-se por escolher esta arquitetura como ponto de partida. Figura 26 mostra um exemplo da arquitetura da UNET.

Figura 26 – Exemplo da arquitetura UNET.



Fonte: (RAKHLIN; DAVYDOW; NIKOLENKO, 2018)

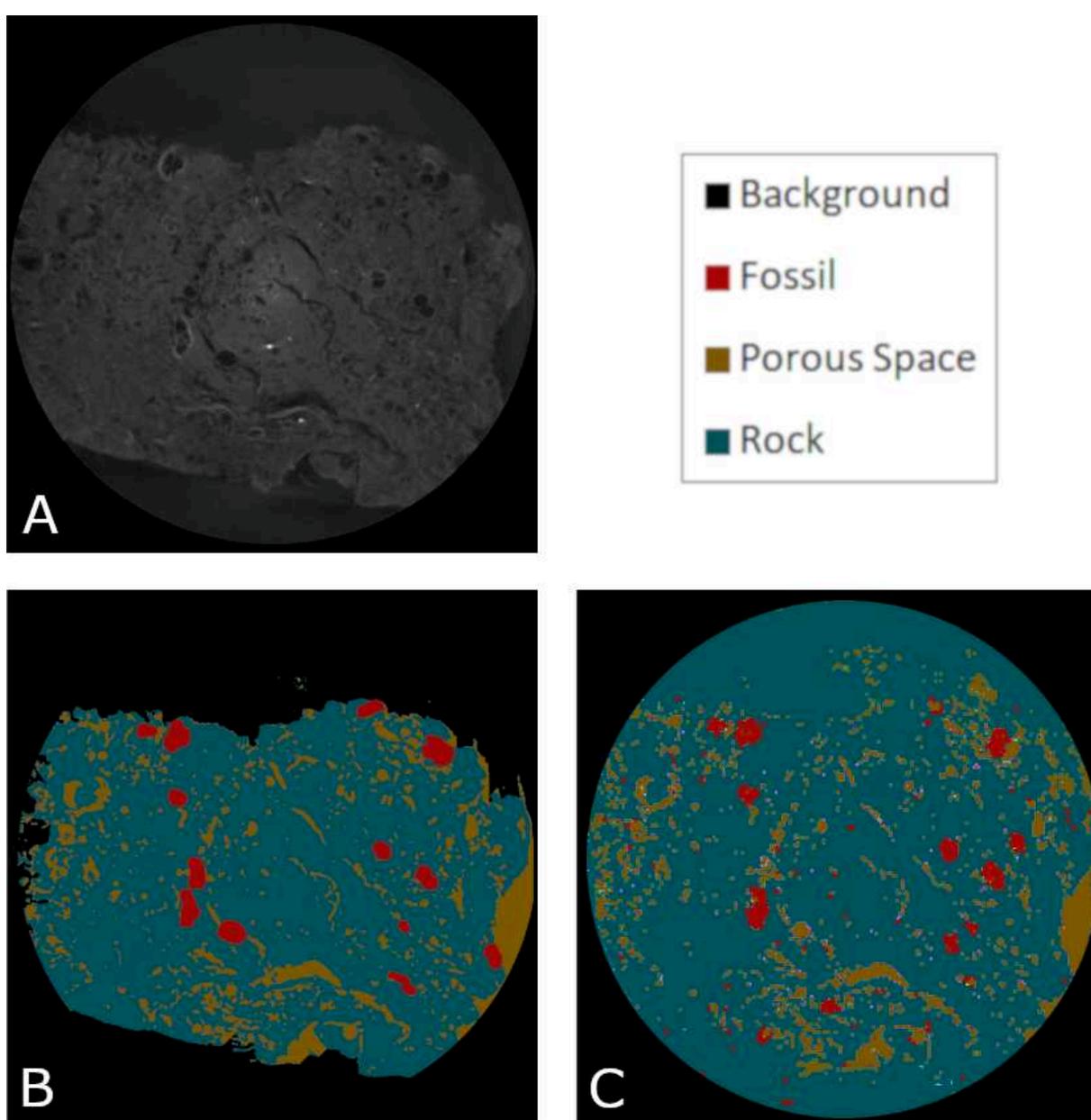
Desta forma, para os testes iniciais a seguinte estrutura foi montada: UNET associada com Resnet-34 (HE *et al.*, 2015), uma amostra de rocha com microfósseis (digitalizada com micro-CT totalizando 1000 fatias, as quais foram marcadas conforme a localização dos microfósseis, realizadas por especialistas da área de paleontologia) e a *binary cross entropy* como *loss function*. Essa estrutura está disponível no *framework* do *fastai*¹, que é uma estrutura sob Pytorch que contém vários modelos, métodos e últimos aprimoramentos do estado da arte.

Nesta estrutura inicial, testou-se primeiramente o uso das marcações apenas dos microfósseis, realizando uma separação binária entre microfóssil e não-microfóssil.

¹ <https://www.fast.ai/>

Com esta estrutura algumas estratégias como *data augmentation* e *transfer learning* foram aplicadas a fim de minimizar o efeito de possuir uma base de dados pequena. Entretanto, o coeficiente IOU, utilizado para avaliação dos resultados, ficou estagnado em 40–45%. Depois, realizou-se o aumento do número de classes, o que desmembrou a classe não-microfóssil em espaço poroso, rocha e plano de fundo. Nesta parte, os valores de IOU subiram e estagnaram em 75–76%. A figura 27 mostra o resultado obtido após o treinamento e a marcação realizada para a respectiva fatia.

Figura 27 – Resultado da segmentação de microfósseis obtido com a abordagem de 4 classes. (A) Imagem digitalizada original. (B) GT gerado manualmente pelos paleontólogos. (C) UNET + ResNet 34.



Os dados do Micro-CT apresentaram um desafio: os requisitos de memória

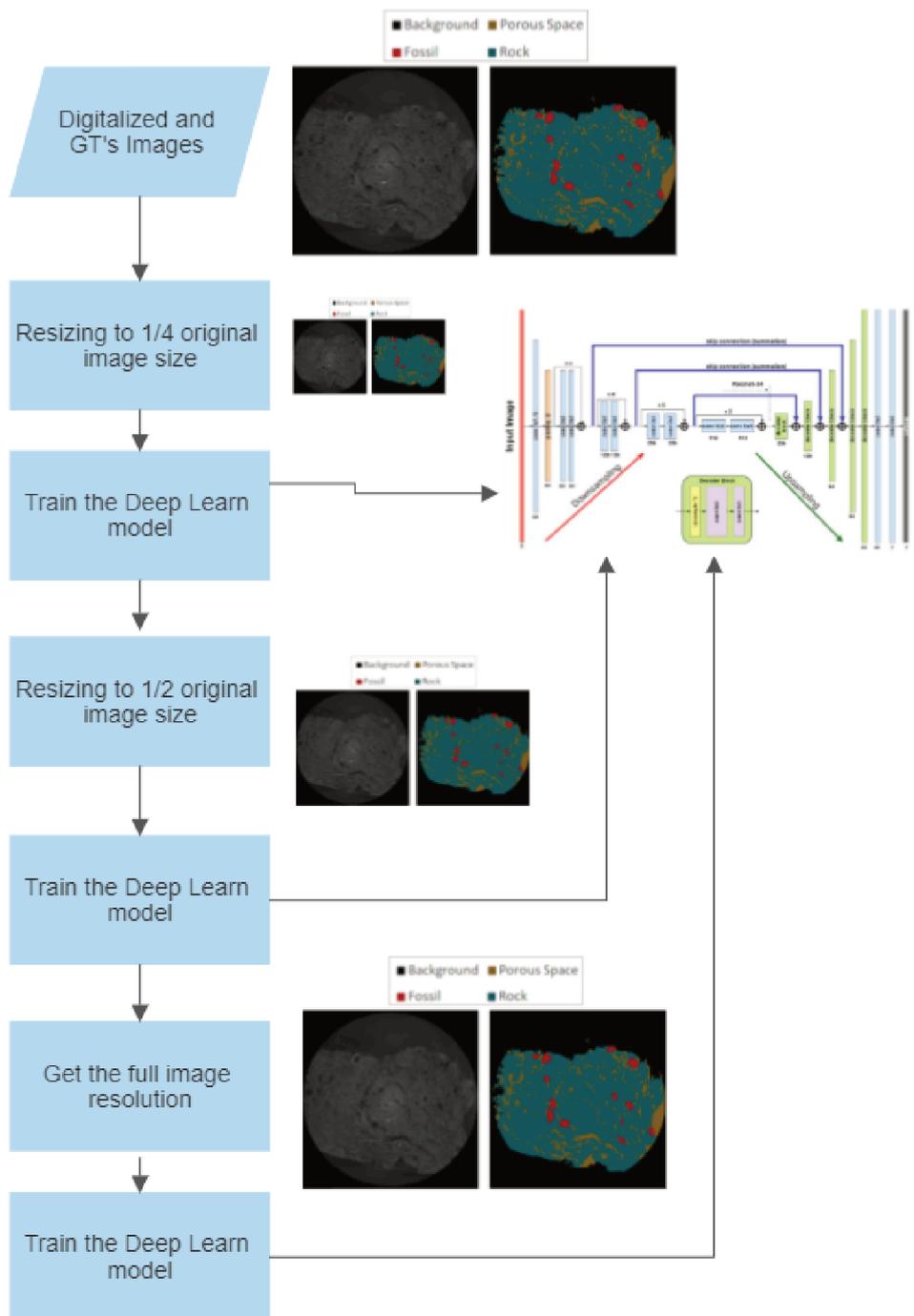
impostos pelo tamanho do conjunto de dados empregado e pela arquitetura UNET limitariam fortemente o *batch size* do conjunto de treinamento. Para superar essa limitação e poder trabalhar inicialmente com *batch sizes* maiores e treinar a rede em um ritmo mais rápido, empregou-se uma **abordagem de treinamento de resolução de imagem progressiva**. A Figura 28 fornece uma visualização desta abordagem de treinamento realizado em **três ciclos**: primeiro, inicia-se o *transfer learning* com o conjunto de dados em 1/4 da resolução original da imagem de MicroCT; depois, treina-se o modelo DL e redimensiona-se o conjunto de dados para 1/2 da resolução original; treina-se novamente e, por fim, emprega-se a resolução completa do volume do MicroCT para um ciclo final de treinamento para ajuste fino. O esboço dessa estratégia foi originalmente apresentado por Jeremy Howard como uma comunicação informal durante uma aula sobre RNC disponível em <https://course.fast.ai/videos/?lesson=1>).

Outra estratégia para ajustar um modelo é a estratégia de *Differential Learning Rates*, também apresentada informalmente por Jeremy Howard durante uma palestra da mesma série de cursos *fastai*. Essa abordagem para múltiplas taxas de aprendizado específicas para cada camada, à medida que as camadas na arquitetura empregada aprofundam-se, são justificadas pela seguinte lógica: Ao executar o *transfer learning* seguido de ajuste fino, nas primeiras camadas, o modelo pré-treinado sendo adaptado aprenderá características genéricas de baixo nível do novo conjunto de dados usado no *transfer learning*. Essas características de baixo nível são provavelmente muito semelhantes aos do conjunto de dados original, independentemente do contexto da imagem. Portanto, não há a necessidade de empregar altas taxas de aprendizado nessas primeiras camadas. No entanto, à medida que as informações aprofundam-se na arquitetura, as combinações de características tornam-se mais complexas e específicas ao conjunto de dados e são mais diretamente conectadas ao contexto de aplicação. Consequentemente, taxas mais altas de aprendizado em camadas mais profundas são desejáveis para permitir que a rede adapte-se melhor aos recursos específicos do contexto.

Desta forma, ainda utilizando a abordagem com 4 classes, os hiper-parâmetros foram ajustados, utilizando técnicas tais como as citadas acima e as presentes em (XIE *et al.*, 2018), e *data augmentation* e *batch size* foram melhor explorados obtendo o valor de 98 % IOU. A figura 29 mostra o GT de uma fatia 2D e o respectivo resultado utilizando o ajuste dos hiper-parâmetros.

Os experimentos realizados resultaram em um ambiente experimental onde empregou-se a UNET como modelo base (parte de codificação) associado com outros modelos na parte de decodificação (restnet18, ResNet34, ResNet50, ResNet101), *binary cross entropy* como *loss function* e o IOU para avaliação dos resultados. A Tabela 1 mostra os resultados de IOU para cada método empregado e a figura 30 mostra a imagem original, seu GT e os resultados de predição para todas as arquiteturas

Figura 28 – Abordagem de treinamento de resolução de imagem progressiva.



testadas.

Após realizar a segmentação, extraiu-se a máscara originária da predição da ResNet34 e aplicou a mesma sob a imagem original. O resultado deste processo é a identificação e extração de vários microfósseis, presentes na amostra sendo analisada, de forma automática e não-destrutiva. As Figuras 31, 32, 33 mostram os resultados de sobreposição das máscaras de segmentação, a identificação de uma amostra de microfóssil (destacada com o retângulo vermelho) seguida por sua versão ampliada e

Figura 29 – O GT (A) e o resultado obtido da segmentação de microfósseis (B) com 4 classes, ajuste dos hiper-parâmetros e melhor exploração de *data augmentation* e *batch size*.

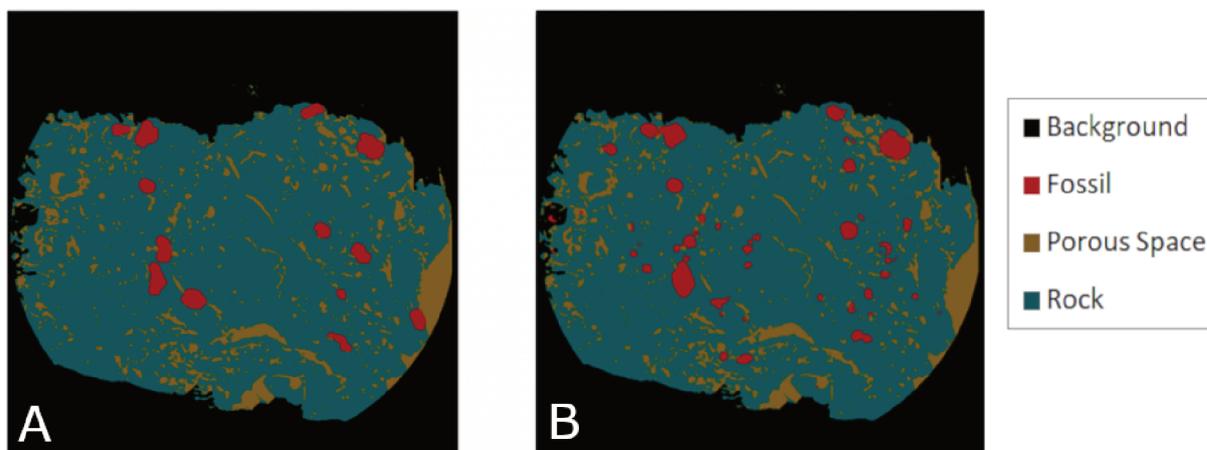


Tabela 1 – Desempenho da segmentação em termos de valor de IOU. Cada método foi avaliado em um conjunto de 1000 imagens contendo marcações dos dados de microfósseis.

Método	Valor de IOU
Contornos Ativos	0.20
UNET + ResNet34	0.76
UNET + ResNet18 + otimização dos hiper-parâmetros	0.97
UNET + ResNet101 + otimização dos hiper-parâmetros	0.97
UNET + ResNet34 + otimização dos hiper-parâmetros	0.98
UNET + ResNet50 + otimização dos hiper-parâmetros	0.98

a correlação dessa versão ampliada com as outras duas versões da mesma amostra (isoladas fisicamente e digitalizadas com o Versa XRM-500 MicroCT e o Zeiss Discovery Estereoscópio V20). O código utilizado para os experimentos e a base de dados de microfósseis com marcação dos especialistas está publicamente disponível em: <http://www.lapix.ufsc.br/microfossil-segmentation/>

Figura 30 – (A) Fatia 2D da amostra original digitalizada. (B) GT referente a fatia 2D. (C) Resultado da predição UNET + ResNet18 + otimização dos hiper-parâmetros. (D) Resultado da predição UNET + ResNet101 + otimização dos hiper-parâmetros. (E) Resultado da predição UNET + ResNet34 + otimização dos hiper-parâmetros. (F) UNET + ResNet50 + otimização dos hiper-parâmetros.

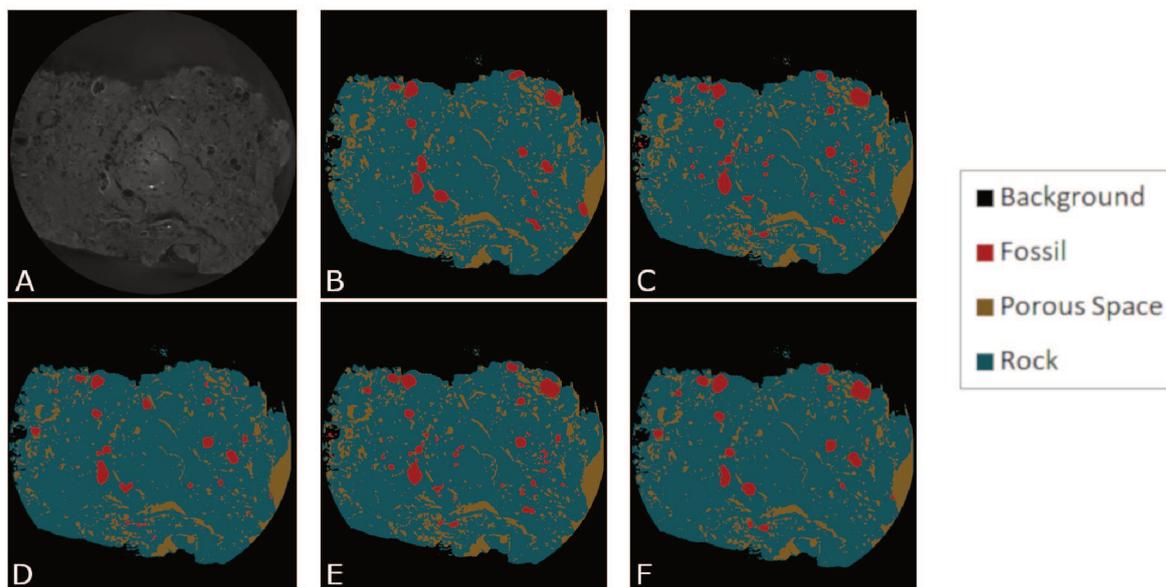


Figura 31 – Resultado da aplicação da máscara de segmentação obtida sobre a imagem digitalizada. (A) imagem da fatia 2D, com contraste ajustado, após a aplicação da máscara resultante. Destacado por um retângulo vermelho encontra-se um microfóssil específico. (B) Microfóssil destacado extraído e ampliado para visualização. (C) Microfóssil isolado fisicamente, digitalizado com o Versa XRM-500 MicroCT. (D) Aquisição multidimensional do microfóssil cibicoides com o estereoscópio Zeiss Discovery V20.

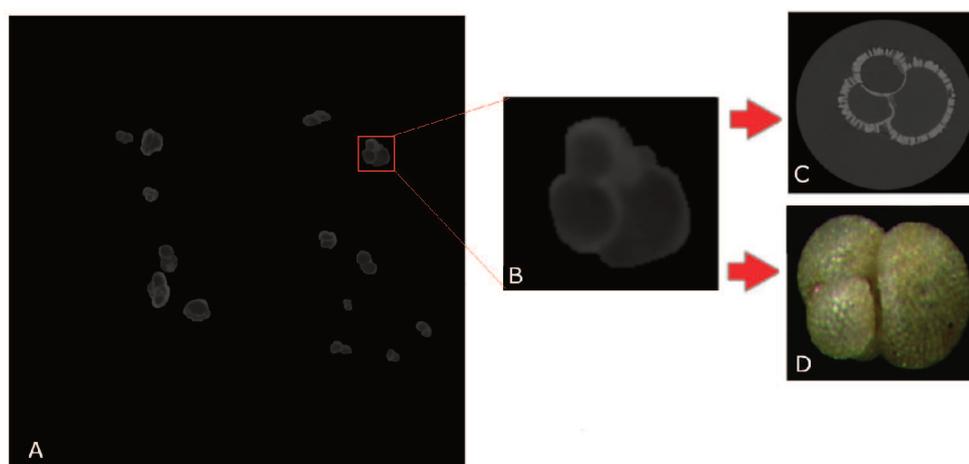


Figura 32 – Resultado da aplicação da máscara de segmentação obtida sobre a imagem digitalizada. (A) imagem da fatia 2D, com contraste ajustado, após a aplicação da máscara resultante. Em vermelho encontra-se um microfóssil específico destacado. (B) Microfóssil destacado extraído e ampliado para visualização. (C) Microfóssil isolado fisicamente, digitalizado com o Versa XRM-500 MicroCT. (D) Aquisição multidimensional do microfóssil cibicidoides com o estereoscópio Zeiss Discovery V20.

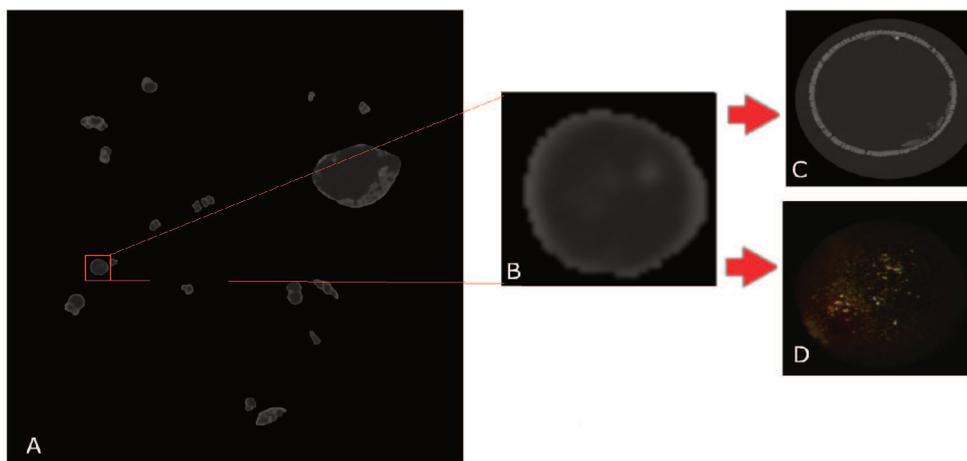
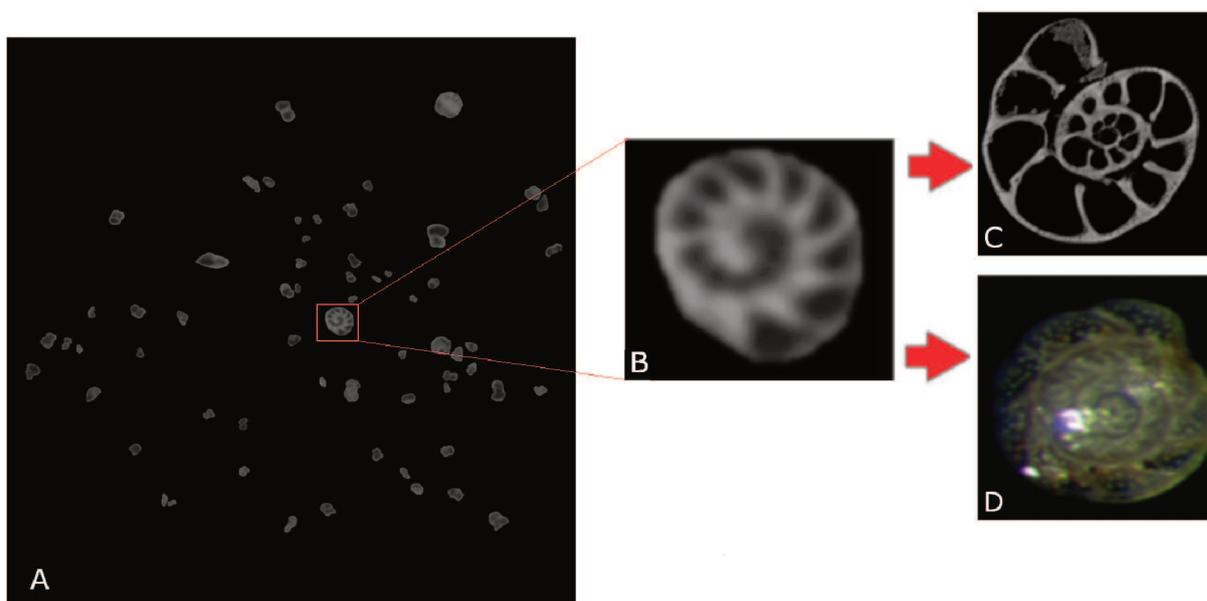


Figura 33 – Resultado da aplicação da máscara de segmentação obtida sobre a imagem digitalizada. (A) imagem da fatia 2D, com contraste ajustado, após a aplicação da máscara resultante. Em vermelho encontra-se um microfóssil específico destacado. (B) Microfóssil destacado extraído e ampliado para visualização. (C) Microfóssil isolado fisicamente, digitalizado com o Versa XRM-500 MicroCT. (D) Aquisição multidimensional do microfóssil cibicidoides com o estereoscópio Zeiss Discovery V20.



6 CONCLUSÃO

Neste trabalho foi apresentado um *pipeline* não-destrutivo de processamento de imagens para identificação de microfósseis em rochas carbonáticas que permite a identificação dos microfósseis de forma automática sem necessidade de executar um isolamento físico prévio. Adicionalmente, os métodos de visão computacional empregados foram desenvolvidos e avaliados. A avaliação foi realizada quantitativamente e automaticamente contra um padrão ouro (GT) gerado manualmente por paleontólogos. O IOU utilizado como valor de comparação entre os padrões gerados pelos especialistas apresentou uma taxa de 98 % de sobreposição dos resultados.

Um aspecto de suma relevância para a utilização deste método na paleontologia, mais especificadamente micropaleontologia, está no caráter não-destrutivo das amostras. No processo de estudo micropaleontológico uma etapa imprescindível é a preparação das amostras, objetivando separar os microfósseis dos demais constituintes da rocha e/ou sedimento. Neste processo realizado em laboratório, as amostras são desagregadas fisicamente, trituradas ou moídas, e posteriormente é feita a desagregação química, com adição de reagentes (e.g. peróxido de hidrogênio e ácido acético). Tanto a desagregação física quanto química podem alterar ou até mesmo destruir características dos microfósseis a serem analisados. Nesta premissa, o método do imageamento é crucial para a visualização das características morfológicas as mais fidedignas possíveis, permitindo o reconhecimento taxonômico dos indivíduos (KACHOVICH; SHENG; AITCHISON, 2019).

Além do aspecto citado acima, outro fator relevante para a exequibilidade em estudos com microfósseis está na abordagem tafonômica, que representa a análise da preservação dos microrganismos ao longo do tempo geológico, bem como aspectos de fossilização, preservação, e até mesmo posição na qual os microfósseis estão depositados (preservados) nas rochas. Ressalta-se que estudos com a abordagem tafonômica são fundamentais para reconstituições de condições paleoambientais e/ou processos de alteração diagenética ao longo do tempo geológico.

A exequibilidade desse método, além de ser uma ferramenta científica, é diretamente aplicável à prospecção do petróleo. O uso desta ferramenta é fortemente indicado em casos em que é extremamente difícil a recuperação de microfósseis ao longo de seções e/ou intervalos específicos onde o material (rocha) é muito compacto e até mesmo, quando apresenta incipiente alteração diagenética.

Algumas observações podem ser realizadas sobre os resultados obtidos: (i) a importância de empregar hiperparâmetros apropriados, como taxa de aprendizado, *weight decay*, *momentum* e *batch size*: com essa otimização de hiperparâmetros, obteve-se uma melhoria de mais de 20 % no IOU. (ii) o aumento em uma arquitetura de rede desenvolvida não implica em melhores resultados: É possível observar que

a ResNet34 mostra os mesmos resultados que a ResNet50 e um resultado melhor quando comparado com a ResNet101. No entanto, aqui se tem uma limitação de hardware: a ResNet50 e a ResNet101 não podiam ser executados com a resolução de imagem completa na GeForce GTX 1080 Ti de 11 GB disponível, mesmo com o *batch size* de 1. Mesmo assim, a ResNet34 requer menos tempo de execução e hardware. (iii) Outra limitação imposta pelo hardware associada a natureza 3D da imagem, foi a não possibilidade de fazer uma análise utilizando uma arquitetura que levasse em conta a parte 3D. A segmentação acabou sendo realizada de forma 2D fatia por fatia, com a rede aprendendo os conceitos de posicionamento 3D com base nesta forma de amostragem fracionada. (iv) Analisando as imagens de resultados obtidas e comparando visualmente com GT (Figura 30), ainda foram observados alguns pequenos erros. No entanto, entende-se que isso pode ser atenuado pela adição de mais amostras de treinamento, juntamente com os GTs de especialistas, para o conjunto de treinamento ao aplicar essa rede pré-treinada a outras amostras. Além disso, sempre há novas melhorias no estado da arte que podem ser testadas visando reduzir ainda mais esses pequenos erros. A Figura 31 mostra o microfóssil isolado digitalizado e sua identificação correlacionada na amostra.

Em suma, utilizar o *pipeline* proposto, sem a necessidade de isolar fisicamente o microfóssil, tem o potencial de permitir que os paleontólogos analisem aspectos específicos da amostra, como, por exemplo, a forma que o microfóssil se depositou, importante para algumas aplicações na indústria de óleo e gás. Adicionalmente, o *pipeline* tem o potencial de melhorar a forma de trabalho do paleontologista, dado que ao em vez de gastar tempo para isolar fisicamente o microfóssil ele recebe o microfóssil extraído e pode executar outras análises, como por exemplo identificação de classe e orientação de deposição dos espécimes.

6.1 CONTRIBUIÇÕES

O desenvolvimento do *pipeline* com métodos automáticos, ou semiautomáticos, visando automatizar o processo de análise de microfósseis em amostras de rochas, tem o intuito de também realizar a análise de um grande volume de amostras sem a necessidade de um grande esforço manual e exaustivo, como é realizado atualmente. Logo, pode-se pontuar como principais contribuições deste trabalho os seguintes resultados:

- O método para identificar e extrair microfósseis de uma rocha de forma automática.
- Análise de grande volume de amostras com uma iteração mínima entre usuário e método.

- Construção de uma base de dados para experimentos voltados para microfósseis.
- Construção de uma *baseline* para métodos de identificação e extração de microfósseis em rochas.
- Elaboração de métodos para o auxílio da análise Tafonômica de microfósseis.
- Desenvolvimento de uma metodologia para trabalhar com microfósseis (*pipeline* e avaliação).

6.1.1 Publicações

No decorrer do desenvolvimento deste trabalho as seguintes publicações foram realizadas:

- 3D Segmentation Algorithms for Computerized Tomographic Imaging: a Systematic Literature Review (CARVALHO; SOBIERANSKI; WANGENHEIM, 2018).
- 3D object recognition and classification: a systematic literature review (CARVALHO; WANGENHEIM, 2019);
- Automated Microfossil Identification and Segmentation Using a Deep Learning Approach (CARVALHO; FAUTH *et al.*, 2019).

6.2 TRABALHOS FUTUROS

Adicionalmente às contribuições anteriormente mencionadas, esta pesquisa abre espaço para alguns trabalhos futuros, tais como:

- Ampliação da aplicabilidade dos métodos desenvolvidos neste trabalho para análise de fósseis de escalas maiores ou menores, fósseis e nanofósseis respectivamente;
- Análise Tafonômica digital de microfósseis em amostras de rochas;
- Melhoria da arquitetura empregada bem como adição de outros métodos de melhoramento de performance buscando diminuir ainda mais o erro de identificação dos microfósseis nas rochas.

REFERÊNCIAS

- BADRINARAYANAN, Vijay; KENDALL, Alex; CIPOLLA, Roberto. Segnet: A deep convolutional encoder-decoder architecture for image segmentation. **IEEE transactions on pattern analysis and machine intelligence**, IEEE, v. 39, n. 12, p. 2481–2495, 2017.
- BARBARIN, Nicolas. **La reconnaissance automatisée des nannofossiles calcaires du Cénozoïque**. 2014. Tese (Doutorado) – Aix-Marseille.
- BEAUFORT, L; DOLLFUS, D. Automatic recognition of coccoliths by dynamical neural networks. **Marine Micropaleontology**, Elsevier, v. 51, n. 1-2, p. 57–73, 2004.
- BEN-HUR, Asa; ELISSEEFF, Andre; GUYON, Isabelle. A stability based method for discovering structure in clustered data. **Pacific Symposium on Biocomputing**. **Pacific Symposium on Biocomputing**, BioWulf Technologies LLC, 2030 Addison st. Suite 102, Berkeley, CA 94704, USA., p. 6–17, 2002.
- BOAS, F Edward; FLEISCHMANN, Dominik. CT artifacts: causes and reduction techniques. **Imaging in medicine**, Open Access Journals, v. 4, n. 2, p. 229–240, 2012.
- BROWN, Lisa Gottesfeld. A Survey of Image Registration Techniques. **ACM Comput. Surv.**, Association for Computing Machinery, New York, NY, USA, v. 24, n. 4, p. 325–376, dez. 1992. ISSN 0360-0300. DOI: 10.1145/146370.146374. Disponível em: <https://doi.org/10.1145/146370.146374>.
- BUENO, Gloria *et al.* Automated Diatom Classification (Part A): Handcrafted Feature Approaches. **Applied Sciences**, v. 7, p. 753, jul. 2017. DOI: 10.3390/app7080753.
- CARVALHO, L.E; FAUTH, G. *et al.* Automated Microfossil Identification and Segmentation Using a Deep Learning Approach. **bioRxiv**, Cold Spring Harbor Laboratory, 2019. DOI: 10.1101/661694. eprint: <https://www.biorxiv.org/content/early/2019/06/06/661694.full.pdf>. Disponível em: <https://www.biorxiv.org/content/early/2019/06/06/661694>.
- CARVALHO, LE; SOBIERANSKI, Antonio Carlos; WANGENHEIM, Aldo von. 3D segmentation algorithms for computerized tomographic imaging: a systematic literature review. **Journal of digital imaging**, Springer, v. 31, n. 6, p. 799–850, 2018.
- CARVALHO, Luis Eduardo Ramos de; WANGENHEIM, Aldo von. 3D object recognition and classification: a systematic literature review. **Pattern Anal. Appl.**, v. 22, n. 4, p. 1243–1292, 2019. DOI: 10.1007/s10044-019-00804-4. Disponível em: <https://doi.org/10.1007/s10044-019-00804-4>.
- CHARLES, J. J. Automatic recognition of complete palynomorphs in digital images. **Machine Vision and Applications**, v. 22, n. 1, p. 53–60, jan. 2011. ISSN 1432-1769.

DOI: 10.1007/s00138-009-0200-4. Disponível em:
<https://doi.org/10.1007/s00138-009-0200-4>.

CULVERHOUSE, Phil F *et al.* Automatic classification of field-collected dinoflagellates by artificial neural network. **Marine Ecology Progress Series**, v. 139, p. 281–287, 1996.

DONGEN, Stijn Van. **Performance Criteria for Graph Clustering and Markov Cluster Experiments**. [S.l.], 2000.

DUMOULIN, Vincent; VISIN, Francesco. A guide to convolution arithmetic for deep learning, 2016. arXiv: 1603.07285 [stat.ML].

FOWLKES, E. B.; MALLOWS, C. L. A Method for Comparing Two Hierarchical Clusterings. **Journal of the American Statistical Association**, American Statistical Association, v. 78, n. 383, p. 553–569, 1983.

FRANCUS, P. **Image Analysis, Sediments and Paleoenvironments**. [S.l.]: Springer Netherlands, 2007. (Developments in Paleoenvironmental Research). ISBN 9781402020612. Disponível em:
<https://books.google.com.br/books?id=SRGDk2Gbj4UC>.

GOODFELLOW, Ian; BENGIO, Yoshua; COURVILLE, Aaron. **Deep Learning**. [S.l.]: MIT Press, 2016. <http://www.deeplearningbook.org>.

HE, Kaiming *et al.* Deep Residual Learning for Image Recognition. **CoRR**, abs/1512.03385, 2015. eprint: 1512.03385.

HIBBETT, David. Automated Taxon Identification in Systematics: Theory, Approaches and Applications. **The Quarterly Review of Biology**, v. 84, n. 3, p. 295–296, 2009. DOI: 10.1086/644681. eprint: <https://doi.org/10.1086/644681>. Disponível em:
<https://doi.org/10.1086/644681>.

HSIANG, Allison Y. *et al.* Endless Forams: >34,000 Modern Planktonic Foraminiferal Images for Taxonomic Training and Automated Species Recognition Using Convolutional Neural Networks. English. **Paleoceanography and Paleoclimatology**, John Wiley & Sons, Ltd, v. 34, n. 7, p. 1157–1177, jul. 2019. ISSN 2572-4517. DOI: 10.1029/2019PA003612.

KACHOVICH, Sarah; SHENG, Jiani; AITCHISON, Jonathan C. Adding a new dimension to investigations of early radiolarian evolution. **Scientific Reports**, v. 9, 1 2019. DOI: 10.1038/s41598-019-42771-0.

KASS, Michael; WITKIN, Andrew; TERZOPOULOS, Demetri. Snakes: Active contour models. **International Journal of Computer Vision**, v. 1, n. 4, p. 321–331, jan. 1988.

ISSN 1573-1405. DOI: 10.1007/BF00133570. Disponível em:
<https://doi.org/10.1007/BF00133570>.

KITCHENHAM, Barbara. **Procedures for Performing Systematic Reviews**. [S.l.], 2004.

LECUN, Yann *et al.* Gradient-based learning applied to document recognition. *In*: PROCEEDINGS of the IEEE. [S.l.: s.n.], 1998. p. 2278–2324.

LIU, Shihming; THONNAT, Monique; BERTHOD, Marc. Automatic classification of planktonic foraminifera by a knowledge-based system. **Proceedings of the Tenth Conference on Artificial Intelligence for Applications**, p. 358–364, 1994.

LORENSEN, William E.; CLINE, Harvey E. Marching Cubes: A High Resolution 3D Surface Construction Algorithm. **SIGGRAPH Comput. Graph.**, ACM, New York, NY, USA, v. 21, n. 4, p. 163–169, ago. 1987. ISSN 0097-8930. DOI: 10.1145/37402.37422. Disponível em: <http://doi.acm.org/10.1145/37402.37422>.

MARCHANT, Ross *et al.* Classification of down-core foraminifera image sets using convolutional neural networks. **bioRxiv**, Cold Spring Harbor Laboratory, 2019. DOI: 10.1101/840926. eprint:
<https://www.biorxiv.org/content/early/2019/11/13/840926.full.pdf>.

MARTIN, R.E. **Taphonomy: A Process Approach**. [S.l.]: Cambridge University Press, 1999. (Cambridge Paleobiology Series). ISBN 9781316582671. Disponível em:
<https://books.google.com.br/books?id=dcjLCgAAQBAJ>.

MITRA, R. *et al.* Automated species-level identification of planktic foraminifera using convolutional neural networks, with comparison to human performance. **Marine Micropaleontology**, v. 147, p. 16–24, 2019. ISSN 0377-8398. DOI:
<https://doi.org/10.1016/j.marmicro.2019.01.005>. Disponível em:
<http://www.sciencedirect.com/science/article/pii/S0377839818301105>.

MOLINA, Eustoquio; ED. **Micropaleontologia (2ª edición)**. [S.l.]: Prensas Universitarias de Zaragoza. Colección Textos Docentes, 2004.

NAGENDAR, Gattigorla *et al.* Neuro-IoU: Learning a Surrogate Loss for Semantic Segmentation. *In*: BMVC. [S.l.: s.n.], 2018. p. 278.

O'MAHONY, Niall *et al.* Deep Learning vs. Traditional Computer Vision. *In*: SPRINGER. SCIENCE and Information Conference. [S.l.: s.n.], 2019. p. 128–144.

OTSU, N. A Threshold Selection Method from Gray-Level Histograms. **IEEE Transactions on Systems, Man, and Cybernetics**, v. 9, n. 1, p. 62–66, jan. 1979. ISSN 0018-9472. DOI: 10.1109/TSMC.1979.4310076.

PENSAMENTO, Livre. **O método hipotético-dedutivo**. [S.l.: s.n.], Acessado em: 17/06/2016. Disponível em:

<https://livrepensamento.com/2013/10/01/o-metodo-hipotetico-dedutivo/>.

PONTI, Moacir Antonelli; COSTA, Gabriel B Paranhos da. Como funciona o deep learning. **arXiv preprint arXiv:1806.07908**, 2018.

POPPER, Karl. **The Logic of Scientific Discovery (Routledge Classics)**. 2. ed. [S.l.]: Routledge, 2002. p. 544. ISBN 8130908115.

RAHMAN, Md Atiqur; WANG, Yang. Optimizing intersection-over-union in deep neural networks for image segmentation. *In*: SPRINGER. INTERNATIONAL symposium on visual computing. [S.l.: s.n.], 2016. p. 234–244.

RAKHLIN, Alexander; DAVYDOW, Alex; NIKOLENKO, Sergey. Land Cover Classification from Satellite Imagery with U-Net and Lovász-Softmax Loss. *In*: IEEE. 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). [S.l.: s.n.], 2018. p. 257–2574.

RAND, William M. Objective Criteria for the Evaluation of Clustering Methods. **Journal of the American Statistical Association**, American Statistical Association, v. 66, n. 336, p. 846–850, 1971.

REZATOFIGHI, Hamid *et al.* Generalized intersection over union: A metric and a loss for bounding box regression. *In*: PROCEEDINGS of the IEEE Conference on Computer Vision and Pattern Recognition. [S.l.: s.n.], 2019. p. 658–666.

RONNEBERGER, Olaf; FISCHER, Philipp; BROX, Thomas. U-Net: Convolutional Networks for Biomedical Image Segmentation. **CoRR**, abs/1505.04597, 2015. eprint: 1505.04597.

SCHULZE, Katja *et al.* PlanktoVision - an automated analysis system for the identification of phytoplankton. **BMC Bioinformatics**, v. 14, n. 1, p. 115, mar. 2013. ISSN 1471-2105. DOI: 10.1186/1471-2105-14-115. Disponível em: <https://doi.org/10.1186/1471-2105-14-115>.

SIJBERS, J *et al.* Watershed-based segmentation of 3D MR data for volume quantization. **Magnetic Resonance Imaging**, Elsevier, v. 15, n. 6, p. 679–688, 1997.

SINGH, Asheesh. Micropaleontology in petroleum exploration. *In*: 7TH International Conference and Exposition of Petroleum Geophysics. [S.l.: s.n.], 2008. p. 14–16.

SØRENSEN, T. A method of establishing groups of equal amplitude in plant sociology based on similarity of species and its application to analyses of the vegetation on Danish commons. **Biol. Skr.**, v. 5, p. 1–34, 1948.

SOUZA CARVALHO, Ismar de. **Paleontologia**. [S.l.]: Interciência, 2004. v. 1. ISBN 85-7193-107-0.

SUTTON, M.D.; RAHMAN, I.; GARWOOD, R.J. **Techniques for Virtual Palaeontology**. 1. ed. [S.l.]: Wiley, 2014. (New Analytical Methods in Earth and Environmental Science). ISBN 9781118591130.

TIPPER, John C. Computer applications in paleontology: Balance in the late 1980s? **Computers & Geosciences**, v. 17, n. 8, p. 1091–1098, 1991. ISSN 0098-3004.

WU, Yi; LIM, Jongwoo; YANG, Ming-Hsuan. Online object tracking: A benchmark. *In: PROCEEDINGS of the IEEE conference on computer vision and pattern recognition*. [S.l.: s.n.], 2013. p. 2411–2418.

XIE, Junyuan *et al.* Bag of tricks for image classification with convolutional neural networks. **arXiv preprint arXiv:1812.01187**, 2018.

YU, S. *et al.* Feasibility study of automatic identification of planktic foraminifera by computer vision. **Journal of Foraminiferal Research**, v. 26, n. 2, p. 113–123, abr. 1996. ISSN 0096-1191. DOI: 10.2113/gsjfr.26.2.113. eprint: <https://pubs.geoscienceworld.org/jfr/article-pdf/26/2/113/3022161/113.pdf>. Disponível em: <https://doi.org/10.2113/gsjfr.26.2.113>.

**ANEXO A – REVISÃO DA LITERATURA SOBRE MÉTODOS DE VISÃO
COMPUTACIONAL PARA ANÁLISE DE IMAGEM DE FÓSSEIS**

Revisão da literatura sobre métodos de visão computacional para análise de imagem de fósseis

Carvalho, L.E. e von Wangenheim, A

I. INTRODUÇÃO

A área da paleontologia é um campo de muito estudo que tem como base um passado cheio de informações. Com base na análise do passado foram feitas as primeiras hipóteses de como surgiu a vida, como se deu a evolução humana e as principais transformações do mundo. A paleontologia é uma ciência antiga, existem relatos da época antes de cristo e seu entrelaçamento com a área da computação se deu com maior afinco a partir da década de 90.

Em um *review* elaborado por Tipper [1], onde uma análise dos estudos paleontológicos utilizando aplicações computacionais no fim da década de 80 é apresentada, o autor fez uma pesquisa nas principais revistas de paleontologia da época e percebeu que em algumas destas existia, de forma escassa, um artigo que descrevia alguma aplicação de métodos computacionais, em outras existiam poucos e em algumas uma verdadeira enxurrada de artigos. Outros jornais e livros, gerais bem como explicitamente paleontológicos, também relevantes, contavam a mesma história: a inserção de métodos computacionais na paleontologia parece, na melhor das hipóteses, desigual.

Com a industrialização dos tomógrafos médicos houve um incentivo na pesquisa utilizando imagens tomográficas aplicadas a paleontologia, o que impulsionou o desenvolvimento de softwares voltados à análise das mesmas. Alguns exemplos são: DRISHTI, SPIERS, SEG3D, IMAGEJ, MIMICS, VGSTUDIO MAX, AVIZO, AMIRA, Geomagic, Rhinoceros, Imaris, ITK-SNAP e TurtleSEG.

Com a popularização dos tomógrafos e o desenvolvimento de ferramentas de análise, a paleontologia, que antes considerava o uso dos métodos computacionais por muitas vezes desnecessária, obteve uma ferramenta indispensável para suas análises, seja utilizando métodos de análise de imagens ou utilizando simulações de propriedades físicas, por exemplo. A aplicabilidade da então intitulada paleontologia virtual, vai desde as análises da evolução de animais, plantas e microrganismos até a criação de simulações sobre o habitat dos seres de determinada época, podendo ser aplicada nos mais diversos nichos, como por exemplo na parte de exploração de petróleo.

Com foco na análise dos métodos computacionais para análise de imagem tomográficas de fósseis, este documento tem como propósito detalhar todos os procedimentos relacionados para a revisão sistemática da literatura, para o tópico em questão, para o período de 2006-2016 (julho), e realizar uma análise do estado da arte, tendo em vista o entendimento sobre os métodos computacionais geralmente empregados para análise de imagem de fósseis.

Este relatório técnico está dividido nas seguintes seções. A seção de metodologia, define quais foram os passos executados para a realização deste trabalho. A seção de descrição da pesquisa, detalha os parâmetros de pesquisa utilizados bem como critérios de seleção empregados na seleção dos trabalhos a serem analisados. A seção de trabalhos analisados, sumariza e evidencia os principais aspectos dos trabalhos selecionados para análise, com foco nas técnicas de Visão Computacional (VC) utilizadas. Por fim, a análise, discussão e conclusão sobre os trabalhos selecionados é realizada na seção de discussão e conclusão.

II. METODOLOGIA

Este trabalho teve como base a metodologia proposta por Kitchenham[2] para conduzir o processo de pesquisa. Esta metodologia, para análise sistemática da literatura, tem como objetivo, como descrito pela própria autora, avaliar e interpretar todas as pesquisas relevantes disponíveis à uma particular pergunta de pesquisa, tópico ou fenômeno de interesse. Três guias utilizados para pesquisa na área da saúde serviram como base para a metodologia de Kitchenham [3] [4] [5] [6]. A principal característica ou propriedade desta metodologia é a de manter a procura reproduzível, quando todos os passos descritos forem repetidos e as mesmas palavras-chave forem utilizadas.

Para a análise dos trabalhos resultantes da aplicação da metodologia de Kitchenham, foram utilizados como referência os métodos de VC empregados.

III. DESCRIÇÃO DA PESQUISA

Inicialmente, com base em Kitchenham, definiu-se a tópico de pesquisa, **métodos computacionais aplicado a imagens tomográficas de fósseis**. Com base neste tópico, definiram-se as palavras-chave e utilizou-se três bases de artigos científicos (ScienceDirect, ACM e IEEE) para a pesquisa. As palavras-chave utilizadas em cada base e os filtros de pesquisa foram: **ScienceDirect:** *pub-date > 2005 and pub-date < 2017 and TITLE-ABSTR-KEY(fossil) and TITLE-ABSTR-KEY(tomography)*.

ACM: *(+ "fossil" tomography) e Published years: 2006 - 2016*).

IEEE: *((fossil) AND tomography) e Year: 2006 - 2016*.

Obteve-se um total de 250 artigos, os quais foram analisados os *abstracts*. Na leitura dos *abstracts*, utilizaram-se os seguintes **critérios de exclusão**, afim de selecionar os trabalhos a serem analisados na íntegra:

- Artigos escritos em outras línguas diferentes da inglesa;

- Artigos repetidos e que não fazem referência ao estudo de fósseis com base na análise de imagem.
- Artigos que não possuem uma descrição de etapas computacionais realizadas (aplicação de métodos de análise de imagem, filtros, segmentações, entre outros).
- O trabalho relatado, na parte computacional, se resume apenas ao uso de um software para visualização e medição do fóssil.

Com a aplicação dos critérios de exclusão, obteve-se um total de 18 artigos, os quais foram lidos na íntegra, sumarizados e analisados.

IV. TRABALHOS ANALISADOS

A. *A virtual world of paleontology*

O artigo proposto por Cunningham et. al [7] é um *survey*, o qual fornece um panorama sobre o estudo de fósseis, como é feita a extração do fóssil da rocha e sua posterior caracterização em 3D. Inicialmente, o autor descreve um breve histórico sobre a evolução das técnicas de análise e reconstrução 3D de fósseis, utilizando gesso entre outros, expondo os problemas destas técnicas (tempo de reconstrução, destruição do fóssil e processo totalmente laboratorial, por exemplo). No final deste panorama, o autor chega na parte de utilização de Tomografia Computadorizada (TC), a qual entra na parte de caracterização não-destrutiva de fósseis, uma tabela com outras técnicas não destrutivas de caracterização é mostrada no artigo). Inicialmente, dado o principal foco que era a área médica, a análise tomográfica de fósseis era voltada para fósseis macroscópicos devido a baixa resolução e ao pouco poder de investigação em materiais muito densos. Com o tempo e o desenvolvimento dos Micro-TC e nano-TC (alta energia e resolução) possibilitou-se a análise de um número maior de fósseis. Algumas características especiais dos outros métodos de análise não destrutiva (técnicas para análise da superfície dos fósseis ou de tecidos biológicos, por exemplo) são brevemente citadas.

Por fim, o autor comenta sobre algumas ferramentas de análise digital que foram desenvolvidas impulsionadas pela pesquisa na área, alguns exemplos são destas ferramentas são: Amira e Avizo (ferramentas pagas) e Drishti (ferramenta de código livre).

B. *Combined methodologies for three-dimensional reconstruction of fossil plants preserved in siderite nodules: *Stephanospermum braidwoodensis* nov. sp. (Medullosales) from the Mazon Creek lagerstätte*

Com o estudo sobre a identificação de uma nova espécie de óvulo de *Medullosan* encontrado em Mazon Creek Carboniferous lagerstätte, o autor [8] mostra uma combinação de métodos para efetivar o trabalho de identificação. Para tal identificação utilizou-se uma combinação de métodos de digitalização, Microtomografia Computadorizada (Micro-CT) e em um preciso seccionamento orientado pelos resultados do Micro-CT. A reconstrução 3D do óvulo correlacionou geometrias de diferentes camadas com características de tecidos coletadas das seções oriundas do processo de seccionamento,

permitindo o posicionamento de cortes seriais em posições pré-determinadas do holótipo.

Para realizar o processo de separação do fóssil da matriz, realizou-se um processo de segmentação manual, ou seja, para cada fatia era selecionado manualmente um valor de intensidade que pertencia a um determinado segmento. Uma limpeza manual das fatias também foi realizado para remover artefatos e ruídos. Estrutura anatômicas distintas foram atribuídas para máscaras individuais afim de serem renderizadas como *isosurfaces* separadas, um processo iterativo de edição e renderização foi realizado com a finalidade de produzir um modelo preciso com cores falsas e mostrar a inter-relação entre características anatômicas distintas dentro do espaço tridimensional.

Após esse processo de estudo, o fóssil passou por uma reconstrução física e foi novamente digitalizado e as imagens resultantes foram cortadas, rotacionadas, tiveram suas bordas melhoradas e equalizadas para fins de comparação.

Os autores demonstraram que a combinação do Micro-CT seguido por um preciso seccionamento é uma forma efetiva de fornecer informação de alta qualidade sobre a morfologia e anatomia de fósseis de plantas, entretanto esta combinação é cara e demorada.

C. *A proof of concept demonstration of the automated laser removal of rock from a fossil using 3D X-ray tomography data*

O trabalho desenvolvido por du Plessis [9] tem como foco a identificação dos limites do fóssil em uma rocha, afim de remover as partes que não dizem respeito ao mesmo com um laser. Para tanto é utilizado um Micro-CT afim de gerar um mapa de contorno evidenciando a diferença entre rocha e fóssil.

Esses dados são inseridos na estação de trabalho de processamento do laser, a qual é programada para remover a rocha do fóssil. Todo um processamento de remoção de ruídos com filtro gaussiano e correção da malha de pontos é realizado a fim de evidenciar os limites entre rocha e fóssil. O autor demonstra o resultado obtido para uma amostra, mas evidencia alguns desafios, tais como a densidade do fóssil e da rocha muitas vezes não é o suficiente para separar os limites entre ambos e a remoção da rocha no fóssil por outros ângulos.

D. *Computed 3D visualisation of an extinct cephalopod using computer tomographs*

No trabalho proposto Lukeneder [10] o autor faz uma análise morfológica de um cefalópode extinto, para tanto ele utiliza TC e dados paleontológicos, tais como formato, modo de vida e o habitat do animal em questão. A figura 1 mostra a metodologia empregada no trabalho.

Análises combinadas baseadas em objetos e vários outros programas ajudaram o autor a entender detalhes da morfologia assim como mudanças sofridas pelo material do fóssil. Os resultados desta análise são modelos geométricos 3D. Estes demonstram um grande número de aplicações, técnicas de análise e algumas possíveis limitações. O autor cita várias referências do uso de modelos 3D geométricos em estudos

paleontológicos, técnicas de análise aplicadas a uma grande gama de fósseis bem como as limitações destes modelos.

Por fim, ele demonstra como utiliza as informações obtidas para criar um modelo 3D do cefalópode, mostrando o mesmo movimentando-se em seu habitat natural. Técnicas de renderização e modelagem foram utilizadas para alcançar o resultado final (softwares tais como NUBS, Rhinoceros, 3D Studio max e 3D Coat foram utilizados).

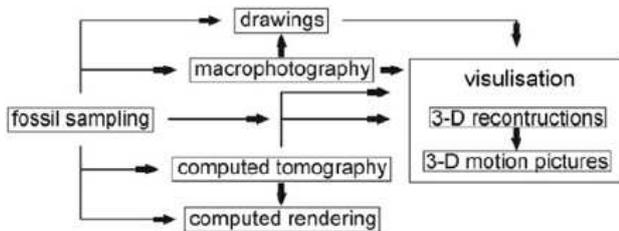


Fig. 1. Resumo da metodologia utilizada por Lukeneder.

E. Changes in fossil assemblage in sediment cores from Mingulay Reef Complex (NE Atlantic): Implications for coral reef build-up

O estudo realizado por Douarin [11] tem como foco analisar a formação de um recife de coral de águas frias. Para tanto foi utilizado Tomografia Computadorizada (TC) para analisar os cores extraídos de um recife da costa oeste da Escócia, buscando reconstruir e quantificar a relativa abundancia e preservação dos fragmentos do coral nos cores, identificar faces distintas do recife e, conseqüentemente, fornecer dados que poderiam indicar um modelo conceitual do crescimento do recife.

O autor diferencia 3 faces (camadas) evidenciadas pelo TC. A primeira é a camada rica em coral que foi caracterizada pela grande quantidade de clastos e coral e poucos fósseis. A segunda camada é caracterizada pelo grande número de fragmentos de fósseis, contendo não apenas bivalves, mas também outros grupos da fauna (briozoários, braquiópodes, equinodermes e gastrópodes). A terceira camada é caracterizada pela quase ausência de fragmentos de corais e outros fósseis, sendo formada basicamente por sedimentos.

Com base nessa separação proveniente das tomografias, o autor propõe um modelo de desenvolvimento do recife no tempo e no espaço.

F. Endocranial volume of *Australopithecus africanus*: New CT-based estimates and the effects of missing data and small sample size

Neubauer mostra a importância de se estudar o volume endo-cranial afim de avaliar a evolução do cérebro na espécie em questão [12]. Para realizar tal estudo é utilizado imagens de TC, e sobre estas realizado um trabalho de preparação, espalhamento da imagem e morfometria geométrica com base em semi-pontos de referência afim de gerar e reconstruir os moldes internos e medir os volumes endo-cranias dos mesmos.

Os três passos seguidos para gerar esses moldes internos virtuais são:

- Segmentação : segmentando a cavidade cranial a qual contém ar, o osso fossilizado, a matriz da rocha e o material utilizado para reconstrução física.
- Espelhamento da imagem: uma cópia completa do fóssil/molde foi espelhada e alinhada com a imagem original para preencher as regiões faltantes.
- Reconstrução geométrica: com base nos moldes internos virtuais das amostras de referência de chimpanzés, estima-se o formato das regiões que ainda estão faltando.

Para realizar esse processo são medidos 29 pontos de referência e 58 semi pontos de referência nas curvas endo-craniais, assim como 358 semi pontos de referência na superfície endo-cranial. Um total de 445 pontos tridimensionais foi obtido para cada um dos indivíduos modernos. Para espécimes fósseis, foi medido semi pontos e pontos de referência localizados em regiões preservadas. De acordo com estes pontos de referência disponíveis, foi utilizado uma interpolação pelo método de *thin-plate spline* para mapear as regiões faltantes correspondentes do molde interno de referência para o molde interno incompleto do *A. africanus*. A amostra de referência incluiu 60 moldes internos de chimpanzés e outro molde interno previamente analisado. Por fim o autor faz uma análise estatística, com muitas e com poucas amostras, de estimar o volume endo-cranial.

As simulações demonstraram que com poucas amostras se torna problemático estimar o volume endo-cranial. Mostram também que incluindo o tamanho da amostra e incluindo a reconstrução de espécimes parcialmente preservadas irá fornecer uma melhor estimativa da média do volume endo-cranial da espécie.

G. Convergence analysis of a finite element skull model of *Herpestes javanicus* (Carnivora, Mammalia): Implications for robust comparative inferences of biomechanical function

O trabalho idealizado por Tseng [13] mostra uma análise de convergência para modelos 3D de elementos finitos para crânios de mangusto. Elementos finitos é uma técnica aplicada para análise comparativa da função cranio-dental em vertebrados extintos e existentes. Originalmente a análise de elementos finitos foi um método de engenharia desenvolvido para resolver problemas complexos do mundo real envolvendo estrutura continuas, utilizando representações virtuais, simplificadas e discretas destas estruturas. Com avanço da tecnologia e do poder computacional e a contínua redução do preço de softwares e hardwares necessários para conduzir a análise de elementos finitos esta técnica foi mais frequentemente utilizada na área da biologia. Para biólogos, análise de elementos finitos são formas para prever e testar implicações funcionais das variações nas propriedades das estruturas e materiais de um organismo. Construir modelos de elementos finitos de estruturas biológicas complexas que refletem a realidade é uma tarefa desafiadora, pois os elementos finitos utilizados para retratar a morfologia são relativamente simples blocos de construção (triângulos, tetraedros, entre outros). Análises

de sensibilidade, verificação e validação são os métodos mais comuns para melhoria dos modelos de elementos finitos.

Por fim o autor conduz uma análise de convergência em uma série de modelos de elementos finitos de crânios de *H. janvanicus* construídos utilizando um protocolo previamente descrito que utiliza um algoritmo de cobrimento de músculo para simular a contração muscular do fechamento da mandíbula. Nesta análise o autor não achou uma convergência consistente das 23 medidas biomecânicas entre os modelos de alta resolução. Concluindo que simples modelos em alta resolução não são suficientes para análise comparativa de funções biomecânicas, pois incertezas nos valores de saída não são necessariamente baixas em modelos com grande quantidade de elementos finitos.

H. Micro-CT of fossils preserved in amber

Visando mostrar a importância do Micro-CT para o campo de fósseis em âmbar, Dierick [14] utiliza exemplos que tipicamente apresentam problemas para outros métodos ópticos padrões utilizados para investigação. Os dados dos fósseis em âmbar foram reconstruídos com o Octopus, o qual possui uma adaptação especial do algoritmo FDK descrito em *Principles of Computerized Tomographic Imaging*. Uma correção moderada do efeito de *beam hardening* foi aplicada onde necessária para corrigir este artefato proveniente da natureza policromática do raio-X. Outro problema corrigido é que frequentemente o pedaço de âmbar é maior que o fóssil. No contexto de tomografia, isto é importante porque em teoria o objeto completo deveria permanecer no campo de visão através da aquisição. Isto limita a ampliação e consequentemente resolução espacial que pode ser alcançada com o fóssil. Na chamada região de escaneamento de interesse, a ampliação é escolhida de tal forma que o fóssil esteja no campo de visão em vez do âmbar. Teoricamente isto significa que o objeto não pode ser reconstruído exatamente. Mas a natureza dos fósseis em âmbar possibilita reduzir o artefato resultante com um procedimento simples de não preenchidos com zeros na direita e na esquerda antes do filtro de fourier mover os erros de truncamento fora da área de interesse. Para reduzir os artefatos do escaneamento com região de interesse, foi adaptado este processo de tal forma que o dado de projeção fosse primeiramente preenchido com a média de uma banda de fora dos dados de projeção e posteriormente com zeros. Graças ao fato que a matriz do âmbar é relativamente homogêneo este procedimento mostra ser funcional para este caso. Para visualizar o resultado da digitalização foi utilizado o VGStudio Max. Por fim com base nos exemplos, o autor conclui que o Micro-CT é uma ferramenta valiosa no estudo de fósseis em âmbar, devido ao fato de que mais detalhes taxonômicos são preservados de forma que uma identificação positiva pode ser alcançada com o uso de apenas Micro-TC.

I. Middle Pleistocene human facial morphology in an evolutionary and developmental context

O trabalho apresentado por [15] estuda tendências temporais na morfologia facial afim de melhor entender a polaridade das características e descrever o padrão alométrico do formato da

face. Para tanto foi realizado um processo de digitalização e análise de medidas. O modelo tridimensional da superfície foi extraído dos escaneamentos da superfície ou dos dados de CT e foram processados utilizando Geometric Studio ou OptoCat. Para os dados de TC a superfície foi extraída utilizando Avizo. Os pontos de referência foram digitalizados nos modelos de superfície utilizando o Landmark Editor. Dados dos pontos de referência e dos semi pontos de referência definindo curvas e superfícies foram digitalizados por um observador. As coordenadas 3D destes pontos ao longo da curva foram digitalizados para todas as espécies e a malha dos pontos de referência da superfície foi digitalizado em um indivíduo modelo utilizando o Landmark Editor. Seguindo um protocolo que garante que todas as espécies tem o mesmo número de pontos de referência e semi pontos de referência, aproximadamente, nas localizações correspondentes. Os semi pontos de referência inicialmente equidistantes foram deslizados ao longo das tangentes para as curvas e dos planos tangentes para as superfícies de forma a minimizar a energia de curvatura da interpolação thin-plate spline (TPS) entre cada espécime e a configuração *Procrustes consensus*.

Após o deslizamento, os pontos de referência e semi pontos de referência foram tratados da mesma forma na análise estatística subsequente. Para converter as coordenadas dos pontos de referência em variáveis de formato, uma Generic Procrustes Analysis (GPA) genérico foi realizado. GPA remove os efeitos de translação e rotação nos dados de coordenadas e padroniza cada espécime para unidade do tamanho do centroide. Para a parte de dados faltantes, dado que métodos de geometria morfométrica requer que toda as espécimes tenham o mesmo número de pontos homólogos, uma pequena reconstrução de dados foi necessária para alguns fósseis. Primeiramente, a simetria bilateral foi explorada através do espelhamento da superfície do lado melhor preservado ao longo do plano sagital médio utilizando Geomagic e Avizo. Depois, se o dado faltante ocorreu bilateralmente em um espécime ou ao longo de sua linha média, pontos de referência foram estimados utilizando reconstrução geométrica via interpolação TPS. Para explorar a variabilidade temporal na morfologia facial, uma Análise de Componentes Principais (ACP). Foi utilizado deformação por TPS do formato médio do *Procrustes* ao longo dos dois principais componentes para visualizar a mudança de formato. Adicionalmente, para identificar quais indivíduos são mais similares, Nearest Neighbors foram calculados utilizando a distancia *Procrustes* entre indivíduos. Em morfometria, é feita uma distinção entre formato e forma. formato refere a propriedade geométrica de um objeto que é independente de seu tamanho médio, posição e orientação, enquanto que a forma de um objeto inclui ambos seus formato e tamanho. Dado que esse estudo foca em alometria, a inclusão do tamanho na análise estatística é essencial. Para quantificar o tamanho facial, utilizou-se o tamanho do centroide variável. Depois, foi realizado um ACP no espaço da forma do *Procrustes*. Espaço da forma o tamanho geométrico de cada espécime o que é valioso por causa da relação entre formato e tamanho podem ser facilmente exploradas.

Por fim o autor conclui com base nas comparações e análise estatísticas que se pode dividir humanos *Pleistocene*

em grupos temporais de acordo com sua morfologia facial, entretanto o limite entre *MPhs* e *Neanderthals* é menos óbvio.

J. Digitizing, modelling and 3D printing of skeletal digital models of *Palaeoloxodon tiliensis* (Tilos, Dodecanese, Greece)

O artigo elaborado por Mitsopoulou [16] demonstra a metodologia utilizada, figura 2, para digitalização, modelagem e impressão 3D do esqueleto do fóssil em questão. Dados tafonômicos foram combinados com métodos matemáticos apropriados e análise alométrica a fim de determinar as medidas dos ossos perdidos e estimar corretamente as proporções relativas dos elementos do esqueleto. Tomografia Computadorizada (TC) e digitalização via escaneamento a laser foram utilizados para capturar a morfologia da superfície e criar modelos 3D que são ajustados com as dimensões corretas derivadas da análise matemática. Neste trabalho, atlas dos modelos 3D do *P. tiliensis* e *Palaeoloxodon antiquus* foram combinados em um modelo 3D único que quantifica diferenças morfológicas pela escala de cor, assim minimizando erros de observação.

Por fim, os modelos gerados do esqueleto do fóssil foram impressos com uma tecnologia de prototipação rápida. Como grande vantagem de se utilizar modelos 3D e réplicas de impressão 3D tem-se a facilidade de uma maior interação entre instituições científicas, minimizando custos e riscos relacionados com a transferência de fósseis insubstituíveis. Como resultado positivo deste trabalho, tem-se o uso desta metodologia em atividades educacionais em escolas, universidades e museus.

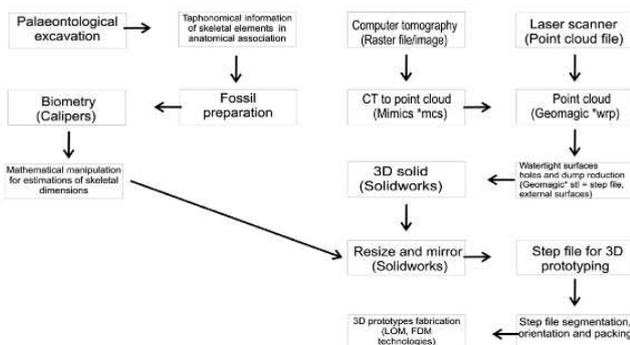


Fig. 2. Resumo da metodologia utilizada por Mitsopoulou.

K. Miocene *Clypeaster* from Valencia (E Spain): Insights into the taphonomy and ichnology of bioeroded echinoids using X-ray micro-tomography

No estudo apresentado por Rahman [17] análise tafonômica (termo para descrever o estudo da transição de restos, peças ou produtos de organismos, na biosfera para a litosfera, criando um conjunto de fósseis) e micro tomografia foram aplicadas em espécimes de *Clypeaster*, possibilitando a identificação de vários fatores de preservação, incluindo incrustação, bioerosão, abrasão e rachaduras pós-deposicional. A excelente preservação e características morfológicas, como revelados

pelo Micro-TC, acompanhado com a observação que (em certos casos) conchas articuladas bivalves estão ainda presentes dentro de câmaras, permitindo a identificação do fóssil como o bivalve *Rocellaria*, reafirmando esses animais de tubo como perfuradores, mutuários e construtores de criptas. Para fornecer um visão geral da preservação dos fósseis, uma análise tafonômica, gravando três atributos principais (fragmentação, abrasão e atividade biológica, como por exemplo bioerosão e incrustação) foi realizada em 75 espécimes de *Clypeaster*. A reconstrução tomográfica foi realizada utilizando o software Nikon CT Pro, fornecendo um conjunto de dados baseado em fatias com um tamanho de *voxel* de 60 micrômetros.

O modelo 3D virtual foi construído utilizando o SPIERS, da seguinte forma: primeiramente, *pixels* são atribuídos para um dos dois segmentos com base em seu brilho: *pixels* cinza claro e cinza escuro representam o equinoide e o sedimento são atribuídos para o segmento da matriz; *pixels* pretos representando os espaços vazios internos são atribuídos para o segmento "molde". Regiões individuais de interesse foram manualmente definidas para ambos os segmentos; a alta densidade dos tubos tornou possível distinguir todos os traços, mas 19 do exemplos melhores preservados foram atribuídos para regiões de interesse distintas. Por fim, múltiplas *isosurfaces* representando diferente regiões de interesse/segmentos foram gerados para produzir uma reconstrução interativa virtual, a qual foi submetida a filtros de suavização e de remoção de ilhas para minimizar o ruído. O resultado foram visualizados utilizando a técnica de Raytracing presente no software Blender. Com base nesta reconstrução foi realizada uma análise que vai desde a análise tafonômica até a parte de partes faltantes e morfologia. Destas análise o autor conclui que as espécies *Clypeaster* apresentam grande variação tafonômica; o Micro-CT quando aplicado a um único indivíduo revela intensa bioerosão; a excelente preservação e características morfológicas do fóssil, como mostrado pelo Micro-CT, permite identificar o gênero *Rocellaria*.

L. Magnetic resonance spectroscopy and imaging for the study of fossils

O trabalho elaborado por Giovannetti [18] mostra como Espectroscopia de Ressonância Magnética (ERM) e Imagem de Ressonância Magnética (IRM) ganharam espaço como ferramenta de análise para exame de amostras paleontológicas, através do fornecimento de informações, de forma não destrutiva, da estrutura e composição dos fósseis. Enquanto IRM é capaz de revelar a geometria tridimensional do fóssil, o ERM pode fornecer informações sobre a composição química das amostras. De forma geral TC fornece uma boa visualização de tecidos rígidos existentes e de fósseis e IRM é a técnica padrão ouro para investigar tecidos suaves e a morfogênese de esqueletos. O autor descreve o funcionamento e o uso do ERM e do IRM em aplicações que partem desde a análise na parte química voltada para a área petrolífera até a identificação da evolução do esqueleto humano, levantando vantagens e desvantagens de ambos os métodos.

M. Protostylid expression at the enamel-dentine junction and enamel surface of mandibular molars of Paranthropus robustus and Australopithecus africanus

No trabalho apresentado por Skinner [19] tem-se a análise das características dentárias de um fóssil de homínido, utilizando para tal imagens de Micro-TC. Esta análise auxilia no entendimento de como se deu o desenvolvimento das características morfológicas dentárias e melhora interpretações sobre a variabilidade destas características dentro e ao longo da classe. O processo para esta análise, partindo da parte de digitalização, se resume na aplicação do filtro mediana, seguido pela média do filtro *least variance*, o que implica classes de tecidos mais homogêneas e aloca *pixels* com valores médios de intensidade nas divisões entre tecidos aos tecidos apropriados. Em fósseis de dentes onde a dentina e o esmalte diferem substancialmente no grau de mineralização, o processo de filtragem resulta em distribuições de valores de *pixels* para cada tecido que não se sobrepõem, permitindo simples segmentação de tecido. Depois da segmentação os fósseis foram reconstruídos como modelos de superfície baseados em triângulos utilizando Amira e por fim esses modelos reconstruídos foram medidos utilizando uma ferramenta de medição existente no Amira. Com base nessas medições foi realizado uma análise estatística comparando as diferenças morfológica dentárias entre as duas espécies de homínidos sendo estudadas, o que resultou na constatação de uma diferença significativa na estrutura morfológica entre estas duas espécies, a qual pode dar suporte para a filiação taxonômica para essas duas classes.

N. Quantitative estimation of bioturbation based on digital image analysis

O estudo elaborado por Dorador [20] faz uma breve revisão da literatura sobre trabalhos que realizaram uma determinação quantitativa da modificação das características dos sedimentos primários, denominada *bioturbation*, a qual é essencial na geociência. Depois apresenta um compilado de diferente métodos utilizando o Adobe Photoshop software CS6, para estimativa digital que é uma parte do IDIAP (Ichnological Digital Analysis Images Package), a qual é recomendada para amostras de núcleos. Os métodos apresentados, Similar Pixel Selection Method (SPSM), Magic Wand Method (MWM) e Color Range Selection Method (CRSM), possuem vantagens e desvantagens dependendo das características do sedimento (por exemplo, composição, cor, textura, porosidade, entre outros) e características Icnológicas (tamanho dos traços, material de preenchimento, entre outras). O IDIAP fornece uma estimativa da quantidade traços fósseis produzidos por um *ichnotaxon*, por um inteira *ichnocoenosis* ou até mesmo por um completo *ichnofabric*. O SPSM permite acha *pixels* na imagem inteira com valores dentro alcance dos *pixels* registrados anteriormente em uma seleção de uma região em particular. Isto se torna útil quando objetos precisam ser selecionados com um alcance de cor variável porque uma área com um certo número de *pixels* foi selecionada. Em casos onde objetos não podem ser definidos por uma cor, este método tem a vantagem, entretanto a grande desvantagem deste método é

que a sensibilidade selecionada não pode ser modificada, ou seja, o método não é útil se a diferença dos *pixels* é baixa.

O MWM seleciona *pixels* próximos que possuem variações de cores similares ao pixel anteriormente selecionado. A porcentagem máxima da diferença entre o pixel de referência e o pixel selecionado é definido por um parâmetro chamado de tolerância. A desvantagem é que pode consumir muito tempo especialmente se existem objetos pequenos na imagem. Entretanto, o MWM oferece uma grande vantagem de execuções múltiplas. Este método é útil especialmente quando objetos são grandes e possuem mais que uma cor. O CRSM é baseado na localização dos *pixels* na imagem toda que são similares a um único pixel selecionado manualmente. Compartilha a grande vantagem do SPSM, a seleção se estende por toda a imagem, mas no caso do CRSM apenas um pixel representativo pode ser selecionado, enquanto no SPSM todos os *pixels* registrados na área selecionada são considerados. O CRSM permite escolher uma seleção ou um alcance de pixel desejado pela modificação do parâmetro denominado *fuzziness*. Este método é o mais rápido e mais útil quando existem muitos pequenos objetos de cor igual ou similar.

O. Solid modeling of fossil small mammal teeth

No trabalho realizado por Marschallinger[21] mostra-se a combinação de Micro-CT, *Object Base Image Analysis* (OBIA) e modelagem solida para uma compreensiva descrição da arquitetura interna de um fóssil de um pequeno dente de um mamífero. O Micro-CT como método de análise inerentemente tridimensional, rápido e sem destruição do objeto sendo analisado gera uma pilha de imagens em tons de cinza de alta resolução. Com a finalidade de pavimentar o caminho para modelos 3D sólidos dos dentes de mamíferos, as imagens de Micro-CT parcialmente ruidosas com vários artefatos são processadas pelo OBIA, incorporando conhecimento externo de um especialista sobre relações topológicas no dente. Formulando o conhecimento de um especialista em micro paleontologia como um conjunto de regras permite o processamento reprodutível automatizado de uma variedade de tipos de dentes de mamíferos.

O OBIA produz uma pilha de imagens que são classificadas de acordo com os componentes principais dos dentes dos mamíferos. Encaminhando esta saída para um sistema de modelagem de *voxel*, o acesso é fornecido não só para as superfícies dos dentes de mamíferos, mas também para as estruturas internas destes. Os modelos 3D resultantes visam servir como uma base original para classificação de dentes de mamíferos, análise automática e comunicação de resultados de pesquisa, porque os modelos sólidos gerados são quantitativos, totalmente reproduzíveis e podem ser facilmente compartilhados entre paleontologistas via internet.

P. Structural analysis of the femoral diaphyses of an early modern human from Tianyuan Cave, China

O estudo desenvolvido por Wei[22] mostra uma análise estrutural da diáfise femoral de um fóssil humano encontrado em uma caverna na china denominado Tianyuan 1. Inicialmente o fóssil foi digitalizado utilizando um Micro-CT e, assim como

em outros fósseis que foram digitalizados para comparação, foi digitalizado com parâmetros específicos visando a análise do fêmur. Depois da reconstrução dos fósseis, os mesmos foram segmentados e renderizados utilizando Avizo, empregando limiares específicos para a imagem do fóssil Tianyuan 1 e para os fósseis de comparação.

Uma vez renderizados, os modelos digitais dos fêmur foram separados utilizando o módulo *Label Field* do Avizo. Devido a falta de algumas partes do fêmur esquerdo do Tianyuan 1 foi necessário uma reconstrução antes de ser possível extrair os dados necessários da parte estrutural. As etapas de reconstrução envolveram os seguintes passos: Primeiro, o fêmur direito foi alinhado utilizando um protocolo padrão. Depois, a superfície 3D do fêmur esquerdo foi espelhada e alinhada com a superfície do fêmur direito ao longo de seu eixo principal. Esta etapa envolveu um conjunto de pontos de referência para transladar e rotacionar os ossos até obter uma correspondência razoável assim como foi utilizado a opção *Rigid Alignment* das superfícies para minimizar a diferença entre as superfícies 3D. Por fim, as porções do fêmur esquerdo que estão faltando no fêmur direito foram sobrepostas nas imagens de raios-X do fêmur direito.

Após o processo de reconstrução, realizou-se o processo de criação do mapa de cores para visualização da distribuição da espessura do osso cortical da diáfise. Primeiro, o fêmur dos exemplares utilizados para comparação foram posicionados no Avizo utilizando um protocolo padrão. Depois, um destes exemplares foi reescalado para estimar o comprimento biomecânico do fêmur do Tianyuan 1. Em seguida, o fêmur do Tianyuan 1 foi alinhado ao fêmur reescalado utilizando a ferramenta de alinhamento do Avizo e os pontos de referência, rodando 50 iterações afim de fazer o casamento dos pontos de referência. Então, foram extraídos dados de espessura do osso cortical em intervalos regularmente espaçados. Em todas as 61 seções foram extraídas diáfises, as quais foram padronizadas e salvas como imagem. Por fim, valores de espessura foram padronizados pelo comprimento biomecânico do fêmur e mapas de cores 2D foram criados. Em cada mapa 2D a cor de cada pixel no mapa representa uma terceira dimensão e corresponde a uma espessura da diáfise do osso cortical. Com base nestes mapas 2D uma análise estrutural da diáfise do fêmur do Tianyuan 1 é realizada, a qual oferece suporte a este indivíduo pertencer a classe *Homo sapiens*.

Q. Three-dimensional microCT analysis of the Ediacara fossil Pteridinium simplex sheds new light on its ecology and phylogenetic affinity

O artigo apresentado por Meyer [23] faz uma análise 3D utilizando Micro-CT do fóssil *Pteridinium simplex*, o qual está frequentemente preservado em 3D em moldes em quartzito de granulação grossa. Para tal análise a amostra foi cortada em 3 pedaços, sendo duas delas analisada em Micro-CT. Os dados de escaneamento foram processados para reduzir artefatos provenientes do processo de digitalização e armazenados em arquivos TIFF 16 bits. Depois, os dados gerados foram processados utilizando ferramentas de processamento de imagem, segmentação e renderização de volume para

criar reconstruções 3D das características de interesse. Devido a grande tamanho do volume de dados do dado original as imagens foram reduzidas em termos de resolução e número de bits para representação (16 para 8 bits) utilizando imageJ, afim de facilitar a renderização. Os volumes foram então segmentados em duas classes utilizando uma combinação manual de pintura de máscara e ferramentas semiautomáticas disponíveis no Seg3D. As máscaras de interesse foram exportadas como uma pilha de imagens, redimensionadas utilizando o imageJ, para uma resolução menor buscando renderização em tempo real, e convertidos para NRRD utilizando o UNU.

Por fim, os segmentos NRRD foram compostos e renderizados utilizando H3D, onde cada segmento foi representado como uma *isosurface* e unicamente colorido de forma falsa para uma fácil visualização. Utilizando o H3D, os dois pedaços foram dimensionados e costurados para restaurar sua configuração pré-corte. Com base nesta reconstrução, as análises morfológicas foram realizadas o que permitiu chegar a um morfologia básica do fóssil em questão, salientando a importância do uso do Micro-CT como uma forma efetiva de estudo de espécimes preservadas tridimensionalmente dentro de depósitos de fluxo de massa.

R. IPKB: A Digital Library for Invertebrate Paleontology

O autor apresenta uma base de conhecimento para paleontologia com foco em invertebrados [24]. Caracterizado como um esforço para digitalizar e compartilhar o tratado em paleontologia invertebrada. Este tratado é a mais autoritária compilação de registros de fósseis de invertebrados, entretanto a versão PDF deste é um simples compilado de artigos e seu conteúdo não possui uma organização que facilite a sua consulta. Para solução deste problema foram extraídos textos e imagens do tratado, armazenados em uma base de dados e construído um sistema eficiente para navegação e procura.

Na parte de processamento de imagens, descreve-se um processo de segmentação e identificação de contornos e sua posterior comparação com *templates*. Para o processo de segmentação é utilizado um método similar ao watershed, para o método de identificação de contornos foi utilizado o sobel e para a parte de reconhecimento é medido a distância entre a imagem e os *templates* e o *template* com a distância mínima (*Sum of square error*) é classificação a ser utilizada. Por fim o autor conclui falando da biblioteca digital e outras partes desta biblioteca, como acesso, indexação, procura, resultados obtidos e trabalhos futuros.

V. DISCUSSÃO

De forma geral todos os trabalhos selecionados utilizam técnicas de processamento de imagens para a análise das imagens de fósseis, entretanto grande parte destes depende de uma exaustiva iteração manual por parte do operador do método. Visto que o estudo que originou esta revisão está direcionado na geração de um encadeamento de métodos, os quais necessitem o mínimo possível da iteração do operador, para fins de identificação e classificação de fósseis em imagens tomográficas, os trabalhos analisados servem para fornecer

uma visão geral dos métodos geralmente aplicados para os estudos voltados para esta área.

Também pode-se apontar que a revisão proposta, seguindo a metodologia de Kitchenham, demonstra: (i) uma busca reproduzível com uma metodologia sistemática para revisar os métodos computacionais aplicados em imagens de fósseis no período de 2006 até julho de 2016; (ii) um resumo de todos os trabalhos analisados que tiveram uma correspondência com os critérios estabelecidos, especificando os principais métodos de VC, o tipo de imagem utilizado e os experimentos realizados por cada trabalho; (iii) uma visão geral analisando, com base nos trabalhos selecionados, a área de VC aplicada as imagens de fósseis.

A. Conclusão

A revisão sistemática da literatura apresenta um visão geral em relação a área de métodos computacionais aplicados a imagens de fósseis. Esta pesquisa incluiu artigos publicados entre 2006 e julho de 2016 em 3 bases de dados de artigos científicos (Science Direct, ACM and IEEEExplore). No geral, os trabalhos analisados mostraram técnicas de VC para a análise das imagens de fósseis, entretanto grande partes destes depende de iteração manual por parte do operador do método. Desta forma, os métodos avaliados são custoso em termos de tempo de operação e inviabilizam a execução de uma análise automática.

Esta revisão com a análise dos método computacionais e experimentos realizados sob as imagens de fósseis é útil para pesquisas que trabalham com imagens de fósseis e necessitam de métodos de VC para uma melhor compreensão e caracterização da imagem.

REFERENCES

- [1] J. C. Tipper, "Computer applications in paleontology: Balance in the late 1980s?" *Computers & Geosciences*, vol. 17, no. 8, pp. 1091 – 1098, 1991.
- [2] B. Kitchenham, "Procedures for Performing Systematic Reviews," Joint Technical Report TR/SE-0401, Tech. Rep., 2004.
- [3] N. Health, M. R. C. (Australia), and N. Staff, *How to Review the Evidence: Systematic Identification and Review of the Scientific Literature*, ser. Handbook series on preparing clinical practice guidelines. National Health and Medical Research Council, 2000. [Online]. Available: <https://books.google.com.br/books?id=G8JmAAAACAAJ>
- [4] —, *How to Use the Evidence: Assessment and Application of Scientific Evidence*, ser. Handbook series on preparing clinical practice guidelines. National Health and Medical Research Council, 2000. [Online]. Available: <https://books.google.com.br/books?id=7ioKGQAACAAJ>
- [5] C. Collaboration, *Cochrane Reviewers' Handbook*, ser. Version 4.2.1. National Health and Medical Research Council, 2003.
- [6] U. of York. NHS Centre for Reviews and Dissemination, *Undertaking systematic reviews of research on effectiveness: CRD's guidance for those carrying out or commissioning reviews*, ser. CRD report. NHS Centre for Reviews and Dissemination, University of York, 2001. [Online]. Available: <https://books.google.com.br/books?id=2rhpAAAAMAAJ>
- [7] J. A. Cunningham, I. A. Rahman, S. Lautenschlager, E. J. Rayfield, and P. C. Donoghue, "A virtual world of paleontology," *Trends in Ecology & Evolution*, vol. 29, no. 6, pp. 347 – 357, 2014.
- [8] A. R. Spencer, J. Hilton, and M. D. Sutton, "Combined methodologies for three-dimensional reconstruction of fossil plants preserved in siderite nodules: *Stephanospermum braidwoodensis* nov. sp. (medullosales) from the mazon creek lagerstätte," *Review of Palaeobotany and Palynology*, vol. 188, pp. 1 – 17, 2013.
- [9] A. du Plessis, J. Steyn, D. Roberts, L. Botha, and L. Berger, "A proof of concept demonstration of the automated laser removal of rock from a fossil using 3d x-ray tomography data," *Journal of Archaeological Science*, vol. 40, no. 12, pp. 4607 – 4611, 2013.
- [10] A. Lukeneder, "Computed 3d visualisation of an extinct cephalopod using computer tomographs," *Computers & Geosciences*, vol. 45, pp. 68 – 74, 2012.
- [11] M. Douarin, D. J. Sinclair, M. Elliot, L.-A. Henry, D. Long, F. Mitchison, and J. M. Roberts, "Changes in fossil assemblage in sediment cores from mingulay reef complex (ne atlantic): Implications for coral reef build-up," *Deep Sea Research Part II: Topical Studies in Oceanography*, vol. 99, pp. 286 – 296, 2014, biology and Geology of Deep-Sea Coral Ecosystems: Proceedings of the Fifth International Symposium on Deep Sea Corals.
- [12] S. Neubauer, P. Gunz, G. W. Weber, and J.-J. Hublin, "Endocranial volume of australopithecus africanus: New ct-based estimates and the effects of missing data and small sample size," *Journal of Human Evolution*, vol. 62, no. 4, pp. 498 – 510, 2012.
- [13] Z. J. Tseng and J. J. Flynn, "Convergence analysis of a finite element skull model of herpestes javanicus (carnivora, mammalia): Implications for robust comparative inferences of biomechanical function," *Journal of Theoretical Biology*, vol. 365, pp. 112 – 148, 2015.
- [14] M. Dierick, V. Cnudde, B. Masschaele, J. Vlassenbroeck, L. V. Hoorebeke, and P. Jacobs, "Micro-ct of fossils preserved in amber," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 580, no. 1, pp. 641 – 643, 2007, proceedings of the 10 th International Symposium on Radiation Physics ISRP 10.
- [15] S. E. Freidline, P. Gunz, K. Harvati, and J.-J. Hublin, "Middle pleistocene human facial morphology in an evolutionary and developmental context," *Journal of Human Evolution*, vol. 63, no. 5, pp. 723 – 740, 2012.
- [16] V. Mitsopoulou, D. Michailidis, E. Theodorou, S. Isidorou, S. Rousiakis, T. Vasilopoulos, S. Polydoros, G. Kaisarliis, V. Spitas, E. Stathopoulou, C. Provatidis, and G. Theodorou, "Digitizing, modelling and 3d printing of skeletal digital models of palaeoloxodon tiliensis (tilos, dodecanese, greece)," *Quaternary International*, vol. 379, pp. 4 – 13, 2015, mammoths and their Relatives: {VIth} International Conference, Grevena-Siatista, Greece, part 1.
- [17] I. A. Rahman, Z. Belaústegui, S. Zamora, J. H. Nebelsick, R. Domènech, and J. Martinell, "Miocene clypeaster from valencia (e spain): Insights into the taphonomy and ichnology of bioeroded echinoids using x-ray micro-tomography," *Palaeogeography, Palaeoclimatology, Palaeoecology*, vol. 438, pp. 168 – 179, 2015.
- [18] G. Giovannetti, A. Guerrini, and P. A. Salvadori, "Magnetic resonance spectroscopy and imaging for the study of fossils," *Magnetic Resonance Imaging*, vol. 34, no. 6, pp. 730 – 742, 2016.
- [19] M. M. Skinner, B. A. Wood, and J.-J. Hublin, "Protostyloid expression at the enamel-dentine junction and enamel surface of mandibular molars of paranthropus robustus and australopithecus africanus," *Journal of Human Evolution*, vol. 56, no. 1, pp. 76 – 85, 2009.
- [20] J. Dorador and F. J. Rodríguez-Tovar, "Quantitative estimation of bioturbation based on digital image analysis," *Marine Geology*, vol. 349, pp. 55 – 60, 2014.
- [21] R. Marschallinger, P. Hofmann, G. Daxner-Höck, and R. A. Ketcham, "Solid modeling of fossil small mammal teeth," *Computers & Geosciences*, vol. 37, no. 9, pp. 1364 – 1371, 2011.
- [22] P. Wei, I. J. Wallace, T. Jashashvili, C. M. Musiba, and W. Liu, "Structural analysis of the femoral diaphyses of an early modern human from tianyuan cave, china," *Quaternary International*, vol. 434, Part A, pp. 48 – 56, 2017, quaternary Biostratigraphy in East Asia: A Multidisciplinary Research Approach on Gigantopithecus Fauna and Human Evolution. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S104061821501085X>
- [23] M. Meyer, D. Elliott, A. D. Wood, N. F. Polys, M. Colbert, J. A. Maisano, P. Vickers-Rich, M. Hall, K. H. Hoffman, G. Schneider, and S. Xiao, "Three-dimensional microct analysis of the ediacara fossil pteridium simplex sheds new light on its ecology and phylogenetic affinity," *Precambrian Research*, vol. 249, pp. 79 – 87, 2014.
- [24] Y. Meng, J. Li, P. Denton, Y. Chen, B. Luo, P. Selden, and X.-w. Chen, "Ipkb: A digital library for invertebrate paleontology," in *Proceedings of the 12th ACM/IEEE-CS Joint Conference on Digital Libraries*, ser. JCDL '12. New York, NY, USA: ACM, 2012, pp. 101–110.

**ANEXO B – RELATÓRIO TÉCNICO - REVISÃO DA LITERATURA PARA
ALGORITMOS DE SEGMENTAÇÃO 3D APLICADOS A IMAGENS
TOMOGRÁFICAS**



INCoD
Instituto Nacional para Convergência Digital



ISSN 2236-5281

Relatório Técnico
INCoD/LAPIX.28.2017.P.v01

Revisão da literatura para algoritmos de segmentação 3D aplicados a imagens tomográficas

Autores:

Luis Eduardo R. De Carvalho
Aldo von Wangenheim

Versão 1.0

Status: Final

Distribuição: Externa

DEZEMBRO - 2017

© 2017 **INCoD** – Instituto Nacional para Convergência Digital

Todos os direitos reservados e protegidos pela Lei 9.610 de 19/02/1998. Nenhuma parte deste documento, sem autorização prévia por escrito do Instituto, poderá ser re-produzida ou transmitida sejam quais forem os meio empregados: eletrônicos, mecânicos, foto-gráficos, gravação ou quaisquer outros.



Instituto Nacional para Convergência Digital

Universidade Federal de Santa Catarina - UFSC
Campus Universitário João David Ferreira Lima - Trindade
Departamento de Informática e Estatística - Sala 320
Florianópolis-SC - CEP 88040-970

Fone / FAX: +55 48 3721-9516 R.17

www.incod.ufsc.br

ISSN 2236-5281

Relatório Técnico do Instituto Nacional para Convergência Digital/
Departamento de Informática e Estatística, Centro Tecnológico,
Universidade Federal de Santa Catarina. -- v.1, n.1 (2017).--
Florianópolis: INE, UFSC, 2017-

Semestral
Resumo em português

ISSN 2236-5281

1. Convergência digital. 2. Segmentação 3D.
3. Imagens tomográficas 3D. 4. Revisão sistemática I. Universidade
Federal de Santa Catarina. Departamento de Informática e
Estatística.

Revisão da literatura para algoritmos de segmentação 3D aplicados a imagens tomográficas

Carvalho, L.E. e von Wangenheim, A

I. INTRODUÇÃO

Segmentação é o processo onde um dado sinal de entrada I é dividido em suas partes ou regiões constituintes [1]. Estas regiões devem compartilhar alguma propriedade local tal como intensidade de pixel, ou voxel quando se tratando de volumes, continuidade, variância, informação de textura, entre outros, geralmente guiadas por algum critério de similaridade utilizado para supervisionar o processo de fusão [2]. Este conceito genérico vai além da ideia de dimensionalidade e podendo ser aplicado em amostras bi-dimensionais ou volumétricas, dados de domínio real simples ou multivariados, valores com múltiplos alcances ou dados normalizados, entre outros.

Os estágios iniciais de segmentação de imagem foram iniciados com o caso de imagens bi-dimensionais e em tons de cinza. Com o aumento do poder computacional estes métodos foram rapidamente estendidos para suportar informações de cor. De fato, vários dos métodos encontrados na literatura são meras extensões de métodos desenvolvidos muitos anos atrás. Com a popularização dos métodos de aquisição de dados volumétricos por meio de dispositivos não destrutivos, tais como Tomografia Computadorizada (TC), Ressonância Magnética (RM) e suas derivações, métodos de segmentação 3D foram desenvolvidos para aplicações médicas, análise de materiais, geologia, inspeção de objetos em geral e outras aplicações. Apesar de uma quantidade significativa de métodos de segmentação 3D serem aplicados fatia por fatia, existem abordagens especificamente desenvolvidas para o caso 3D com base na computação do voxel. Com foco na análise dos métodos de segmentação 3D, este documento tem como propósito detalhar todos os procedimentos relacionados para a revisão sistemática da literatura, para o tópico de segmentação 3D aplicado à imagens tomográficas, para o período de 2006-2016 (julho), e realizar uma análise do estado da arte sobre o assunto, tendo em vista as possíveis aplicações destes métodos em outras áreas.

Este relatório técnico está dividido nas seguintes seções. A seção de metodologia, define quais foram os passos executados para a realização deste trabalho. A seção de descrição da pesquisa, detalha os parâmetros de pesquisa utilizados bem como critérios de seleção empregados na seleção dos trabalhos a serem analisados. A seção de trabalhos analisados, categoriza e detalha os trabalhos selecionados para análise, com foco na técnica de segmentação utilizada, nos experimentos realizados e na avaliação dos resultados. A análise e discussão sobre os trabalhos selecionados é realizada na seção de discussão. Por fim, a última seção fornece a conclusão e as futuras

perspectivas para a área.

II. METODOLOGIA

Este trabalho teve como base a metodologia proposta por Kitchenham[3] para conduzir o processo de pesquisa. Esta metodologia, para análise sistemática da literatura, tem como objetivo, como descrito pela própria autora, avaliar e interpretar todas as pesquisas relevantes disponíveis à uma particular pergunta de pesquisa, tópico ou fenômeno de interesse. Três guias utilizados para pesquisa na área da saúde serviram como base para a metodologia de Kitchenham [4] [5] [6] [7]. A principal característica ou propriedade desta metodologia é a de manter a procura reproduzível, quando todos os passos descritos forem repetidos e as mesmas palavras-chave forem utilizadas. Depois, para a análise e categorização dos resultados, utilizaram-se como referência os métodos tradicionais de segmentação e a frequência dos métodos de segmentação nos trabalhos analisados, tendo como base o conceito principal de construção dos métodos de segmentação sendo analisados. Por fim, uma tabela foi construída em ordem cronológica, resumindo os aspectos principais dos trabalhos avaliados.

III. DESCRIÇÃO DA PESQUISA

Inicialmente, com base em Kitchenham[3], definiu-se a tópico de pesquisa, **métodos de segmentação 3D aplicados a imagens tomográficas**. Com base neste tópico, definiram-se palavras chave e utilizou-se três bases de artigos científicos (ScienceDirect, ACM e IEEE) para a pesquisa. As palavras-chave utilizadas em cada base e os filtros de pesquisa foram: **ScienceDirect:** *pub-date > 2005 and pub-date < 2016 (july) and TITLE-ABSTR-KEY(3D segmentation) and TITLE-ABSTR-KEY(computer tomography)*.

ACM: (*"3D segmentation" +computer +tomography*) and *Published years: 2006 - 2016 (july)*.

IEEE: (*"3D segmentation"*) AND *computer tomography*) and *Year: 2006-2016 (july)*.

Obteve-se um total de 114 artigos, os quais foram analisados os *abstracts*. Na leitura dos *abstracts*, utilizaram-se os seguintes **critérios de exclusão**, afim de selecionar os trabalhos a serem analisados na íntegra:

- Artigos escritos em outras línguas diferentes da inglesa;
- Artigos repetidos e que não fazem referência ao estudo de métodos de segmentação 3D;
- O artigo não possui uma descrição de etapas computacionais realizadas (pré-processamento, pré-segmentações, entre outros);

Com a aplicação do critério de exclusão, obteve-se um total de 67 artigos, os quais foram lidos na íntegra, sumarizados, analisados e categorizados de acordo com o conceito principal de construção do método de segmentação em questão.

IV. TRABALHOS ANALISADOS

A. Métodos de threshold

1) *A novel approach to {CAD} system for the detection of lung nodules in {CT} images*: Javaid [8] apresenta uma abordagem para a detecção de nódulos de pulmões em imagens de tomografia computadorizada. O processo como um todo inicia com a parte de pré-processamento dos volumes tomográficos através da aplicação de um filtro para aumento do contraste. Depois, realiza-se a extração da região do pulmão das tomografias de tórax com a utilização de um thresholding associado ao refinamento do contorno com operações morfológicas. O valor do threshold esta diretamente relacionado com a qualidade da segmentação, logo, este valor é decidido com base nos histogramas do volume sobre análise. Após a extração da região dos pulmões, o próximo passo consiste na localização e segmentação dos nódulos em potencial. Esta etapa consiste em dois passos: no primeiro passo, todos os nódulos suspeitos são detectados e segmentados e no segundo passo um refinamento é realizado para nódulos grudados em vasos sanguíneos e bronquíolos. Para a segmentação inicial nesta etapa utiliza-se o k-means sendo atribuído 3 classes (plano de fundo, tecido do parênquima e nódulos candidatos, bronquíolos e vasos sanguíneos). Para a identificação dos nódulos presos a ramificações utiliza-se o operador de abertura com um formato específico para a quebra as conexões 2D entre os nódulos e as ramificações. Para a remoção das conexões 3D utiliza-se uma comparação pela área de regiões 3D conectadas para cada fatia, ou seja, se a diferença entre as maiores áreas e as menores áreas for maior que um certo limiar, as estruturas com menores áreas são removidas. A detecção de nódulos potenciais é dividida em seis grupos e para cada grupo extraem-se características para fins de classificação e eliminação de falsos positivos.

A classificação se dá através de Support Vector Machine (SVM), para nódulos grandes, e com um esquema com base em regras, para nódulos pequenos. A abordagem proposta foi aplicada em dados de tomografia da base de dados da Lung Image Database Consortium e em imagens clínicas do hospital Mayo. Um conjunto total de 110 volumes tomográficos foi utilizado e analisado por dois radiologistas para averiguação de nódulos. A parte de validação da abordagem se deu com base nas medidas de sensibilidade, especificidade, acurácia e no número de falsos positivos por caso.

B. Métodos com base em grafos

1) *Live-Wire-Based 3D Segmentation Method*: Wieclawek [9], propõe uma metodologia de segmentação 3D, com base em um método live-wire 2D, com duas significantes modificações, o mapa de custo com base em wavelet e o procedimento de clusterização através do Fuzzy C-Means (FCM). Estas duas modificações inseriram um aumento na acurácia e uma redução no custo computacional. O método do

live-wire 2D tem como base a análise das fronteiras e consiste de algumas etapas. A primeira etapa é o cálculo do mapa de custo da imagem, o qual reflete a propriedade de borda de cada pixel. Depois a imagem é representada por um grafo em que cada vértice corresponde a um pixel e referencia um elemento no mapa de custo e as arestas ligam os vértices com cada um de seus oito vértices vizinhos. No próximo passo, caminhos ótimos são encontrados no grafo. Para selecionar uma borda desejada, seleciona-se um ponto semente e um ponto livre e de forma iterativa é feita a segmentação. O processo termina quando encontra-se um contorno fechado, o qual permite a extensão do mesmo para as demais fatias na abordagem 3D. A extensão para o método 3D utiliza a mesma noção já apresentada e adiciona o uso de operadores morfológicos, preenchimento de buracos, a clusterização por meio do FCM para a fatia de referência e a próxima fatia e por fim a detecção do contorno para a próxima fatia, finalizando o processo quando não há o surgimento de mais nenhuma estrutura segmentada. O método proposto foi testado em imagens de tomografia computadorizada e ressonância magnética de pulmões e da junta do joelho e avaliado com base na sua resistência ao ruído.

2) *Segmentation of trabecular bones from Vertebral bodies in volumetric CT spine images*: Um framework de segmentação 3D de ossos trabecular em imagens de tomografia computadorizada é apresentado por Aslan [10]. Esse framework e sua posterior evolução [11], a qual tem o acréscimo do matched filter com a finalidade de localizar automaticamente a região vertebral, utilizam a aplicação separada de dois métodos, um de corte em grafos e o outro de crescimento local de volume, para segmentar os ossos trabecular e cortical. No método de corte em grafo cada vértice no grafo representa um voxel no volume e o peso de cara aresta é definido com base na propriedade regional dos segmentos, integrando a combinação linear de gaussianas com Markov Gibbs Random Fields (MGRF). A segmentação propriamente dita consiste em achar o corte de custo mínimo neste grafo, o qual é computado via um algoritmo denominado s-t Min-Cut/Max-flow. O método de crescimento local do volume utiliza um critério global para segmentar os componentes da imagem. O critério utilizado tem como base a intensidade média e o desvio padrão na vizinhança (26 conexo) e o procedimento do algoritmo consiste em verificar cada voxel da superfície externa e se o valor de intensidade for maior que um threshold local então utiliza-se o voxel para iniciar um crescimento de volume local. O framework de segmentação proposto teve sua acurácia medida e foi validado com base em 16 conjuntos de dados de tomografia computadorizada com seus respectivos ground truth e com base no European spine phantom, o qual é um padrão aceito para o controle de qualidade em densitometria de ossos.

3) *Segmentation of the central-chest lymph nodes in 3D {MDCT} images*: No trabalho apresentado por Lu [12], são propostas duas abordagens para segmentação de gânglios linfáticos da parte central do tórax em imagens tridimensionais de tomografia computadorizada. Em ambas as abordagens o usuário fornece uma informação inicial em uma fatia utilizada como referência e partindo desta inicialização as abordagens

realizam a segmentação. Tanto no *single-section live wire* quanto no *single-click live wire* o algoritmo base é o 2D *live wire*, o qual é um algoritmo de detecção de borda que define o problema como um problema de procura ótima em grafo através da análise de contornos ativos. Para começar o processo na fatia utilizada como referência, o usuário seleciona uma semente inicial s_1 situada, grosseiramente, dentro de uma região desejada e então move o mouse para um possível candidato a semente s_2 . O *live wire* utiliza então o algoritmo de *Dijkstra* para procura em grafo e uma função de custo para iterativamente sugerir um caminho ótimo local entre todos os pares de *pixels* entre s_1 e s_2 e se o usuário estiver satisfeito com o caminho sugerido confirma-se o mesmo através do clique do mouse na semente s_2 . Este processo continua até a região desejada estiver definida pelo contorno. Com base neste algoritmo base propõe-se as duas abordagens para segmentação 3D. No *single-section live wire*, o algoritmo base é utilizado para definir o contorno da área desejada na fatia de referência. Este contorno é então projetado para as demais fatias e as sementes iniciais são definidas com base nos *pixels* que constituem a região de contorno. Um conjunto de sementes ajustadas é definido com base nos *pixels* da região definida e o *live wire* itera até convergir ou até alcançar a condição de parada, a qual tem como base a consistência da escala de cinza nas regiões adjacentes dos gânglios linfáticos. Uma grande mudança na distribuição tipicamente indica que a região de interesse desapareceu. O *single-click live wire* possui passos semelhantes o que muda é na definição da região de interesse, de modo que no *single-click* o usuário precisa apenas selecionar um pixel inicial dentro da região de interesse e a região limite vai ser definida automaticamente. As sementes iniciais são definidas pela propagação de raios em várias direções partindo do pixel informado. Os *pixels* com maior gradiente de magnitude ao longo do seu respectivo raio serão os *pixels* semente e o resto do processamento segue o mesmo definido pela primeira abordagem. Os testes foram realizados em imagens de uma base de dados de tomografias de tórax humano, sendo nestes testes medido a acurácia, a reprodutibilidade, o tempo de processamento e a taxa de sucesso.

4) *Fully automatic 3D segmentation of iceball for image-guided cryoablation*: O trabalho apresentado por Liu [13] mostra uma abordagem para segmentação de bolas de gelo em imagens de tomografia computadorizada e de ressonância magnética. A abordagem proposta tem como base a segmentação utilizando corte em grafos incorporando informação a priori sobre formato dos objetos, de forma que cada uma das segmentações das bolas de gelo é inicializada com tendências para o formato modelado utilizando parâmetros derivados experimentalmente, proporcionando assim a separação da bola de gelo em relação as estruturas anatômicas em seu entorno mesmo com valores similares de intensidade. A modelagem do formato a priori é derivado diretamente do modelo de crescimento da bola de gelo, gerando uma série de imagens de máscaras de formato a priori a cada ponto específico do ciclo do procedimento de crioblação. As máscaras são calculadas através de uma função hiperbólica com formato de sino atribuindo os rótulos 1 para pontos pertencentes ao plano

principal e 0 a pontos pertencentes ao plano de fundo e são incorporados ao algoritmo de corte em grafos por meio de termos regionais. O algoritmo de corte em grafos utilizado utiliza a teoria max-flow/min-cut de forma que aplica-se o algoritmo proposto por Boykov e Kolmogory (um algoritmo de aumento de caminho) [14] para encontrar o fluxo máximo o que gera, segundo a teoria, o corte mínimo no grafo. A validação do método proposto se deu com experimentos em 8 pontos de tempo do processo de crioblação em 2 casos, uma de ressonância magnética guiada em tempo real e um de ablações de tumores renais monitorados e os resultados comparados com segmentações manuais segundo coeficiente DICE.

5) *Three-Dimensional Segmentation of Fluid-Associated Abnormalities in Retinal OCT: Probability Constrained Graph-Search-Graph-Cut*: Um método para segmentação de fluídos associados a anormalidades na retina em imagens de Optical Coherence Tomography (OCT) da retina é apresentado por Chen [15]. O método proposto consiste de dois passos principais: inicialização e segmentação. A inicialização é composta da aplicação de alguns passos de pré-processamento sobre a imagem de OCT. Inicialmente as superfícies superior e inferior da retina são determinadas utilizando um algoritmo de segmentação de 11-superfícies. Neste algoritmo a superfície superior corresponde ao limite interno da membrana e a camada inferior corresponde ao epitélio pigmentar da retina. Depois, um fitting de uma superfície sobre a camada inferior é executado e realiza-se uma análise das propriedades de textura e espessura das camadas individualmente com a finalidade de achar "pegadas" da symptomatic exudate-associated derangement (SEAD). Por fim, nesta etapa de pré-processamento, ocorre o achatamento das imagens retiniais de acordo com a superfície thin-plate spline identificada. Após a etapa de pré-processamento vem uma etapa de classificação de voxel pertencente ainda a parte de inicialização. Nesta etapa um segmentação inicial das áreas SEAD, preenchidas com fluídos, é gerada através da aplicação de uma abordagem supervisionada de classificação de voxels treinada com os voxels entre as superfícies superior e inferior da retina previamente segmentadas. Este processo atribui a cada voxel entre as superfícies da retina uma probabilidade entre 0 a 1 que o voxel esta dentro de uma região SEAD. Por fim, para finalizar a parte de inicialização realiza-se uma normalização da probabilidade de pertencer a região SEAD explorando o conhecimento que estas regiões possuem uma distribuição aproximadamente gaussiana. Após esta etapa o mapa de probabilidade normalizado é utilizado para restringir o processo de segmentação com base em grafos. No processo de segmentação dois métodos foram combinados, busca em grafo e corte em grafo, para segmentar as áreas de SEAD formulando assim o problema de segmentação como um problema de minimização de função de energia.

A funcional de energia construída tem como base o custo associado com a segmentações de todas as superfícies, somado ao custo associada as regiões segmentadas e adicionado ao custo associado as restrições entre superfícies e regiões. A função de custo das superfícies utiliza o método de procura em grafo, a função de custo das regiões vem do método

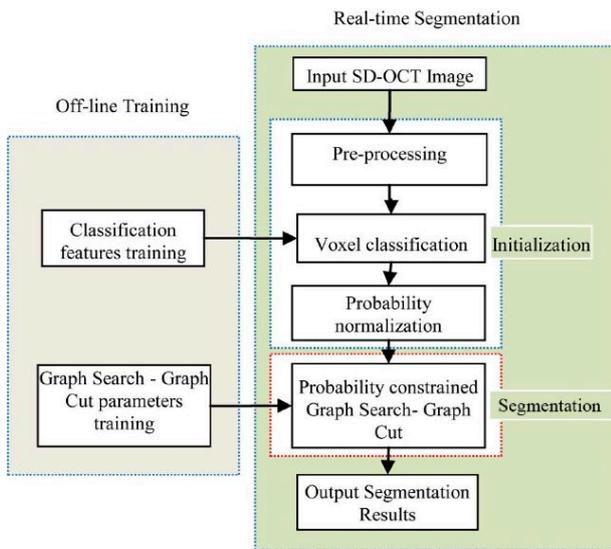


Fig. 1. Fluxograma do método proposto. Extraído de [15]

de corte em grafos e custo das restrições esta atrelado ao posicionamento dos voxels em relação a duas superfícies definidas. Ainda nesta parte a inicialização previamente executada é integrada ao framework pela utilização dos valores de probabilidade, onde voxels com alta probabilidade foram utilizados como sementes fonte e voxels com baixa probabilidade foram utilizados como sementes sink. Com base nestas informações forma-se um grafo s-t, o qual pode ser solucionado com uma técnica de min-cut/max-flow gerando assim o resultado final da segmentação. Uma visão geral do método proposto pode ser visualizado na figura 1. O método proposto foi avaliado em termos de acurácia, medida através dos valores de verdadeiros positivos, falsos positivos e a taxa de diferença relativa do volume. Realizou-se também uma análise estatística de correlação, utilizando regressão linear e Bland-Altman plots, avaliando a relação e a concordância entre as segmentação com método proposto e a manual.

6) *A splice-guided data driven interactive editing*: El-Zehiry [16], apresenta uma abordagem de edição de resultados de segmentação em imagens de tomografia computadorizada e ressonância magnética. Esta ferramenta de edição proposta tem o intuito de corrigir resultados de segmentação tanto 2D quanto 3D através da integração de todos os dados disponíveis (pré-segmentação, dados e entradas do usuário) em um framework de minimização de energia. O algoritmo necessita de uma simples entrada, uma splice 2D e gerando assim atualizações 3D do volume segmentado. A edição por splice é formulada como um problema de minimização de energia onde uma abordagem com base em grafo é utilizada para a otimização dessa energia. Cada voxel esta associado a uma variável binária que representa o rótulo do voxel (se 1 pertence ao objeto de interesse e 0 caso contrário). Cada uma destas variáveis binárias possui um vértice correspondente no grafo e cada aresta é utilizada para ligar vértices vizinhos. O peso de cada aresta reflete os termos da função de energia. Esta

energia possui dois termos concorrentes: o primeiro termo visa a suavidade guiada pelos dados e o segundo termo representa a pré-segmentação. A entrada do usuário é incluída na forma do conjunto de vértices marcados como sementes do primeiro plano e no conjunto de vértices marcados como sementes do plano de fundo (possivelmente vazios). Estas sementes são fornecidas com base em uma interação por splice, a qual utiliza a um conjunto de pontos fornecidos pelo usuário, em uma fatia, para, com base em um teste de cruzamento de raios, calculo de distância e um threshold fornecido, adicionar a semente ao conjunto de sementes do plano de fundo ou do plano principal. Para avaliação do algoritmo proposto uma avaliação quantitativa com 30 conjuntos de dados de ressonância magnética de fígado foi realizada, comparando a acurácia das segmentações editadas com as segmentações realizadas automaticamente. A medida de erro utilizada foi a mean point-to-surface utilizando 2000 pontos, a qual apontou uma melhoria de 15 % no erro médio e de 27% no erro máximo de segmentação.

7) *Left-Ventricle Segmentation of SPECT Images of Rats*:

Um método de segmentação 3D iterativo com base em corte em grafos para segmentação da parede ventricular esquerda em imagens Single Photon Emission Computed Tomography (SPECT) do coração de ratos é apresentado por Jodoín [17]. O método proposto consiste de 3 etapas: a primeira etapa consiste na conversão da imagem de entrada do espaço euclidiano para o espaço spherical-cylindrical (S-C). Esta conversão se da através da identificação do ápice e da base do ventrículo esquerdo, realizada pelo usuário com a inserção de um alvo em T, e do lançamento de raios perpendiculares ao formato de U. Partindo do alvo em T os raios são gerados nas direções radial e cilíndrica no espaço 3D. Cada raio é associado com um ângulo radial, um valor de altura ou um ângulo de elevação. Desta forma cada voxel no espaço euclidiano dentro da área cilíndrica possui um raio, um angulo radial e uma altura ou um angulo de elevação atribuído, resultando em uma imagem 3D S-C com eixos sendo o radial, o ângulo de rotação e a posição de elevação no cilindro ou na parte de hemisfério. O segundo passo é a procura do centro da superfície do miocárdio dentro da imagem 3D S-C. Neste passo foi aplicado o corte em grafos, onde cada vértice esta associado a um voxel da imagem no espaço S-C e esta conectada com os 6 vizinhos mais próximos. Cada peso das arestas foi idealizado de forma que o resultado do corte do grafo corresponde-se aos maiores valores de intensidade, onde é mais provável que a linha central do miocárdio esteja. Por fim, o ultimo passo é localizar as fronteiras do epicárdio e do endocárdio. Para tanto, utiliza-se a linha central calculada anteriormente e estima-se a espessura do miocárdio para cada linha radial através do cálculo da média e da variância de cada linha. Os resultados do método proposto foram comparados segmentações manuais realizadas por 2 especialistas segundo o cálculo do volume sistólico e diastólico do ventrículo esquerdo.

8) *Supervoxels based Graph Cut for Medical Organ Segmentation*: Kitrungrotsakul [18], propõe a um método de segmentação utilizando corte em grafos associado ao uso de *supervoxels*, como estratégia de otimização, para segmentação de órgãos em imagens de tomografia computadorizada. A seg-

mentação utilizando corte em grafos consiste, inicialmente, na representação de cada *voxel* do volume com um nodo do grafo e cada nodo esta conectado com seus nodos vizinhos e sobre esta estrutura busca-se o corte mínimo no grafo necessário para separar o plano de fundo e o primeiro plano. A abordagem proposta faz uma otimização nesta estrutura utilizando a ideia de *supervoxel* que consiste na clusterização dos *voxels* para gerar um conjunto de *supervoxels* e utilizar estes como o nodo do grafo, reduzindo o custo de memória e tempo de computação necessário. Para os experimentos e avaliação dos resultados foram utilizadas 10 tomografias computadorizadas e sobre estas executadas o processo de segmentação. A avaliação quantitativa se deu por comparação dos resultados obtidos com o método proposto com a segmentação realizada por um especialista utilizando o índice DICE como medida.

9) *Automatic segmentation of Computed Tomography (CT) images of domestic pig skeleton using a 3D expansion of Dijkstra's algorithm:* Gangsei [19] apresenta um trabalho para a segmentação e identificação automática de ossos de porcos em imagens de tomografia computadorizada. Para realizar o processo de segmentação e identificação dos ossos, 3 princípios foram aplicados: segmentação pela conectividade; identificação de pontos e linhas; expansão 3D do algoritmo de Dijkstra. Para a segmentação por conectividade foi aplicado um valor de *threshold* para os ossos e o resultado binário foi rotulado. Os objetos conectados dentro da imagem binária foram identificados através do ranqueamento dos volumes ou dos centros de massa. Para identificação dos pontos de referência, projeções 2D do esqueleto foram usadas extensivamente. Depois, ocorre um processo de definição de regiões de interesse com base nos pontos de referência definidos e no conhecimento a priori sobre orientação dos ossos, esta região de interesse deve englobar a região que conecta dois ossos. Dentro da região de interesse o volume é visto como uma pilha 2D de camadas e para cada camada um corte virtual é realizado na "direção do corte". A soma desses cortes constituem a superfície de segmentação. A direção das camadas e a direção do corte variam com a direção principal da superfície de segmentação e foram setadas com base em conhecimento a priori. O primeiro passo do algoritmo de Dijkstra é construir uma matriz de custo, isto é, a matriz que define o custo mínimo acumulado de incluir qualquer *voxel* ao corte virtual. Para a primeira camada, a entrada para a função de custo foi a primeira camada binarizada e para as demais camadas a entrada consiste na soma da matriz de custo da camada anterior com a camada correspondente. O vetor 3D de custos final consiste no vetor de matrizes de custo 2D empilhadas. O próximo passo do algoritmo de Dijkstra é construir o caminho com menor custo, isto é, a superfície de segmentação. O algoritmo começa com a última camada e depois que o caminho em uma camada foi definido, uma região de possíveis caminhos é identificado na próxima camada. Todos os pontos fora dessa região são setados com custo infinito e este processo se repete para todas as camadas subsequentes. O algoritmo apresentado foi aplicado em imagens tomográficas de 485 porcos e os resultados obtidos foram avaliados por inspeção visual.

C. Métodos de Level Set

1) *A Novel Automatic Extraction Method of Lung Texture Tree from {HRCT} Images:* O trabalho apresentado por Jun-Wei [20] propõe um método para a extração da árvore de textura do pulmão em imagens de alta resolução de tomografia computadorizada. Em vista disso, primeiramente mostra-se um modelo implícito de contornos ativos guiado por um modelo de energia denominado Local Binary Fitting (LBF), com parâmetros dinâmicos e modulados pela informação de gradiente da imagem. O LBF tem como base a suposição que a imagem pode ser aproximada localmente por uma imagem binária. Depois, uma técnica de pintura de plano de fundo com base em mapa de intensidade não linear é utilizado para remover a influência do plano de fundo durante a evolução da função de *level set*. A técnica de remoção de plano de fundo tem como base a teoria do sistema de visão humana, na qual há o indício que o reconhecimento de objetos não depende apenas da saliência dos objetos, mas também das características de saliência entre objeto e plano de fundo de tal forma que detalhes internos de um objeto podem ser aprimorados através da redução do contraste entre objeto e plano de fundo. Por fim, foi realizada uma comparação do método proposto com métodos que utilizam modelos com base em borda e região em uma base de dados de tomografias de pulmões, na qual a abordagem proposta mostrou maior efetividade no processo de remoção da textura.

2) *A variational framework for 3D colonic polyp visualization in virtual colonoscopy:* Um algoritmo de segmentação 3D para segmentação do tecido do cólon em imagens de tomografia computadorizada e o posterior uso do resultado para navegação é apresentado por Chen [21] [22]. O método proposto inicia com a remoção do líquido opacificado através de um método de *threshold* e com equalização da intensidade da imagem dentro do cólon. Depois, 4 sementes iniciais são colocadas inteiramente dentro da região do cólon. Então, os contornos fechados centrados nas sementes iniciais são propagados através das fronteiras da região 3D desejada com a evolução iterativa das funções adaptativas da técnica adaptativa de *level set*. Durante cada iteração a informação em cada região é estimada pelos parâmetros da função densidade de probabilidade. No trabalho apresentado foi utilizado uma distribuição gaussiana para os pontos pertencentes a ambas as classes (cólon e não-cólon). Com o resultado da segmentação, gera-se a *iso-surface* através do algoritmo de *marching cube* e por fim extrai-se a linha central para navegação através da solução de uma equação diferencial. O algoritmo foi testado em um conjunto de 22 tomografias computadorizadas de colonografias com diversas patologias e foi validado medindo sua acurácia através do cálculo de sobreposição entre os resultados obtidos pelo método proposto e a segmentação realizada manualmente por um especialista.

3) *Prior-Based Automatic Segmentation of the Carotid Artery Lumen in TOF MRA (PASCAL):* O estudo apresentado por Hutter [23] mostra um algoritmo de segmentação 3D totalmente automático em imagens de angiografias por ressonância magnética sem contraste, o qual utiliza conhecimentos a priori sobre padrões irregulares de fluxo sanguíneo

na área de bifurcação da carótida para resolver os problemas de suscetibilidade do vaso afetado por tal fluxo. O algoritmo proposto é dividido em 3 etapas: a primeira etapa inicia com a geração da projeção de máxima intensidade nos eixos sagital e transversal das fatias provenientes do escaneamento. Depois, as regiões de interesse são identificadas através de um *threshold* adaptativo e regiões ruidosas são eliminadas da procura por candidatos através da procura por áreas conectadas na direção das fatias. A segunda etapa é a detecção do esqueleto da árvore do vaso e da bifurcação, para tanto é diferenciada a carótida da artéria vertebral, nas primeiras fatias da cauda, através da pré-segmentação pelo tamanho e conectividade com a bifurcação. O esqueleto consiste da localização de pontos centrais e das fatias de bifurcação e é extraída da seguinte forma: o resultado do *threshold* da fatia anterior ($k-1$) é analisada utilizando o ajuste de elipsoides. Isto fornece o eixo maior $e1_{k-1,v}$, o eixo menor $e2_{k-1,v}$, a área $a_{k-1,v}$ e o centro $m_{k-1,v}$ para cada segmento de vaso detectado, onde $v = 0$ indica o maior segmento. *Voxels* no eixo maior são utilizados como sementes para a segmentação por crescimento de regiões na fatia k . Para cada fatia é calculada uma função $f(k)$ e a bifurcação é determinada como a fatia com maior $f(k)$. Por ultimo, vem a aplicação da evolução do método 3D de *level set*, o qual tem como base a minimização de funções. Os resultados obtidos da aplicação do método proposto são provenientes de testes em 11 conjuntos de imagens de voluntários sendo estes avaliados de forma quantitativa.

4) *3D Segmentation of the Lung Based on the Neighbor Information and Curvature*: O trabalho apresentado por Qi [24] mostra um método para segmentação automática de pulmões em imagens de tomografia computadorizada. Para tanto é utilizado um *framework* de estimativa denominado *Maximum a posteriori* (MAP), o qual combina informações prévias de vizinhança e informações da imagem em tons de cinza para extrair as bordas do pulmão. Para construir o modelo de informações prévias de vizinhança utilizou-se uma análise de componentes principais em um conjunto de imagens de treinamento. Este modelo é formulado em termos de *level set functions* e as superfícies evoluem de acordo com a equação de Euler-Lagrange associada. Por fim, é utilizado um algoritmo para refinar as bordas grosseiramente geradas pelo MAP, o qual consiste de dois passos: o primeiro passo é a detecção automática e adaptação grosseira da região do hilo pulmonar; o segundo passo é refinar esta adaptação grosseira com base na combinação de informações de probabilidade e curvatura. Experimentos foram conduzidos em imagens reais de pulmões de um hospital da china.

5) *Semi-automatic 3D segmentation of carotid lumen in contrast-enhanced computed tomography angiography images*: Hemmat [25] apresenta um método de segmentação do lúmen da artéria carótida em imagens de Tomografia Computadorizada de Angiografia Contrastada (TCAC). Inicialmente tem-se a seleção de uma região de interesse para evitar computações desnecessárias e a utilização do *mean shift* para aumentar a homogeneidade da intensidade da região do lúmen ao longo do caminho do vaso sem qualquer dano as bordas das estruturas. Depois do *mean shift*, objetos muito grandes ou muito pequenos são removidos com operadores

morfológicos e os buracos são preenchidos. No próximo passo, com a ajuda de 3 sementes informadas, são extraídas as linhas centrais utilizando um algoritmo de marcha rápida em menor caminho com o auxílio de matrizes hessianas 3D. Por fim, utiliza-se as linhas centrais associadas a um método 3D de *level set function* para realizar a segmentação do lúmen da carótida. A validação do método proposto foi realizada através da aplicação do mesmo em 14 volumes de TCAC e da subsequente comparação dos resultados obtidos com um *Ground Truth* (GT) gerado por especialistas. A comparação com o GT se deu com as seguintes métricas: número de *voxels*; sobreposição dos volumes (*jaccard* e porcentagem de correspondência); medida de concordância entre dados *DICE similarity*;

D. Método com base em modelo

1) Markov Random Fields:

a) *A tightly coupled region-shape framework for 3D medical image segmentation*: Huang [26], apresenta um *framework* para segmentação de imagens médicas de ressonância magnética. Este *framework* híbrido combina 3D Markov Random Fields (MRF), para a parte de modelagem restrições de rotulamento com base em região, com um modelo deformável, para a parte de restrições com base em formato. No 3D MRF cada voxel do volume é um nodo observável o qual é conectado com um nodo escondido representando o rótulo da região do *voxel*. Cada nodo escondido é também conectado com seis ou mais nodos vizinhos escondidos. A segmentação pode ser vista então como um problema de estimar a solução, maximize-a-posterior (MAP), do MRF utilizando uma função de compatibilidade entre nodos escondidos e uma função de *Expectation Maximization* (EM). Para integrar a parte de modelos deformáveis, atrela-se cada rótulo da região a superfície 3D do objeto utilizando assim uma teoria de modelo gráfico para juntar rigidamente as duas abordagens. Modelos deformáveis podem ser vistos frequentemente em *frameworks* probabilísticos, nos quais a energia interna leva a uma distribuição a priori enquanto a energia externa pode ser convertida para um modelo sensor com probabilidade condicional. Ao em vez de condicionar a imagem diretamente ao modelo deformável, assume-se que o condicionamento é indireto através das restrições de rotulamento de região. Com base nesta forma de acoplamento dos dois modelos utiliza uma função Belief Propagation (BP) para fazer a inferência na parte do MRF e *Finite-Element Method* (FEM) para inferir na parte dos modelos deformáveis, suavizando assim as fronteiras grosseiramente estimadas com o MRF e introduzindo um formato a priori no modelo híbrido proposto. Os testes realizados com o *framework* híbrido proposto foram, inicialmente, em imagens sintéticas com um objeto conhecido e em imagens médicas de ressonância magnética de uma cabeça com o crânio parcialmente removido com a finalidade de expor o cérebro. A avaliação se deu de forma qualitativa com a opinião de especialistas.

b) *Model Based 3D Segmentation and OCT Image Undistortion of Percutaneous Implants*: Uma forma automática de segmentação 3D para implantes percutâneos é proposto por Müller [27]. Com base em imagens geradas através

de uma técnica não invasiva, *Optical Coherence Tomography* (OCT), o autor propõe uma abordagem de segmentação que utiliza *Markov Random fields* para a estimativa da superfície da pele no espaço 3D e, para a segmentação do implante, é utilizado um modelo de distorção de refração associada a superfície da pele previamente segmentada. Para a análise morfométrica de imagens de OCT existe a necessidade de realizar a correção da distorção óptica causada pelo índice de refração do tecido. Esta correção é realizada utilizando os implantes subcutâneos para a estimativa, com base em modelo, do índice de refração utilizando uma transformação de *Hough* generalizada. Vários experimentos foram realizados no modelo não distorcido incluindo uma avaliação quantitativa com a finalidade de mostrar a competitividade da proposta contra a segmentação realizada manualmente por especialistas.

c) *3D Segmentation in CT Imagery with Conditional Random Fields and Histograms of Oriented Gradients*: Bhole [28] aborda o problema de segmentação 3D em imagens tomográficas de abdomens para identificar os órgãos nesta região. Para tal é proposto e avaliado diferentes modelos e estratégias de modelagem para segmentação 3D com base em *Markov Random Fields* (MRF) e seu homólogo discriminativo conhecido como *Conditional Random Fields* (CRF). No trabalho em questão, avalia-se também a utilidade das características com base em histogramas da orientação do gradiente ou características de HOG. Com base nas imagens de tomografia computadorizada uma grade 3D de MRF e CRF é construído. Utiliza-se Variational Message Passing (VMP) para a aprendizagem e a inferência por *max-product* (MP) para a predição do modelo. Estas abordagens para inferência e aprendizagem permite aprender termos emparelhados em cadeias randômicas que não são sub-modulares, sendo portanto muito flexíveis. Os experimentos foram realizados com foco em órgãos abdominais e segmentação de regiões e o método proposto foi avaliado em um conjunto de estruturas anatômicas encontradas dentro de uma base de dados de imagens de fígado.

2) Modelos deformáveis:

a) *Assessment of Airway Remodeling in Asthma: Volumetric Versus Surface Quantification Approaches*: O trabalho apresentado por Saragaglia [29], desenvolve uma abordagem para quantificação volumétrica das paredes das vias aéreas em imagens geradas por um tomógrafo computadorizado com multi-detecores, explorando uma metodologia de segmentação 3D com base em um modelo de malha deformável específica do paciente. O método proposto inicia a partir da segmentação 3D do lúmen da via aérea obtido pelo método utilizado no estudo comparativo, o qual realiza a quantificação bronquial através de uma análise 2D, com um método de segmentação que combina morfologia matemática e pareamento de contornos com base em energia, aplicado à cada fatia do volume de forma independente. Com base nesta segmentação inicial um modelo de malha da parede interna da via aérea é construída. Com foco de preservar a topologia e a geometria dos pequenos brônquios envolvidos no processo de quantificação e para obter uma malha regular, uma abordagem restrita da triangulação de Delaunay é aplicada com um critério específico e adaptativo de distância. A

superfície externa da parede dos brônquios é então segmentada através da deformação do modelo de malha da parede interna, condicionalmente aos dados da imagem e as restrições de formato. Uma equação é atribuída para cada vértice da malha permitindo-a mover em um campo de força governado por forças internas e externas resultando assim em um modelo deformável da malha o que permite uma segmentação mais precisa da região. A abordagem proposta foi avaliada utilizando imagens sintéticas e dados reais do tomógrafo e um estudo comparativo com a técnica de quantificação de superfícies 2D/3D desenvolvida foi conduzido.

b) *3D Segmentation of the Liver Using Free-form Deformation Based on Boosting and Deformation Gradients*:

A abordagem apresentada por Zhang [30], visa utilizar a segmentação 3D para alcançar uma segmentação confiável do fígado em imagens de tomografia computadorizada. O método proposto combina conhecimento a priori de intensidade e formato para gerar forças as quais são aplicadas a um modelo deformável. Para tanto, o processo inicia com a codificação do volume em uma *octree* com base na entropia dos dados. Uma classificação dos nodos na *octree* atribui cada nodo em um dos dois grupos: dentro ou fora do fígado. Este processo estima grosseiramente a localização do fígado no volume. Após a inicialização, o dado de profundidade do gradiente de deformação para o formato médio é calculado para cada vértice. Se o dado de profundidade é menor que um *threshold* a força do formato é aplicada, caso contrário, a força da intensidade é aplicada. Após o cálculo de força para cada vértice, todo o campo de força em cada vértice da malha é suavizada por um peso médio dentro de um anel centrado no vértice em questão. Então uma deformação de formato livre com 27 pontos de controle é aplicada com base nas forças calculadas e a iteração continua até alcançar uma estabilidade. Todo o cálculo de forças, extração de características e de refinamento de bordas é explicado com mais detalhes. Os testes para o método proposto foram realizados em 10 casos reais de dados de tomografia computadorizada e foram validados através de um padrão ouro realizado por um especialista em radiologia, o qual realizou a segmentação manual do fígado para os 10 casos.

c) *3D segmentation of soft organs by flipping-free mesh deformation*: Ding [31], apresenta um trabalho sobre segmentação de órgãos em imagens médicas de tomografia computadorizada. Para tanto é proposto um algoritmo de segmentação 3D com base em modelos deformáveis para segmentar órgãos moles tais como a bacia e o fígado. O algoritmo com base em modelos deformáveis segmenta o volume de entrada através da deformação iterativa de um modelo de malha 3D para registrá-lo para a extração de características da imagem. Em cada iteração, o algoritmo procura por uma possível correspondência entre os vértices da malha e as características da imagem em longas distâncias. A correspondência detectada é refinada antes da deformação para evitar giros. Para a parte de deformação utiliza-se o método laplaciano devido a facilidade de incorporar restrições geométricas. Em suma, o algoritmo pode ser resumido em 5 passos: Extração das características da imagem, procura por correspondência, detecção de possíveis giros, prevenção de possíveis giros e deformação da malha. O

algoritmo proposto foi testado em 8 conjuntos de tomografia computadorizada e comparado contra outros algoritmos (GVF snake, level set e corte em grafo) em termos de resistência a ruído, acurácia e eficiência com base em um ground truth.

d) *Automatic Detection of Lung Nodules in CT Datasets Based on Stable 3D Mass-spring Models*: Um sistema de auxílio a detecção para encontrar nódulos pulmonares em tomografias computadorizadas é proposto por Cascio [32]. Para tanto, o sistema como um todo consiste de 6 etapas que são: interpolação isotrópica, identificação do volume do parênquima pulmonar, escolha das sementes, detecção e segmentação anatômica das regiões de interesse através do 3D *Mass-Spring Model* (MSM), extração de características para cada região de interesse e a execução de um classificador com base em redes neurais. A interpolação isotrópica é utilizada com a finalidade de reduzir possíveis erros devido a representação anisotrópica dos *grids*. A identificação do volume do parênquima pulmonar, realizado com o uso de um algoritmo 3D de crescimento de regiões e um processo morfológico de dilatação, é uma tarefa crucial que tem por finalidade definir a região de interesse reduzindo assim o tempo computacional e o número de falsos positivos. Depois para a escolha das sementes é realizado uma subtração de *voxels* envolvendo os resultados binários da dilatação e do processo de crescimento de regiões e a posterior multiplicação da subtração com a imagem de entrada o que resulta em uma matriz que será utilizada para a identificação das estruturas internas do volume pulmonar através de uma procura de valores com intensidade máxima. Este detector de picos é aplicado para gerar a lista de sementes. Com a lista de sementes calculada, inicia-se o processo de segmentação das regiões anatômicas através do MSM, o qual com modelos com base na física da iteração de uma rede de pontos de massa interconectados por molas elásticas com o propósito de descrever o formato dos objetos, visa realizar a segmentação e a descrição dos objetos anatômicos com base neste modelo deformável. Todo o processo de inicialização, cálculo de energia interna e externa, representação e evolução do modelo é explicado com detalhes. Uma vez que o formato e o volume dos nódulos são obtidos utiliza-se a extração de características para fins de utilização no processo de classificação dos nódulos, visando a redução do número de falsos positivos. Os resultados obtidos foram comparados com outros métodos através da aplicação da abordagem proposta sobre a mesma base de dados, comparando-os com base na eficiência e no número de falsos positivos.

e) *An Intelligent Tool for Anatomical Object Segmentation Using Deformable Surfaces*: Um método de segmentação de objetos em imagens tridimensionais com intervenção inicial do usuário é apresentado por Delibasis [33]. O trabalho descreve abordagem proposta para implementação do algoritmo de superfícies ativas/deformáveis, utilizando um esquema explícito para a evolução da equação do modelo de superfícies ativas. Este modelo é construído em uma malha retangular o que permite operações aritméticas simples para calcular derivadas parciais. Para calcular as forças externas, utilizadas no cálculo de evolução da equação, ao modelo foi utilizado *Vector Field Convolution* (VFC) o qual é gerado utilizando um *kernel* K contendo vetores que apontam através

da origem de K. Os resultados da utilização do método foram apresentados através da aplicação em dados sintéticos 3D e imagens de tomografias computadorizadas de objetos anatômicos e a avaliação quantitativa e qualitativa destes mostram que o método proposto é capaz modelar estruturas anatômicas de forma precisa requerindo apenas uma limitada intervenção do usuário.

f) *Greedy algorithm based deformable simplex meshes using gradient vector flow as external energy*: O estudo apresentado por Shi [34] mostra uma abordagem para segmentação 3D de imagens médicas de tomografia computadorizada. Esta abordagem combina *deformable simplex meshes*, um algoritmo guloso e o Generalized gradient vector flow (GGVF) para realizar a segmentação com base em modelos deformáveis. Inicialmente, modela-se a superfície do objeto com o *deformable simplex meshes* e com base nas informações geométricas se calcula a energia interna necessária para a parte de deformação. Depois, utiliza-se um algoritmo guloso como método de evolução para guiar o *deformable simplex meshes* para o objeto de interesse. Em cada iteração procura-se uma janela cúbica sobre um vértice da malha e nesta janela computa-se a energia para cada vértice. A energia é calculada com base na energia interna, definida pela geometria da malha, e na energia externa calculada pela aplicação do GGVF, o qual utiliza uma fórmula variacional para propagar os vetores de gradiente de uma imagem de entrada em regiões homogêneas através de um processo de otimização iterativo. A deformação da malha é repetida até que o movimento dos vértices em uma iteração seja menor que um valor determinado. Os testes da abordagem proposta foram realizados em conjuntos de dados sintéticos e reais. A avaliação quantitativa se deu com a aplicação do Median Radial Error (MRE) sobre os resultados de segmentação em comparação com o *ground truth*.

3) Contornos ativos:

a) *Automatic 3D segmentation of CT images based on active contour models*: Uma abordagem para segmentação do fígado em imagens de tomografia computadorizada é apresentada por JIANG [35]. No método proposto, o primeiro passo consiste da aplicação de um threshold global para separar os segmentos do fígado, bacia, estômago, entre outros do plano de fundo. Com o resultado binário obtido, aplica-se uma operação morfológica de erosão para remover pequenos tecidos e a operação de dilatação para manter a forma original do fígado. Como próximo passo tem-se a utilização do operador sobel para a identificação das bordas da imagem resultante do passo anterior. Os pontos da borda detectada devem ser utilizados em uma direção específica de modo a iniciar o algoritmo de contornos ativos. Utilizando a borda como uma curva fechada e contínua para a inicialização do algoritmo de contornos ativos obteve-se o contorno ativo do fígado para a fatia em questão. Para as demais fatias utiliza-se o contorno ativo encontrado na fatia vizinha como inicialização encontrando assim, para cada fatia, os contornos ativos subsequentes e, conseqüentemente, a segmentação do fígado é realizada. Os experimentos para a abordagem proposta foram realizados em 5 conjuntos de tomografias computadorizadas de abdômen e a validação se deu através da comparação da área do fígado obtida pelo método automático com a área segmentada

manualmente por especialistas.

b) *B-Spline Explicit Active Surfaces: An Efficient Framework for Real-Time 3-D Region-Based Segmentation*: Barbosa [36] propõe um aprimoramento, utilizando b-splines, de um framework de segmentação com base em contornos ativos para segmentação de imagens de tomografia computadorizada e ultrassom. O aprimoramento proposto parte do framework apresentado por Duan [37], o qual apresenta o volume através de uma interface de representação explícita denominada active geometric functions (AGF). A ideia principal deste framework é modelar interface como uma função explícita o que implica que geometricamente uma das coordenadas dos pontos da interface são expressas como uma função das coordenadas restantes. O aprimoramento proposto expressa as funções explícitas como uma combinação linear de funções b-spline base e o problema de segmentar um objeto do plano de fundo torna-se um problema de minimização de uma função de energia com respeito aos coeficientes da b-spline, levando em conta características locais e globais do volume. O método em questão pode ser dividido em duas etapas: inicialização e o processo de minimização da função de energia. Na inicialização tem-se toda a definição de sistema de coordenada a ser utilizado, definição da origem do sistema de coordenadas escolhido, inicialização de variáveis e parâmetros. Na parte de minimização tem-se todo o cálculo de características da imagem, a computação do gradiente de energia e atualização dos coeficientes b-spline e a re-amostragem dos pontos de interface buscando a redução do valor de energia. Visando avaliar o aprimoramento proposto vários experimentos foram realizados em dados sintéticos e reais. Inicialmente teve-se a comparação com o framework base proposto por Duan em termos de acurácia (medido com *the root mean square error*) e tempo de execução. Depois, realizou-se uma comparação com um método de level set comparando em termos de coeficiente DICE em relação a uma segmentação manual. Por fim, utilizou-se o método na segmentação de tomografias computadorizadas de tumores no fígado, na qual foi realizado uma comparação com o framework base em termos de erro de sobreposição, diferença de volume, distâncias média e máxima e tempo de execução.

c) *Forensic-Case Analysis: From 3D Imaging to Interactive Visualization*: Urschler [38], propõe um framework para a análise de casos forenses, no qual um dos algoritmos básicos é a segmentação. O algoritmo de segmentação utilizado é um algoritmo interativo de segmentação de primeiro plano/plano de fundo formulado como uma minimização de energia. Denominado como Geodesic Active Countour (GAC), é um modelo que separa os planos, primeiro e de fundo, através da definição de uma hiper-superfície na divisa entre os planos. O GAC pode incluir conhecimento a priori (distribuição do valor de cinza, textura e restrições de formato) e, no trabalho apresentado, foi utilizado uma formulação contínua de minimização de energia em um framework variacional. Para tanto, dado uma imagem, busca-se a representação binária da mesma dividindo-a entre plano de fundo e primeiro plano, a qual é alcançada através da resolução da minimização pela derivação e resolução das equações de Euler-lagrange associadas ao modelo. Os resultados da aplicação da abordagem proposta foram obtidos através

da aplicação do *framework* em dois casos de análise forense, um com imagens de tomografia computadorizada e o outro com imagens de ressonância magnética.

d) *High resolution volume quantification of the knee joint space based on a semi-automatic segmentation of computed tomography images*: Mezlini [39] apresenta um método de segmentação para medição das variações locais 3D do espaço articular nos joelhos em imagens de tomografia computadorizada com alta resolução periférica. A abordagem proposta divide-se em uma série de etapas: a primeira etapa consiste na reorientação manual das imagens do joelho devido a variedade de orientações presentes nas imagens. Depois, tem-se a aplicação de uma etapa de pré-processamento nas imagens transversais, onde ocorre a aplicação de um filtro de média, para suavizar a imagem e aprimorar as bordas. A aplicação deste filtro com um disco de 8 pixels permite encontrar um estável quantil de níveis de cinza separando o osso de tecidos suaves. Este quantil é utilizado como valor de threshold para binarizar a imagem. Sobre a imagem binarizada aplicam-se operações de abertura e fechamento para reconectar as regiões de osso, preencher buracos e eliminar resíduos de tecidos suaves. Esta etapa gera uma máscara das partes que contém osso. A próxima etapa consiste na segmentação 3D das imagens coronais, a qual consiste na geração de uma máscara 3D com base em um hysteresis threshold 3D e operadores 3D de fechamento para eliminar tecidos residuais, ruídos e extrair o volume ósseo. Como próximos passos tem-se a seleção de um volume de interesse, realizada manualmente, e o posterior rotulamento das diferentes partes do osso. Por fim ocorre a extração do espaço articular utilizando uma técnica de contornos ativos e a medição do volume do espaço articular. O método proposto foi validado com base na reprodutibilidade de diferentes operações manuais realizadas por dois usuários em 5 conjuntos de dados de joelhos e a confiabilidade do método se deu pelo cálculo do root mean square coefficient variation. A eficiência e robustez foram demonstradas com base na aplicação do método em diferentes imagens de tomografia computadorizada com diferentes resoluções e geometrias de raios.

4) *Modelo com base em formato*:

a) *3D segmentation of cell boundaries from whole cell cryogenic electron tomography volumes*: Moussavi [40] propõe um algoritmo recursivo denominado BLASTED (Boundary Localization using Adaptive Shape and Texture Discovery) para extrair automaticamente os limites da célula utilizando um *Conditional Random Field (CRF)*, framework no qual pontos de fronteira e formato são inferidos em conjunto, em imagens de *cryon-electron tomography*(cyro-ET). A segmentação da célula foi realizada através: do tratamento da textura das fronteiras como uma variável não localmente aprendida; da proposta de pontos candidatos utilizando características físicas que dependem do formato estimado bem como a aprendida textura da fronteira; e da classificação robusta dos pontos candidatos utilizando a textura aprendida com contexto de formato local e global. Para tanto, o processo inicia com um fatia rotulada manualmente visando encontrar todos os subseqüente contornos e conjunto de pontos que coletivamente estão na membrana da célula. Com esse con-

torno inicial é previsto na próxima fatia os pontos candidatos que podem pertencer ao contorno, com base em um *template* atualizado a cada passo, e sobre esse pontos candidatos aplica-se um processo de inferência que classifica qual desses pontos estão no contorno. A parte da recursividade do algoritmo vem da partição da inferência 3D da fronteira da célula em uma organização 2D fatia por fatia recursiva. O Método proposto foi comparado com outros algoritmos e testado em um conjunto de 20 cyro-ET, utilizando duas métricas com base no *ground truth* das imagens utilizadas.

b) *A Model-Based Validation Scheme for Organ Segmentation in CT Scan Volumes:* Badakhshannoory [41] propõe uma abordagem de segmentação 3D de órgãos em imagens de tomografia computadorizada. Para tanto, um algoritmo com base em modelo é utilizado. O algoritmo possui duas fases (treinamento e teste). Na fase de treinamento, utilizando os *ground truth* fornecidos por um conjunto de dados de treinamento, um espaço é construído, o qual descreve o órgão. O espaço do órgão é um grupo de autovetores que são gerados através da aplicação do método de Principal Component Analysis (PCA) sobre o conjunto de dados de treinamento. Na parte de teste, quatro processo principais são executados : o primeiro passo é o pré-processamento. Nele ocorre um alinhamento do volume tomográfico e a estimativa de um ponto de início, o qual depende de características únicas do órgão de interesse e do conhecimento geral do órgão. O segundo passo é a geração de possíveis candidatos, a qual ocorre através da segmentação de cada fatia em um número grande de regiões sobrepostas entre as quais uma região poderia ser a região verdadeira correspondendo ao órgão de interesse em uma fatia específica. O terceiro passo é o refinamento dos possíveis candidatos. Nesta etapa ocorre a redução do espaço de procura por meio da utilização de restrições que incorporam relacionamentos entre regiões dos órgãos nas fatias consecutivas do volume. Por fim, o ultimo passo é a seleção dos melhores candidatos. Este procedimento mede a qualidade de todos os candidatos gerados de acordo com o espaço do órgão gerado na fase de treinamento e identifica o melhor candidato por meio da medição da similaridade do candidato e o órgão. A figura 2 mostra um fluxograma da abordagem proposta. A validação e teste da abordagem proposta se deu no conjunto de dados de fígado, nos quais o espaço do órgão foi gerado através de máscaras 3D dos volumes de fígado do conjunto de 20 volumes de treinamento do MICCAI's 2007 grand challenge workshop. Para a parte de avaliação dos resultados foram utilizadas medidas de sensibilidade, erro (Mean-square) e similaridade (Dice coefficient) com base em padrões ouro.

c) *Towards Robust and Effective Shape Prior Modeling: Sparse Shape Composition:* O estudo apresentado por Zhang [42], abrange alguns desafios existentes na inferência e no refinamento do formato dos objetos na imagem. Visando solucionar esses desafios o autor propõe um *framework* para modelagem denominado Sparse Shape Composition (SSC). Neste *framework* é selecionado um conjunto de formatos esparsos de um repositório de formatos, compondo-os em um só modelo com a finalidade de inferir e refinar um formato fornecido como entrada. Este modelo é formulado como um problema de aprendizagem utilizando *L1 norm*

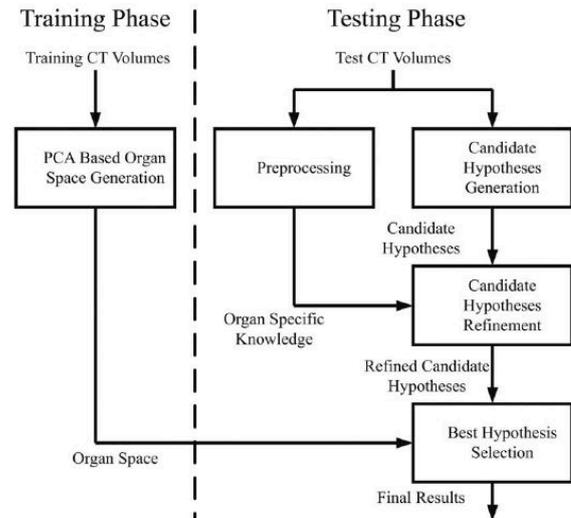


Fig. 2. Diagrama de fluxo do método proposto. Extraído de[41]

relaxation e pode ser solucionado de forma eficiente por um método de *Expectation-Maximization* (EM). O método proposto é independente da dimensão dos dados, ou seja, funciona tanto para contornos 2D quanto para malhas 3D. Na parte de experimentos, o método foi validado em algumas aplicações de imagens médicas tais como: localização 2D do pulmão em imagens de raio-x do tórax, segmentação 3D do fígado em escaneamentos de tomografia computadorizada de baixa dosagem, segmentação 3D de múltiplas estruturas do cérebro de roedores em imagens de ressonância magnética, rastreamento em tempo real do ventrículo esquerdo em imagens de ressonância magnética e reconstrução de tomografias computadorizadas em alta resolução. Netas aplicações o SSC mostrou maior robustez e exatidão comparado com algumas abordagens amplamente utilizadas.

d) *Statistical Modeling of the Eye for Multimodal Treatment Planning for External Beam Radiation Therapy of Intraocular Tumors:* Rügsegger [43] apresenta uma modelagem estatística de formato para modelagem precisa do olho para tratamento de tumores intraoculares. Para realizar a modelagem, inicialmente é construído um atlas representado o formato médio contendo apenas o contorno através dos formatos de treinamento. Este atlas é marcado com pontos de referência antes de seus pontos serem propagados para cada formato de treinamento. Com base nos pontos de referência uma modelagem estatística do formato pode ser computado utilizando uma análise de componentes principais. Para automaticamente encaixar o modelo do olho em novas pilhas de imagens estende-se o modelo para utilizar uma técnica de *Active Shape Model* (ASM), a qual possibilita a correspondência automática do modelo estatístico em outras modalidades de imagem de uma forma robusta e exata. Com base no modelo construído inicia-se um processo iterativo de procura do melhor encaixe dos pontos de perfil do modelo com os pontos de perfil das imagens dos pacientes através

da minimização da distância de mahalanobis com o perfil do modelo. Depois de um número confiável de casamentos entre pontos foi encontrado para cada ponto de perfil o modelo foi encaixado aos pontos o mais próximo possível. Os resultados das segmentações manual e automática foram comparadas em tomografias computadorizadas de 17 pacientes medindo o erro médio entre as segmentações.

E. Crescimento de região

1) *Assessment of 18F {PET} signals for automatic target volume definition in radiotherapy treatment planning:* Davis [44], propõe um método com base em modelo para determinar um *threshold* relativo para delineamento grosseiro do volume de um tumor em imagens de *Positron Emission Tomography* (PET). O método proposto começa com a definição de um *threshold* absoluto. Este é definido com base em pontos pertencentes a região de interesse e ao plano de fundo. O processo como todo inicia com a seleção manual de um ponto dentro da região alvo e um segundo ponto pertencente ao plano de fundo. Com base nesses dois pontos estima-se um valor de *threshold* absoluto. Um algoritmo de crescimento de regiões começa do primeiro ponto e produz um volume alvo preliminar com base na estimativa do *threshold* absoluto, entretanto neste processo preliminar pode-se ter incluído estruturas não alvo. Para descartar estas estruturas um terceiro ponto é informado dentro uma estrutura não alvo. Dois algoritmos de crescimento de região iniciam dos pontos de sinal máximo, calculado anteriormente, e do ponto recentemente informado. As duas regiões crescem simultaneamente e o algoritmo para antes que as duas regiões se juntem em uma única região. A separação das duas regiões é realizada utilizando um algoritmo de *Voronoi*. Apenas os *voxels* que estão dentro da região alvo e encontram-se dentro da metade do *full-width at half-maximum* da resolução da imagem PET são incluídos na estimativa final do verdadeiro nível de sinal máximo. Um processo similar é adotado para a estimativa final do sinal médio do plano de fundo. Então calcula-se, com base nos dois valores obtidos, o *threshold* absoluto final utilizado para a determinação do volume alvo. A validação foi realizada através da submissão de imagens PET de esferas com diâmetro conhecido para segmentação automática dos volumes. Adicionalmente, imagens de pacientes com câncer de mama e de timo foram utilizadas.

2) *Automatic rib segmentation and labeling in computed tomography scans using a general framework for detection, recognition and segmentation of objects in volumetric data:* Um estudo para segmentação e rotulação automática da caixa torácica em imagens de tomografia computadorizada do tórax é apresentado por Staal [45]. O sistema proposto consiste de 5 estágios: Detecção de estruturas relevantes; construção de primitivas da imagem; classificação das primitivas; agrupamento e reconhecimento das primitivas classificadas; e segmentação com base nos grupos obtidos. O primeiro estágio consiste na aplicação de um *threshold* obtendo uma imagem binária e na aplicação de uma matriz hessiana com valores obtidos da aplicação de um *kernel* gaussiano sobre a imagem. Esse processo resulta na detecção de estruturas 1D rígidas. Estas estruturas 1D rígidas são utilizadas na etapa de

construção das primitivas em forma de linhas. Um processo de crescimento é utilizado, o qual leva em conta a orientação local das estruturas rígidas. A vizinhança em volta dos *voxels* rígidos é investigada e outros *voxels* rígidos na vizinhança que cumprem certos requerimentos são adicionados ao primitivo. As primitivas que cumprem as restrições são denominados conjuntos convexos. As sementes escolhidas para o processo de crescimento são escolhidas randomicamente dos *voxels* rígidos disponíveis. Depois da extração de um conjunto convexo uma nova semente é escolhida dos *voxels* rígidos remanescentes. Para evitar o crescimento de um conjunto convexo em outro, um *threshold* pode ser setado no tamanho máximo que um conjunto convexo pode atingir. A próxima etapa é a classificação das primitivas, nesta é utilizado um classificador para diferenciar as costelas de outras primitivas. Um classificador mapeia um vetor de características extraídos dos objetos sobre investigação para um número de classe ou para um vetor de probabilidade de pertencer a aquela classe. Este mapeamento deve ser aprendido através de dados de exemplo, onde estes exemplos foram rotulados para as classes apropriadas. Como quarto estágio tem-se o agrupamento de primitivas. O objetivo desta etapa é agrupar as primitivas da costela em linhas centrais de costelas específicas e rotular as linhas centrais em lado da costela e número. Por fim, para completar o processo de segmentação, utiliza-se um algoritmo de crescimento de regiões com base em sementes, onde estas sementes são determinadas com base nos valores médios de tom de cinza de cada linha central. O método foi testado em 20 escaneamentos de tomografia computadorizada e os resultados foram avaliados qualitativamente.

3) *Defining and Computing Stable Representations of Volume Shapes from Discrete Trace Using Volume Primitives: Application to 3D Image Analysis in Soil Science:* O trabalho proposto por Monga [46] tem como propósito descrever imagens de volumes como um conjunto de remendos de volumes primitivos com a finalidade de otimizar um critério assegurando a sua estabilidade e incluindo uma caracterização de sua escala. Esta otimização se dá com um algoritmo de crescimento de regiões aplicado a um grafo de adjacência o qual representa os volumes primitivos e suas relações de adjacência. A abordagem proposta pode ser sintetizada em 3 passos que são: Inicialização através da representação dos dados com base na triangulação de Delaunay; seleção dos formatos primitivos iniciais utilizando esferas de Delaunay; fusão dos primitivos iniciais utilizando um algoritmo de minimização com base em crescimento por regiões. Os experimentos da aplicação da abordagem foram realizados em imagens tomográficas reais de solo visando representar a estrutura geométrica do espaço de poros a qual possibilita caracterizar propriedades do solo tais como: circulação de fluido e o desenvolvimento de micro e macroorganismos. Além disso, poderia ser utilizado para simulação da bioatividade do solo.

4) *Comparison of 3D Segmentation Algorithms for Medical Imaging:* Bulu [47], apresenta um estudo comparativo entre quatro algoritmos de segmentação aplicados a imagens médicas. Os algoritmos comparados são: dois algoritmos de crescimento por regiões um com semente outro sem, Weibull E-SD Fields (WESDF) e thresholding automático de múltiplos

níveis utilizando o OTSU. Dois automáticos (OTSU e o crescimento de regiões sem semente) e dois semi-automáticos (WESDF e o crescimento de regiões com sementes). O algoritmo de crescimento de regiões com semente, inicia seu processo de crescimento de região das sementes informadas e com base em um critério de conectividade extrai regiões conexas de um volume 3D. O WESDF gera, inicialmente, $K \times K \times K$ cubos do volume, os quais são denominado K-Voxel. Cada K-voxel recebe dois valores: Expectancy and Standard Deviation (E-SD). Depois, utiliza-se o índice Weibull para remover o ruído da imagem, tornando os valores de E-SD mais precisos. Após o cálculo do E-SD, a frequência dos voxels que possuem o mesmo E-SD é mostrada. A segmentação propriamente dita assume que os valores de E-SD em uma região são relativamente homogêneos e são diferentes daqueles em outras regiões. O Thresholding utilizando OTSU, consiste na ideia básica de selecionar automaticamente um valor de threshold separando objetos de interesse em uma imagem de seu plano de fundo com base no valor de tons de cinza. A seleção deste threshold tem como base a escolha de um valor que maximize a variância entre classes de um histograma. O crescimento de regiões sem semente tem o funcionamento que se assemelha ao com semente, com a diferença que o processo inicializa com uma região contendo um pixel, geralmente em uma fatia inicial, e o estado de funcionamento consiste em um conjunto de regiões identificadas e um conjunto de pixels não alocados com fronteira com ao menos uma destas regiões, logo com base em um critério de similaridade atribui-se os pixels não alocados a região com maior similaridade. A avaliação e comparação dos quatro métodos se deu em quatro conjuntos de dados diferentes, um para propósitos de testes e os outros com tomografias computadorizadas reais. Algumas medidas, escalonadas de (1 a 4; 1 a 5 ou 1 a 10), foram utilizadas para comparação dos algoritmos. São elas: implementação, Sensitividade ao ruído, interação com o usuário, tempo de execução, resultado obtido e resultado geral. Estas medidas foram atribuídas empiricamente com base na observação dos autores.

5) *3-D Segmentation Algorithm of Small Lung Nodules in Spiral CT Images*: Diciotti [48] apresenta um algoritmo de segmentação 3D de nódulos em tomografia computadorizada de pulmões. O método proposto inicia com a seleção de um volume de interesse que tem como referência um voxel central informado pelo usuário, assumindo assim que o nódulo está completamente incluso no volume delimitado. Depois, o voxel é super amostrado com uma interpolação trilinear para obter um voxel isotrópico e reduzir efeitos de volume parcial. O próximo estágio identifica marcadores candidatos correspondendo a estruturas com formato de partículas no espaço de escala. Cada marcador candidato é caracterizado pela localização, escala e contraste da estrutura associada. O marcador candidato, localizado 2 mm do centro do volume de interesse e com o maior contraste é automaticamente considerado um marcador nodular. Todos os candidatos são mostrados para o usuário para inspeção e devem ser confirmados como marcador nodular, marcador de uma estrutura pulmonar ou um objeto a ser descartado. Com base nestes marcadores o algoritmo de crescimento de regiões utiliza-os

no processo de segmentação buscando evitar uma possível fusão entre nódulos e diferentes estruturas pulmonares. No fim do processo de segmentação, se a segmentação obtida não for satisfatória, pode-se repetir o processo de supervisão do marcador e segmentação. O volume do nódulo segmentado é extraído utilizando um processo simples de contagem de voxel. O processo de crescimento de regiões utilizado junta a representação em níveis de cinza com o conceito de distância geodésica. O método proposto foi avaliado em imagens com pequenos nódulos fantasmas e em tomografias presentes na base de dados da Lung Image Database Consortium em termos da taxa de erro.

6) *Automatic 3D Segmentation of Lung Airway Tree: A Novel Adaptive Region Growing Approach*: Lai [49] propõe um algoritmo de segmentação adaptativo para segmentação da árvore da via aérea pulmonar em imagens de tomografia computadorizada. A abordagem proposta consiste de duas etapas, na primeira etapa o volume é segmentado grosseiramente e dividido em várias subvolumes através de um método de crescimento de regiões com base em volume de interesse. Estes subvolumes segmentados gerados são divididos em três tipos de acordo com as suas topologias. Na segunda etapa, para cada tipo de subvolume aplica-se um tipo de método específico para encontrar as sementes para a segmentação. Para o primeiro tipo, a semente do primeiro tipo 1 encontrado é o centro geométrico da fatia do brônquio no topo e o threshold inicial é determinado com base em trabalho anteriores. Para os outros tipos 1 a semente é a linha central da bifurcação em seu pai tipo 2, e o threshold inicial é o threshold final de seu pai. Quando a região começa a crescer atribui-se o threshold final como sendo o inicial e aumenta-se o threshold final gradativamente até que ocorra um vazamento. Para tipo 2 a semente é a linha central do seu pai tipo 1 e o threshold inicial é o threshold final de seu pai e o ajuste deste valor é feito similarmente ao realizado para o tipo 1. Para o tipo 3, a semente está na linha central da bifurcação em seu pai do tipo 2 e o threshold inicial é o threshold final do seu pai e o ajuste é o mesmo realizado para os demais tipos. Para avaliar quantitativamente a performance da abordagem proposta, definiu-se uma taxa de detecção, a qual leva em conta o número de bifurcações encontradas pela segmentação automática e o número de bifurcações detectados pela segmentação manual.

7) *An automatic method for colon segmentation in {CT} colonography*: O trabalho proposto por Bert [50], tem como propósito a segmentação da parede do cólon em tomografias computadorizadas de abdômen. Para este fim, foi utilizada uma abordagem de três estágios utilizando um algoritmo de crescimento por regiões 3D adaptativo, com uma condição de crescimento autoajustável. O primeiro estágio da abordagem é a segmentação externa, a qual mascara o ar presente fora da superfície do corpo e é realizada utilizando primeiramente um algoritmo de *threshold* seguido de um método de crescimento de regiões 3D. As sementes para esse primeiro crescimento de regiões são resultantes do processo de *threshold* executado. O segundo estágio é a segmentação dos pulmões, a qual mascara o ar dentro dos pulmões e é realizado através do uso do algoritmo de crescimento de regiões 3D com base em sementes pertencentes ao intervalo de menor valor da intensidade da im-

agem e o pico presente no histograma referente aos primeiros 25 mm do volume. O terceiro estágio, após o ar presente no pulmão e na parte externa ter sido mascarado, é a segmentação do cólon, a qual extrai os segmentos do cólon e é realizado através da inicial geração de um histograma do volume no atual estágio com a finalidade de definir as sementes para o processo de crescimento de regiões (3D 6-conexo com crescimento por largura primeiro). Se durante o processo de crescimento for encontrado um valor de *voxel* maior que o valor do limite superior, um novo valor para o limite superior é calculado com base nos valores locais da parede do cólon, parte adaptativa. O método foi avaliado utilizando os resultados da segmentações realizadas por especialistas em 30 tomografias de abdômen coletadas previamente com os resultados obtidos pelo método proposto.

8) *A fully automatic three-step liver segmentation method on LDA-based probability maps for multiple contrast {MR} images*: Um segmentador 3D automático utilizado para segmentação de fígado em imagens de ressonância magnética é apresentado por Gloger [51]. O método proposto busca utilizar toda a informação disponível nos canais de informação da ressonância magnética para formular um *framework* probabilístico. Neste *framework*, aplica-se uma análise linear multi-classe discriminante como técnica de redução de dimensionalidade e gera-se um mapa de probabilidade utilizado no processo de segmentação. O método proposto pode ser dividido em 4 fases:

- Preparação: nesta fase o conhecimento do especialista é integrado ao método através de um treinamento, o qual consiste das informações de contorno de 10 diferentes tipos de formatos de fígado fornecido por radiologistas. Com base neste conhecimento pode-se categorizar os pontos dos conjuntos de dados de ressonância magnética em tecido do fígado, tecido do rim e plano de fundo.
- Geração do mapa de probabilidade: ao utilizar um método de segmentação por crescimento de regiões percebeu-se ocasionais heterogeneidades na intensidade dentro do fígado, entre o lado esquerdo e o lado direito. Com base nestas observações, a geração dos mapas de probabilidade foram subdivididos em 3 casos: mapa de probabilidade para 3 classes (fígado, rim e plano de fundo); mapa de probabilidade para 2 classes (fígado e plano de fundo); mapa de probabilidade para 2 classes (parte esquerda do fígado e plano de fundo). Estes mapas das probabilidades são determinados através dos histogramas da distribuição das coordenadas dos pontos e da distribuição de intensidade do conjunto de dados de ressonância magnética.
- Segmentação: a segmentação pode ser subdividida em 3 etapas. Na primeira etapa aplica-se um algoritmo de crescimento de regiões nas imagens de probabilidade com baixa resolução. Na segunda etapa, após o término da aplicação do crescimento de regiões grosseiro, ocorre o refinamento da segmentação com a aplicação de um crescimento de regiões nas imagens de probabilidade originais. Na última etapa segmenta-se exclusivamente a parte esquerda do fígado com uma técnica de *threshold* no sentido longitudinal do corpo.
- Refinamento : devido a similaridade das probabilidades

dos tecidos próximo ao fígado ocorre a segmentação errada destas regiões. Para corrigir este problema utiliza-se refinamento das bordas com descritores de *Fourier*, análise da curvatura da superfície do fígado e uma análise discriminante linear.

Os testes do método proposto foram realizados em imagens de ressonância magnética de pacientes entre 21 a 79 anos com os mais variados casos de volumetria de fígado.

9) *A 3D segmentation method of lung parenchyma based on CT image sequences*: O trabalho proposto por REN [52] demonstra uma técnica de segmentação 3D para detecção de nódulos no pulmão em imagens de tomografia computadorizada. Para tanto o método é composto de quatro passos principais: o primeiro passo consiste na binarização do volume analisado através de uma técnica de *thresholding* adaptativa, na qual seleciona-se um *threshold* ótimo no intuito de separar voxels de alta intensidade e das estruturas do tórax das demais estruturas. O segundo passo é a extração do parênquima pulmonar através da aplicação de um rotulamento de componentes conexas 3D na imagem binarizada invertida. Na imagem invertida apenas o parênquima pulmonar, a traqueia e os brônquios e os nódulos aparecem salientados. O terceiro passo tem como propósito a remoção da traqueia e dos brônquios utilizando um método de crescimento de regiões, no qual o conhecimento anatômico sobre a traqueia é utilizado para inicializar a semente automaticamente no processo e os critérios para crescimento e parada do processo são estabelecidos com base em um critério de comparabilidade. Por fim, uma sequência de operadores morfológicos é utilizado para suavizar as bordas e preencher os buracos causados por pequenos vasos, traqueia e nódulos. O método proposto foi avaliado com base em 20 conjuntos de tomografias computadorizadas de 20 pacientes utilizando duas medidas de performance (acurácia média e tempo médio de execução).

10) *Automatic 3D segmentation of human brain images using data-mining techniques*: Uma abordagem para segmentação 3D de imagens de tomografia computadorizada do cérebro humano é apresentado por Uher [53]. A primeira parte da abordagem é o pré-processamento, a qual consiste na aplicação de vários filtros (média, máximo, mínimo, gaussiano, detector de bordas) e outras operações tais como equalização de histograma e normalização. O próximo passo é a segmentação, a qual consiste na geração de pontos 3D na área do cérebro e cada um destes voxels é um ponto de início para o processo. Dado um valor de tolerância, obtém-se os valores máximo e mínimo do voxel pela adição e subtração, respectivamente, deste valor ao valor do voxel. Voxels na vizinhança (6 conexo) com valores maiores que o mínimo e menores que o valor máximo de voxel são adicionados ao segmento e os vizinhos dos voxels adicionados são analisados, repetindo-se o processo até não ser possível adicionar mais voxels ao segmento. O próximo passo consiste na extração de características. Nesta etapa características de borda, tamanho relativo, volume do segmento, média, média geométrica, desvio padrão, soma, soma dos quadrados, mínimo, máximo, assimetria, variância e curtose foram utilizadas. Após a extração de características, o modelo é construído. A árvore de decisão foi o algoritmo de aprendizagem utilizado devido sua melhor acurácia em

relação a outros algoritmos de aprendizagem testados. A abordagem proposta foi testada em 262 fatias de tomografias computadorizadas do cérebro e avaliada com base na sua acurácia com valores de falso positivos e verdadeiros positivos.

11) *Digital rock physics benchmarks-Part I: Imaging and segmentation*: Andrä [54] apresenta 4 *benchmarks* para simulação física de rochas as quais suas aplicações incluem a aquisição da imagem (via tomografia computadorizada), a parte de processamento de imagem, o preparo dos experimentos numéricos e a resolução numérica das equações de campo. Uma das etapas na parte de processamento de imagens é a segmentação e no trabalho apresentado são relatados os experimentos com 3 abordagens de segmentação de 3 grupos de pesquisadores diferentes. A primeira abordagem, do grupo de pesquisadores da VSG, tem como primeira parte a remoção de artefatos, tais como ruído, não uniformidade do brilho e franjas de contraste de fase nas bordas dos grãos. Depois, para um melhor resultado na parte de rotulação um processo de 3 etapas foi utilizado. Na primeira etapa o gradiente de magnitude da imagem foi calculado e depois utilizado um *threshold* para identificar todos os limites. Estes limites são marcados para exclusão na segunda etapa. Para o segundo estágio, os *pixels* escuros do espaço de poros e *pixels* brilhantes do mineral sofreram a aplicação de um *threshold*. Todos os *pixels* próximos ou na interface grão-poro foram excluídos da seleção nesta etapa. Na terceira etapa, cada uma destas duas seleções do estágio anterior foram usadas como marcadores para um algoritmo de watershed com base em marcadores e o gradiente de magnitude foi utilizado como entrada do algoritmo. A segunda abordagem, do grupo de pesquisadores de Stanford, tem como base uma metodologia em 3 passos. Inicialmente as imagens são cortadas lateralmente para remover artefatos das bordas aparentes nos cantos da imagem. Depois, um *threshold* é escolhido utilizando o método do OTSU. Um outro *threshold* manual é também utilizado para distinguir ankerita e zircão do quartzo. Por fim, ocorre a rotulação removendo todas as regiões não conectadas com um volume menor que 50 *voxels*. A terceira abordagem, do grupo de pesquisadores do Kongju, consiste na aplicação de um filtro de suavização para reduzir ruídos da imagem de reconstrução e o uso de um único *threshold* para diferenciar mineral e espaço de poros. Após a aplicação de cada abordagem, ocorreu a aplicação do cálculo de porosidade em cada um dos 4 conjuntos de dados e os resultados compilados em uma tabela única fornecendo a comparação da influência de cada uma das abordagens de segmentação no processo de cálculo de uma das propriedades físicas das amostras apresentadas.

12) *Soft computing approach to 3D lung nodule segmentation in {CT}*: Uma abordagem de múltiplos níveis para a segmentação de vários tipos de nódulos pulmonares em imagens de tomografia computadorizada é apresentada por Badura [55]. Nesta abordagem, inicialmente, ocorre a geração da máscara binária com funções relacionadas a alguns tipos pré-definidos de nódulos. Dependendo do tipo de nódulo, esta fase de segmentação grosseira, é realizada seguindo um ou outro caminho. O volume inicialmente é binarizado utilizando *thresholding* com OTSU o que permite uma distinção grosseira entre os nódulos e as regiões em volta. Em casos

onde os nódulos foram considerados de baixa densidade as outras operações de geração da máscara não são realizadas e a segmentação depende apenas da análise de conectividade difusa, para os demais casos continua-se o processo de geração de máscara. Os demais passos são realizados sobre a imagem resultante do OTSU. O próximo passo é a análise da conectividade, 26 conexo, nesta análise se a semente, a qual é informada manualmente, é distinta do componente com maior número de *voxels* então a separação é completa denotando casos de nódulos vascularizados ou bem restritos. Se a análise de conectividade falha então o nódulo está de alguma forma conectado com a pleura (caracterizando nódulos do tipo 3 ou 4) logo uma separação fatia por fatia é indispensável. Esta separação é realizada com uma análise de 8 conexo para cada fatia e a cada sucesso na separação do nódulo 2D com o segmento maior, completa-se o estágio de geração da máscara para a fatia atual. Os próximos dois passos no processo de geração da máscara consistem na separação para os nódulos do tipo 3, através de operadores morfológicos, e do tipo 4 através da reconstrução das bordas do pulmão. Após a geração da máscara, para cada fatia de máscara gerada ocorre a identificação do ponto central dos nódulos e uma geração de sementes dentro e fora do objeto é realizada para tornar a análise de conectividade difusa mais representativa. Com as máscaras e os conjuntos de sementes gerados, inicia-se o estágio de segmentação mais robusta com base na conectividade difusa. O método de segmentação de conectividade difusa pertence a família de técnicas de crescimento de região e descreve a relação entre par de *pixels/voxels* em termos de uma função de pertinência. Seu valor tem como base a afinidade *fuzzy* das relações entre *pixels/voxels* topologicamente adjacentes, reunidos utilizando algoritmos de procura em grafo e com este valor determina-se se pertencem ao mesmo objeto. Por fim os vasos remanescentes são removidos com uma etapa de pós-processamento que consiste na aplicação de operadores morfológicos, transformação de distância e análise de conectividade com um nódulo central. O método foi validado utilizando um conjunto de dados de 23 estudos adquiridos e descritos pela Lung Image Database Consortium e foram avaliados com base em 3 medidas: Taxa de verdadeiros positivos com 100 e 50 % de confiabilidade e taxa falsas descobertas.

13) *Particle tracking during Ostwald ripening using time-resolved laboratory X-ray microtomography*: O trabalho proposto por Werz [56] apresenta o rastreamento de partículas durante o processo de imagens de engrossamento microestrutural do estado semi-sólido de Al-5 wt.% das amostras de Cu em imagens de microtomografia computadorizada. Para realizar o rastreamento, uma sequência de métodos para remoção de ruídos, melhoria do contraste e segmentação foi aplicada. Inicialmente, para assegurar o funcionalmente confiável do watershed, foram aplicados em sequência o filtro adaptativo denominado *Wiener filter*, o qual é um filtro de difusão anisotrópica que preserva bordas, e um filtro *top-hat* as fatias do volume. O próximo passo consiste na remoção de precipitações esféricas da fase matriz incorporadas dentro das partículas de engrossamento. O passo final antes da aplicação do watershed é a utilização da 3D *Euclidean distance transformation*, a qual atribui um número para cada

voxel que é proporcional a distância do *voxel* a fase matriz. Por fim aplicou-se o watershed fechando assim vazios na reconstrução da fase matriz. Uma vez que os vazios dentro dos limites das partículas foram removidos, um ajuste da fração da fase matriz no volume foi realizado pelo alargamento dos limites das partículas. O resultado é um volume em que cada *voxel* remanescente pertence sem ambiguidades a apenas uma partícula, tornando possível determinar a extensão espacial de cada partícula e rotular elas com um número específico. Para a análise do método proposto as amostras foram repetidamente microtomografadas para capturar diferente regimes da evolução microestrutural durante o processo de *ostwald ripening*. Os dados (enrijecimento médio e taxa de crescimento das partículas) obtidos com o método proposto foram comparados com resultados a priori obtidos em estudos experimentais e com simulações computacionais do processo de *ostwald ripening*.

14) *3D Segmentation and Labeling of Fractured Bone from CT Images*: O método proposto por Paulano [57] visa realizar a segmentação e a rotulação de fragmentos de ossos em imagens de tomografia computadorizada. O método em questão tem como base um algoritmo de segmentação 2D por crescimento de regiões e requer uma interação mínima do usuário para inicializar as sementes nos fragmentos do osso na primeira fatia em que aparecerem. Após as sementes terem sido colocadas, ocorre a segmentação das fatias onde as sementes foram inicializadas, de tal forma que para cada semente inicializada ocorre a segmentação e a subsequente subtração de regiões que se sobrepõem, eliminando assim sementes que pertencem a uma região já segmentada. Depois da realização da segmentação para a fatia atual, todas as sementes são propagadas para a fatia subsequente e repete-se a segmentação como descrita anteriormente. Para propagar as sementes primeiro as sementes são herdadas da fatia anterior e se a semente falha, ou seja não pode segmentar uma área considerada área de osso, os seus vizinhos são considerados candidatos a semente. Se todos os vizinhos também falharem a semente é removida. O algoritmo termina quando todas as sementes foram removidas. O método proposto foi testado contra métodos comumente utilizados para segmentar ossos saudáveis ou fraturados em imagens de tomografia computadorizada. Outro teste realizado foi a aplicação do método em conjuntos de imagens contendo diferentes tipos de fraturas.

F. Redes neurais

1) *Automatic clinical image segmentation using pathological modeling, {PCA} and {SVM}*: A abordagem proposta por Li [58], apresenta um sequenciamento de algoritmos para segmentações de imagens clínicas em imagens de tomografia computadorizadas e imagens de raio-x. O processo como um todo é dividido em duas etapas, treinamento e segmentação. Na etapa de treinamento, as imagens selecionadas manualmente são segmentadas utilizando um método variacional de *level set* guiado por uma função de energia projetada para delinear as características patológicas da imagem. A parte de projeto da função de energia utiliza o auxílio de um especialista para modelar a função de acordo com as características patológicas

das diferentes regiões da imagem. Como próximo passo, tem-se a extração de características através da utilização de um método com base em janelas sobre as imagens segmentadas da etapa anterior e a posterior aplicação de um *Principal Component Analysis* (PCA) visando o uso das características extraídas mais relevantes. As características extraídas são utilizadas para o treinamento do *Support Vector Machine* (SVM) e quando há a necessidade de segmentar uma imagem utiliza-se o já treinado SVM para classificá-la. Para avaliar os resultados da abordagem proposta foram utilizadas escaneamentos 2D e 3D do tórax e imagens de raios-X da arcada dentária e os resultados analisados de forma subjetiva.

2) *A New 3D Segmentation Algorithm Based on 3D PCNN for Lung CT Slices*: Chang [59], propõe um algoritmo de segmentação com base em um modelo de Pulse Coupled Neural Network (PCNN) em imagem de tomografia computadorizada de pulmão. O algoritmo de PCNN consiste de três partes: o campo receptivo, o campo modulador e o gerador de pulso. O papel do campo receptivo é receber as entradas de outros neurônios e de fontes externas por dois canais (ligação e alimentação). No campo modulador as entradas dos canais são moduladas e o seu resultado é enviado para o gerador de pulso, o qual é composto por um gerador de pulso e um comparador. O resultado é comparado com um threshold dinâmico para decidir se o neurônio é ativado ou não e se ativado o gerador de pulso terá como saída 1 e o threshold dinâmico é ajustado de acordo. Com base neste modelo aplicado para 2D expande-se o mesmo para 3D seguindo os seguintes passos: constrói-se uma matriz 3D com todas as fatias, seleciona-se o volume de interesse e inicializa-se os parâmetros. Depois, segmenta-se o volume do pulmão com a versão 3D e melhorada do PCNN (o 3D I-PCNN) para obter uma imagem binária. Como próximo passo, extrai-se os campos do pulmão na imagem binária através do uso da erosão e suavização dos contornos. Por fim, multiplica-se o dado binário segmentado 3D e a imagem original para se obter as imagens dos campos do pulmão e renderizá-las em um espaço 3D com um algoritmo de renderização de volume. O algoritmo proposto foi avaliado em uma base de dados de tomografia computadorizada contendo 5 casos. A avaliação se deu com a comparação dos resultados obtidos com o ground truth gerado manualmente por especialistas em termos de mean distance, root means square distance e Tanimoto coefficient.

3) *Automatic detection of small lung nodules in 3D {CT} data using Gaussian mixture models, Tsallis entropy and {SVM}*: Santos [60], propõe uma metodologia para a detecção automática de pequenos nódulos no pulmão em imagens de tomografia computadorizada de baixa dosagem. Esta metodologia representada pela figura 3 é composta de quatro estágios que são: aquisição das imagens de tomografia computadorizada, segmentação dos nódulos candidatos, extração de características e classificação dos candidatos nas classes nódulo e não nódulo. As imagens utilizadas no desenvolvimento do trabalho foram obtidas em uma base de dados denominada Lung Image Database Consortium. Para o segmentação dos nódulos candidatos quatro etapas são utilizadas: a primeira etapa é a segmentação do parênquima pulmonar com a utilização de algoritmos de *threshold* e de crescimento

de regiões. A segunda etapa, consiste na complementação da etapa anterior para remover a traqueia e os brônquios principais através da seleção, em cada fatia do volume, das duas maiores regiões. A terceira etapa segmenta a estruturas dentro dos pulmões utilizando o método de *Gaussian Mixture Models* (GMM) para modelar a distribuição do valor de intensidade dos *voxels* da imagem resultante da etapa anterior distinguindo-os em duas classes: estruturas internas e parênquima pulmonar. Na quarta etapa e ultima etapa, apenas as estruturas com formato e textura que mais se parecem com nódulos são selecionadas através do uso do formato da estrutura encontrada. Como os nódulos são estruturas esféricas, utiliza-se uma matriz hessiana para encontrar estruturas arredondadas evidenciando-as das demais estruturas. Depois, sobre esses possíveis candidatos realiza-se uma extração de características de textura, com o cálculo de entropia proposto por Shannon e Tsallis Q, e uma classificação destas características com *Support Vector Machine* (SVM) para diferenciar os nódulos de outras estruturas selecionadas como candidatos, mas que não correspondem a nódulos. A validação do método proposto é dada pelo número de verdadeiros positivos, falsos positivos, sensibilidade, especificidade, acurácia e pela taxa de falsos positivos por exame.

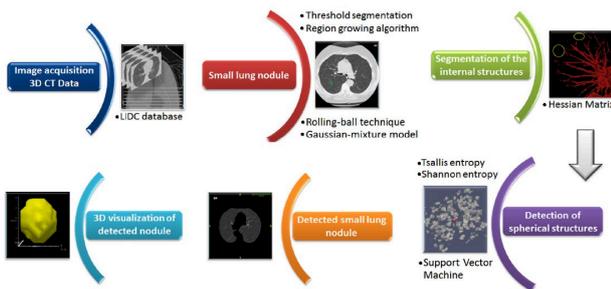


Fig. 3. Fluxograma da metodologia proposta por Santos. Extraído de [60].

4) *Liver vessel segmentation based on extreme learning machine*: O trabalho apresentado por Zeng [61] propõe um algoritmo de segmentação de vasos do fígado com base em *Extreme Machine Learning* (EML) em imagens de tomografia computadorizada. Com este intuito, inicialmente, tem-se a aplicação de filtro anisotrópico para remoção de ruído mantendo a preservação das bordas. Depois, com base no conhecimento de formatos e estruturas geométricas, três filtros (*Saito*, *Frangi* e *offset medialness*) associados com filtro de energia de deformação são utilizados para extrair características estruturais dos vasos. Nesta etapa, os filtros *Saito* e *Frangi* são utilizados para detectar estruturas 3D dos vasos, o *Offset Medialness* para extrair a topologia dos vasos e filtro de energia de deformação para aprimorar estruturas complexas tais como bifurcações e ramificações. Por fim, tendo as características extraídas na etapa anterior, aplica-se uma normalização, um posterior treinamento e o processo classificatório utilizando o EML segmentando assim os vasos do fígado dos *voxels* do plano de fundo. Os experimentos com a técnica proposta utilizaram imagens de tomografia computadorizada de abdômen de 6 conjuntos de dados clínicos

e a avaliação dos resultados se deu com a análise de um radiologista especialista na área, realizando correções nos resultados de segmentação e fazendo o cálculo de três medidas: acurácia, sensibilidade e especificidade.

G. Métodos diversos

1) Association Rule mining:

a) *Integrating discretization and association rule-based classification for Alzheimer's disease diagnosis*: Chaves [62], propõe um sistema de auxílio a diagnóstico para o diagnóstico cedo de casos de Alzheimer's. Este sistema consiste em 3 etapas: A primeira etapa é o procedimento de utilização de uma máscara. Esta máscara é construída com base na utilização de um *threshold* com base em um histograma da imagem média de controle, visando obter a melhor máscara, correspondendo aos maiores *bins* do histograma. Depois, aplica-se uma discretização, a qual consiste no uso de um histograma de igual largura aplicado ao intervalo dos valores de intensidade dentro do limite da máscara dividindo-os em *k bins* de igual largura, para a seleção das regiões de interesse utilizadas na etapa posterior. As regiões de interesse são utilizadas como entrada para o algoritmo de *Association Rule mining* (AR mining), o qual na sua etapa de treinamento, captura padrões de co-ocorrência dentro dos dados de controle e tem como objetivo utilizar regras para classificar as imagens em duas classe (controle e Alzheimer's). Uma regra neste contexto é a relação entre itens de transação com suporte e confiança suficiente. O processo de AR mining emprega um algoritmo para procurar conjuntos de itens com confiança e suporte maiores que uma confiança mínima e um suporte mínimo estabelecidos pelo usuário. O sistema proposto é avaliado utilizando duas base de dados diferentes de *Single Photon Emission Computed Tomography* (SPECT) e *Positron Emission Tomography* (PET) utilizando como medidas a acurácia, a sensibilidade e a especificidade.

2) Componente conexas:

a) *Automatic 3D segmentation of mandible for assessment of facial asymmetry*: Spampinato [63] apresenta um método para segmentação e medição da superfície da mandíbula em imagens de tomografia computadorizada. O algoritmo para segmentação é realizado em duas fases. A primeira fase consiste na extração da arca inferior da mandíbula. Esta etapa identifica uma máscara, a qual é aplicada para filtrar apenas a arcada dentária inferior e pode ser descrita nos seguintes passos: Após uma fase de orientação, aplicam-se as seguintes operações para cada fatia começando pelas fatias de baixo: bit-plane slice coding (a fatia é dividida em várias imagens, onde cada imagem representa a contribuição de um bit específico para a aparência total da imagem), seleção do plano do bit com maior significância, aplicação de operadores morfológicos (erosão e dilatação) e identificação do maior componente conexo em cada fatia. A máscara é selecionada com base no componente conexo com menor taxa, calculada através da razão entre a área do componente conexo e a área do casco convexo do componente). A segunda fase tem como objetivo a extração da arca superior da mandíbula. Para esta fase ocorre a aplicação dos dois primeiros passos utilizados na extração da parte inferior

começando da fatia em que se inicia a arcada dentária superior. Em seguida, os componentes conexos são computados e para ser considerado parte da mandíbula os componentes devem satisfazer três restrições (distância, posição e formato). Uma vez extraída a região da mandíbula aplica-se a renderização da superfície para a reconstrução 3D e posterior medição entre pontos de referência localizados na superfície reconstruída. O método proposto foi avaliado contra medições realizadas por especialistas em 10 volumes tomográficos (5 do grupo de controle e 5 de pacientes com assimetria facial) em termos de média e desvio padrão.

3) Programação dinâmica:

a) *Segmentation of Lung Nodules in Computed Tomography Images Using Dynamic Programming and Multidirection Fusion Techniques*: O trabalho apresentado por Wang [64], propõe um método para segmentação de nódulos em pulmões em imagens de tomografia computadorizada. A abordagem proposta pode ser dividida em duas fases. Na primeira fase foi estendido o modelo 2D de programação dinâmica em um modelo 3D de programação dinâmica, no qual o centro da fatia e o número e o comprimento das linhas radiais foram adaptadas para cada fatia, e uma função de custo interno entre fatias adjacentes foi introduzida. Com base neste modelo, realiza-se uma segmentação fatia a fatia começando de uma fatia do meio e progredindo para as duas extremidades, não apenas na direção axial, mas também nas direções sagital e coronal. Esta segmentação realiza, para cada fatia, a transformação das mesmas em uma imagem composta por uma série de linhas verticais paralelas igualmente espaçadas, transformando, as fronteiras do nódulo, em curvas horizontais. Tendo esta imagem transformada, o problema da segmentação pode ser resolvido através da procura de um caminho ótimo com o mínimo custo utilizando o algoritmo de programação dinâmica. Na segunda fase, combina-se as partes segmentadas, resultantes da etapa anterior, perto do meio do nódulo para cada direção e aplica-se um operador morfológico de abertura para obter a segmentação final. Para avaliar a performance da abordagem proposta foi utilizado um cálculo de sobreposição para quantificar a consistência dos resultados de segmentação com o padrão ouro fornecido por um especialista. Este cálculo de sobreposição foi complementado com a fração dos critérios de verdadeiros positivos e falsos positivos, para os dois conjuntos de dados provenientes do Lung Imaging Database Consortium.

4) Algoritmos genéticos:

a) *Automated tracing of filaments in 3D electron tomography reconstructions using Sculptor and Situs*: Rusu [65] descreve um fluxo de trabalho para segmentar automaticamente filamentos em mapas de *cryon-electron tomography (cryo-ET)*, o qual inclui toda parte de pré-processamento finalizando com a segmentação e comparação dos resultados. Na parte de pré-processamento tem-se as seguintes etapas: A primeira etapa consiste no teste e na utilização de dois filtros de remoção de ruído (*Gaussian-weighted averaging* e *Digital Paths Supervised Variance*); uma etapa de normalização, com a finalidade de aprimorar a aparência das características das estruturas e equalizar uma desigual distribuição de densidade nos mapas; uma etapa de edição do mapa, através da utilização de ferr-

mentas de *clipping* de polígono e de *floodfill* com múltiplos pontos com o intuito de auxiliar na remoção de regiões sem filamentos; uma etapa para correção da degradação da imagem, devido a incompleta amostragem no espaço de *Fourier* realizando uma atenuação dependente da orientação da intensidade dos filamentos. Depois da parte de pré-processamento vem a parte de segmentação dos filamentos utilizando o Volume Trace (VolTrac), o qual utiliza um algoritmo genético para, inicialmente, colocar randomicamente *templates* cilíndricos e com os processo de mutação e *crossover* fazer a procura dos cilindros que maximizam o valor de *fitness*. Por fim, conclui-se o fluxo de trabalho com a validação do resultados, comparando os resultados obtidos para segmentação automática de filamentos com a segmentação manual realizada por especialistas.

5) Conectividade Fuzzy:

a) *Jaw tissues segmentation in dental 3D {CT} images using fuzzy-connectedness and morphological processing*: Lloréns [66], propõe um método para segmentação de tecidos da mandíbula em imagens de tomografia computadorizada. O método apresentado visa reconstruir os tecidos da mandíbula partindo de um conjunto de cortes seriais transversais definidos perpendicularmente ao plano da arcada dentária. A mandíbula é dividida em 5 regiões de acordo com o tecido presente na mesma. O método escaneia cada fatia e estima a qual região cada corte transversal pertence e tenta segmentar os tecidos que estão presentes. Os tecidos mais densos são segmentados utilizando uma técnica de *threshold* associado a inserção das bordas para cada corte transversal. As bordas são estimadas, por operações morfológicas, como sendo o resíduo entre a seção do tecido mais denso e uma erosão dela com um pequeno elemento estruturante. O nervo é segmentado utilizando o *Fuzzy Connectedness Object Extraction (FCOE)*, o qual é um método que inicia com sementes e avalia a finidade entre a semente e o pixel da imagem e com base neste mapa de afinidade e em um *threshold*, delimita os *pixels* que estão conectados no mesmo objeto. As sementes são extraídas com base na visão da projeção pseudo-Ortopantomográfica, onde uma visão mais larga do nervo pode ser observada. Nesta visão, o nervo é definido por função de segundo grau devido ao fato que o mesmo descreve quase uma trajetória quadrática. Com esta função extrai-se as sementes para o algoritmo de segmentação. Quando todos os cortes transversais foram segmentados o volume 3D é reconstruído. A figura 4 mostra um fluxograma do algoritmo proposto. O método foi testado em um conjunto de *ground truth* de 20 pacientes diferentes e avaliado utilizando quatro medidas que são: o índice Jaccard, o coeficiente *DICE* e as distâncias ponto a ponto e ponto a curva.

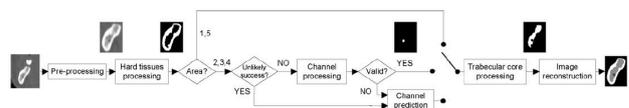


Fig. 4. Fluxograma do algoritmo proposto por Lloréns. Retirado de [66].

6) *Com base em matriz Hessiana:*

a) *Hessian based approaches for 3D lung nodule segmentation:* Gonçalves [67] propõe um método de segmentação de nódulos em imagens de tomografia computadorizada. O método proposto utiliza um processo de segmentação em múltiplas escalas que utiliza o princípio adaptativo central *medialness*. Este princípio adaptativo consiste em uma estratégia com base em matrizes hessianas, as quais fornecem uma boa segmentação em estruturas com formato de partículas (blobs) com no caso dos nódulos pulmonares. Como o foco é a segmentação do nódulo, a estratégia parte da localização de um nódulo detectado de forma a avaliar o método na tarefa de extrair o volume do nódulo. Para tanto, inicialmente, um método de suavização (filtro gaussiano) em múltiplas escalas é utilizada para reduzir o nível de ruído. Depois, computam-se as matrizes hessianas 3D e os correspondentes autovalores os quais são combinados para gerar a máscara de segmentação. Na parte de combinação para geração da máscara 3 abordagens foram testadas. A primeira calculando de *shape index* e *Curvedness* a partir do maior e do menor autovalores. A segunda utiliza os dois maiores autovalores para medir a força da estrutura e razão deles é aplicada para corrigir o desvio do centro da estrutura. E a terceira utiliza a combinação das duas abordagens anteriores através da união da detecção proveniente de cada uma. Uma vez que se obteve o realçamento das estruturas onde os nódulos podem ser claramente identificados o passo final é a aplicação de um *threshold* para selecionar os *voxels* com maiores valores e assim construir as máscaras pela conectividade das estruturas dos nódulos finais segmentados. As diferentes abordagens foram avaliadas em tomografias de diferentes casos de nódulos pulmonares presentes na base de dados *Lung Image Database Consortium and Database Resource Initiative*. Cada tomografia utilizada foi avaliada por 4 radiologistas e apenas nódulos que tiveram um forte consenso entre os 4 especialistas foram considerados. Para medida de comparação das abordagens automáticas com as segmentações realizadas pelos especialistas foi utilizado o índice jaccard.

7) *Com base em Histograma:*

a) *3D segmentation of the cerebrospinal fluid from CT brain scans using local histogram similarity map:* Fabijańska [68] propõe um método de segmentação de regiões preenchidas com líquido cefalorraquidiano em tomografias computadorizadas do cérebro. O método em questão consiste de 7 passos: Inicialmente tem-se a indicação, através do usuário, de uma semente com formato elipsoidal dentro da região de interesse. O segundo passo consiste na normalização dos dados de entrada no intervalo $[0,255]$, evidenciando assim valores correspondendo a tecidos do cérebro e atenuando o ar em volta e os tecidos mais densos. O terceiro passo é o cálculo do histograma H_{sd} de intensidade dos voxels dentro da região elipsoidal previamente informada como semente. Este histograma de referência H_{sd} é comparado com a distribuição local de intensidade em toda a imagem de forma a identificar voxels pertencentes ao objeto segmentado. No passo 4, o histograma local H_{xy} em uma vizinhança circular é coletado para cada voxel do mesmo plano e computa-se o mapa de similaridade do histograma local com o histograma de

referência através da distância de Bhattacharyya e o nível de similaridade é atribuído para cada voxel. Para melhorar a acurácia da região desejada do cérebro o mapa de similaridade é aprimorado com a aplicação da reconstrução pela intensidade seguido por uma operação de abertura com uma elipsoide apropriada como elemento estruturante. Por fim, como ultimo passo, aplica-se o método de thresholding (OTSU) sobre o mapa de similaridade, para produzir o resultado final de segmentação. A acurácia do método de segmentação proposto foi medido através dos coeficientes DICE e Jaccard em 5 casos de tomografias computadorizadas do cérebro.

8) *Ajuste de polinômio:*

a) *3D segmentation of vessels by incremental implicit polynomial fitting and convex optimization:* Biesdorf [69], propõe uma nova abordagem para segmentação 3D de vasos em imagens de Tomografia Computadorizada de Angiografia (TCA). Esta abordagem utiliza um implícito ajuste de polinômio e uma otimização convexa combinados com um esquema de rastreamento incremental. O método utilizado pode ser resumido em 3 passos principais: inicialização, estimativa de parâmetros locais do modelo utilizando otimização convexa e rastreamento ao longo da linha central. Na parte de inicialização é necessária uma estimativa grosseira para o ponto inicial, a orientação local, o raio da região de interesse para o ajuste do modelo, assim como para a intensidade do plano de fundo e do primeiro plano. Essas informações são necessárias apenas para a inicialização do primeiro segmento de vaso, para os demais segmentos de vaso utiliza-se o resultado da segmentação do vaso anterior. Para a estimativa dos parâmetros locais do modelo a função convexa é minimizada em uma região de interesse 3D ao redor do vaso. Para a minimização, uma abordagem de projeção de gradiente descendente é utilizada e a atualização dos parâmetros é computado iterativamente. Para o rastreamento incremental ao longo da linha central do vaso, prevê-se uma nova posição 3D com base na posição da linha central estimada e na orientação local do segmento de vaso atual. A abordagem proposta foi testada em mais de 100 imagens sintéticas e em 10 diferentes casos clínicos de TCA, medindo nestes casos a acurácia com base em parâmetros clínicos relevantes (diâmetros máximo, médio e mínimo dos vasos) assim como o erro médio para a posição da linha central e o coeficiente DICE médio.

9) *Correspondência de template:*

a) *Numerical reconstruction of graphite/epoxy composite microstructure based on sub-micron resolution X-ray computed tomography:* O trabalho proposto por Czabaj [70] apresenta o imageamento, a visualização e a reconstrução numérica de um polímero com fibra reforçada em imagens tomografia computadorizada na escala microscópica. O polímero investigado é o grafite/epóxi e o como primeiro passo tem-se a digitalização na escala microscópica. Depois, para uma correta reconstrução numérica, utiliza-se um algoritmo de segmentação para identificar e estimar a posição correta das fibras individuais no volume. O algoritmo de segmentação em questão, é dividido em duas etapas: na primeira etapa um algoritmo de *Template-Matching* (TM) é utilizado para detectar as coordenadas 2D dos centroides das fibras em cada imagem de corte transversal do volume, a seleção de cada

template foi realizada através da tentativa e erro visando maximizar o número geral de detecções verdadeiras. Na segunda etapa, o algoritmo de atribuição e rastreamento de múltiplas fibras utiliza a detecção do algoritmo de TM para determinar a coordenada 3D de cada fibra individual no volume. Após a segmentação das fibras é utilizado um algoritmo de converter o dados resultantes do processo de segmentação, isto é, coordenadas e diâmetros, em uma detalhada malha 3D de elementos finitos. O algoritmo descrito foi aplicado em um sub-volume de 169^3 de um volume composto de 367 fatias e avaliado qualitativamente, com base na visualização das reconstruções realizadas, e quantitativamente com base na distribuições das fibras e na computação dos ângulos das fibras relativo ao eixo global z.

10) Votação de tensores:

a) *Automatic Segmentation and Quantification of Filamentous Structures in Electron Tomography*: Uma abordagem de segmentação e quantificação de filamentos dos objetos em tomografia 3D com base em elétron é proposta por Loss [71]. Esta abordagem consiste em 3 passos: o primeiro consiste na utilização de um filtro hessiano para aprimoramento local dos filamentos; o segundo passo é a detecção e complementação, como por exemplo o preenchimento de falhas, das estruturas dos filamentos através de *tensor voting*; o terceiro passo é a demarcação da rede de filamentos através de um algoritmo 3D de rastreamento de curvas, o qual permite a detecção e o rotulamento dos filamentos e junções as quais podem ser quantificadas. Este algoritmo de rastreamento de curvas itera entre achar *voxels* para servirem como sementes para o rastreamento e ligar *voxels* adjacentes ao longo de uma direção consistente. O método proposto foi inicialmente avaliado em dados sintéticos com a inserção de ruídos e para sua avaliação quantitativa duas variáveis foram observadas, as taxas de precisão e recordação. Estas variáveis foram definidas com base no número de falsos positivos, falsos negativos e positivos verdadeiros e representam a medida da quantidade de ruído presente na solução e a medida da quantidade de *ground truth* segmentada. Em outro experimento foram utilizadas imagens reais de paredes celulares de plantas que passaram sobre diferentes tratamentos químicos com a finalidade de quantificar propriedades composicionais e morfométricas das paredes celulares das plantas. Os resultados deste experimento foram validados com base em estudos a priori.

11) Mesmo trabalho vários algoritmos:

a) *Computer Tomography 3D Edge Detection Comparative for Metrology Applications*: Ontiverosa [72], faz uma análise de técnicas para detecção de bordas para a aplicação em metrologia em imagens de tomografia computadorizada. Inicialmente, ele explica cada uma das técnicas comumente utilizadas para detecção de bordas (*Atlas-guided*, *Deformable models*, *Detection of discontinuities*, *Pattern classification*, *Region oriented*, *Threshold based*, *Trainable segmentation*, *Watershed transformation*), fazendo referências a alguns métodos para cada técnica. Depois, o autor seleciona dois métodos, justificando a seleção destes e a não seleção dos demais como mostrado na figura 5, para uma análise comparativa através da aplicação das técnicas em uma volume de referência. Nesta avaliação, ambos os métodos foram aplicados mais de 10 vezes

comparando as diferenças de medição em relação ao volume modelo previamente calibrado, a qual mostra o poder de reprodutibilidade do Canny e as pequenas diferenças do *threshold* local.

Methods	Information used	Training	Computational cost	Comments
Atlas-guides	Spatial map	Yes	Low	It was not selected because it is necessary an atlas construction.
Deformable models	Feature vector	No	Medium	It was not selected due to the heavily dependence of the results on the manual interaction to place an initial model and choose appropriate parameters.
Detection of discontinuities	Voxel	No	Medium	It was selected because provides a good surface location capability and significantly reduces the data quantity.
Pattern classification	Feature vector	Yes	Medium	It was not selected because its main goal is just distinguish image features, also can be sensitive to noise and intensity in-homogeneities.
Region oriented	Image patch	No	Low	It was not selected because requires manual interaction to obtain the seed point that must be planted, also segmentation can be sensitive to noise, partial volume effects can cause separate regions to become connected.
Threshold based	Voxel	No	Low	It was selected because it is stable in images with different structures; also there are some works that have been carried out for metrology applications.
Trainable segmentation	Feature vector	Yes	High	It was not selected because they learn by example and has a high computational cost.
Watershed transformation	Image patch	No	Low	It was not selected because this segmentation method generally leads to over-segmentation due to noise and local irregularities of the gradient.

Fig. 5. Tabela com métodos de extração de superfícies. Extraído de [72]

b) *Comparing algorithms for automated vessel segmentation in computed tomography scans of the lung: the {VESSEL12} study*: O trabalho de comparação de algoritmos para segmentação automática de vasos apresentado por Rudyanto [73], visa fornecer uma plataforma para comparar a performance de algoritmos de segmentação no processo de identificação de vasos do pulmão em imagens de tomografia computadorizada. Para tanto, um conjunto de dados de referência contendo anotações sobre 20 tomografias computadorizadas e a proposta de 9 categorias para realização da avaliação foi elaborado. O estudo elaborado tem toda a descrição da parte de aquisição do conjunto de dados (patologia, tipo de imagem, *scanner* e *kernel*, espaçamento, número de fatias e voltagem e amperagem), a definição das categorias de avaliação, a categorização dos métodos de segmentação apresentados (autor, o tipo de algoritmo, se é multi-escala ou não e se possui uma etapa de pós-processamento) e toda a parte de pontuação com toda uma análise estatística dos resultados. Com base em todo o trabalho apresentado, o autor afirma que as principais contribuições do trabalho são: o conjunto de dados de referência com anotações, o sistema quantitativo para pontuação e comparação dos algoritmos e a avaliação dos pontos fortes e fracos dos algoritmos avaliados na presença de algumas doenças pulmonares.

c) *Automatic 3D Pulmonary Nodule Detection in CT Images*: O autor, valente [74], apresenta um *survey* com base em métodos de identificação automática de nódulos em imagens de tomografia computadorizada. O estudo realizado mostra toda a metodologia utilizada que vai dos critérios para a escolha dos trabalhos até uma análise crítica sobre os mesmos e as perspectivas futuras para a área. Os trabalhos selecionados para serem avaliados foram categorizados segundo a presença ou não dos 5 principais estágios normalmente presentes em ferramentas de auxílio a diagnósticos que são: aquisição de dados, pré-processamento, segmentação de pulmões, detecção do nódulo e a redução de falsos positivos. Para cada uma das

5 etapas existe uma descrição sobre o funcionamento de cada uma delas com as técnicas mais utilizadas pelos trabalhos e, em algumas destas etapas, há uma tabela mostrando especificamente o método utilizado por cada um dos trabalhos, como por exemplo, na parte para detecção automática dos nódulos e na parte de técnicas de classificação utilizada para reduzir os falsos positivos. O trabalho apresenta também, além de uma análise de cada trabalho selecionado, uma tabela comparativa com os nomes dos autores do trabalho e características, julgadas importantes para esse tipo de trabalho, presentes ou não em cada um. Algumas dessas características são: a sensibilidade do método, o número de falsos positivos por exame, o número de nódulos utilizados na validação, o tamanho dos nódulos, o tempo de resposta e o tipos dos nódulos. Por fim, o autor faz um resumo do atual estágio da área de identificação automática de nódulos com base na análise dos trabalhos.

V. DISCUSSÃO

De um ponto de vista sistemático, podemos realçar alguns aspectos chave sobre segmentação 3D aplicado à imagens tomográficas:

- O uso de algoritmos de segmentação 3D com base em imagens de TC está ainda com foco maior na área de saúde humana, como pode ser visualizado no gráfico da figura 6. No entanto, existem algumas áreas demonstrando uma tendência de crescimento no uso de segmentadores 3D, como por exemplo na área da geologia e de análise de objetos em geral.
- Existe uma grande quantidade de técnicas de segmentação 3D com diferentes metodologias, o que indica uma constante e crescente pesquisa na área de segmentação 3D.
- O histograma, indicado na figura 7, mostra os métodos principais agrupados de acordo com sua frequência de ocorrência. Deste histograma, pode-se perceber que os métodos com base em modelo foram os mais utilizados no período sendo analisado, o que indica que o uso de modelos para guiar o processo de segmentação, no escopo de segmentação 3D, tem obtido mais resultados positivos. Também pode-se perceber que os métodos de threshold foram utilizados com baixa frequência. No entanto, esta categoria de métodos foram frequentemente apresentados nos trabalhos analisados como processos de suporte ou procedimentos de inicialização para outros métodos, demonstrando a importância de tal técnica.

Também pode-se apontar que a revisão proposta, seguindo a metodologia de Kitchenham, demonstra: (i) uma busca reproduzível com uma metodologia sistemática para revisar os algoritmos de segmentação 3D aplicados em imagens tomográficas no período de 2006 até julho de 2016; (ii) um resumo de todos os trabalhos analisados que tiveram uma correspondência com os critérios estabelecidos, especificando o principal método de segmentação, o tipo de imagem utilizado, os experimentos realizados e os métodos de avaliação de performance empregados por cada trabalho, categorizando cada um dos trabalhos em seus respectivos grupos; uma visão geral analisando, com base nos trabalhos selecionados, a área de segmentação 3D com aplicação em imagens tomográficas.

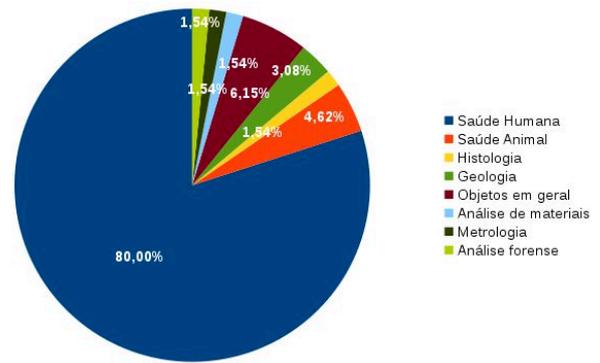


Fig. 6. Gráfico de pizza mostrando a frequência das principais áreas de aplicação dos métodos analisados.

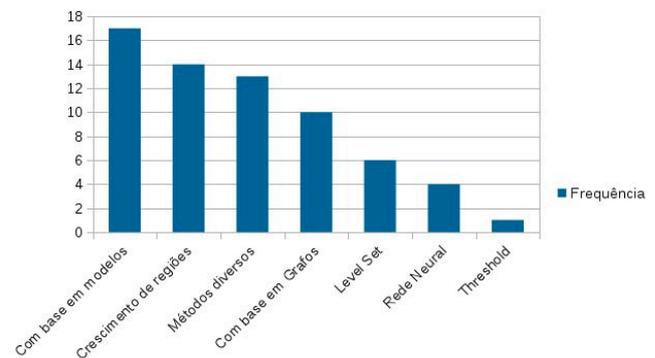


Fig. 7. Histograma mostrando a frequência do método principal utilizado na segmentação com base na análise do agrupamento dos trabalhos.

A. Análise da avaliação de performance

Um importante aspecto no desenvolvimento de algoritmos de segmentação 3D é a sua avaliação. Para realizar tal avaliação, os trabalhos analisados focaram em aspectos específicos tais como: robustez, especificidade, sensibilidade, acurácia, precisão, entre outros. Estes aspectos foram medidos, no geral, com base no número de falso positivos, falso negativos, verdadeiro positivos e verdadeiro negativos. Alguns trabalhos analisados utilizaram uma avaliação qualitativa associada com uma análise quantitativa e outros utilizaram apenas uma avaliação qualitativa. Para a avaliação qualitativa, no escopo dos trabalhos analisados, uma análise subjetiva foi predominante, por exemplo uma análise visual dos resultados obtidos. No entanto este tipo de avaliação não é comumente utilizado para uma ampla comparação entre algoritmos de segmentação. No geral, a avaliação é realizada com base no conhecimento do especialista, isto é, os resultados de segmentação obtidos são comparados com uma segmentação realizada manualmente, geralmente rotulada como *ground-truths* ou padrões ouro, empregando uma métrica específica ou características específicas para avaliação. Nos trabalhos analisados as métricas mais utilizadas foram o *DICE Similarity Coefficient* [75] e o índice Jaccard [76]. Outras métricas para avaliação de métodos de segmentação 3D podem ser encontradas nos trabalhos proposto por Taha [77].

B. Perspectivas Futuras

Com base na informação coletada fornecida pelos trabalhos analisados, podemos observar que os métodos de segmentação 3D (novos ou existentes) devem levar em conta aspectos de performance no intuito de obter resultados com certo grau de eficiência. Execuções em paralelo é explorado ainda de uma maneira sutil, abrindo assim uma grande oportunidade para pesquisa e desenvolvimento. Falando qualitativamente, uma metodologia unificada para avaliação ainda é necessária, uma vez que as metodologias que surgiram utilizam conjuntos de dados próprios nos quais as imagens foram adquiridas em um ambiente totalmente controlado e geralmente, se não disponíveis publicamente, não podem ser reproduzidas e utilizadas para uma avaliação comparativa por outros métodos de segmentação. Podemos também fazer algumas premissas da futura direção que a área de segmentação 3D aplicada a imagens tomográficas está tomando:

- um aumento no uso de métodos de segmentação híbrido, isto é, métodos que utilizam mais que um tipo de segmentação, procurando produzir resultados mais confiáveis;
- emprego dos métodos de segmentação 3D em mais áreas, guiadas pela grande popularização dos sensores de aquisição 3D;
- desenvolvimento de métodos de segmentação 3D paralelos, impulsionados pelo aumento da quantidade de informação disponível e pelo desenvolvimento das *GPUs* e processadores com múltiplos núcleos;
- surgimento de metodologias para validação e avaliação, fornecendo uma forma mais unificada para a comparação de métodos de segmentação 3D;
- uso de conjuntos de dados grandes para a avaliação dos algoritmos de segmentação 3D;
- extensão e aplicação dos métodos de segmentação 2D que ainda não foram testados em um contexto 3D.

VI. CONCLUSÃO

A revisão sistemática da literatura apresenta um visão geral em relação a área de segmentação 3D aplicada a imagens tomográficas. Esta pesquisa incluiu artigos publicados entre 2006 e julho de 2016 em 3 bases de dados de artigos científicos (Science Direct, ACM and IEEEExplore). Avanços na avaliação qualitativa e quantitativa e aplicação em áreas diferentes da área da saúde humana foram demonstradas, mas existem ainda áreas a serem testadas e metodologias a serem formalizadas. No geral, os vários trabalhos analisados mostraram seus potenciais na área da saúde humana, no entanto, o poder de generalização dos métodos apresentados é ainda uma questão em aberto. Um experimento testando este poder seria um possível experimento de comparação das habilidades dos métodos em outras áreas. Este experimento seria possível apenas através de um esforço conjunto entre os autores, para formalizar uma metodologia para avaliação e validação em um *framework* específico, o qual testaria todos os métodos em imagens com a mesma condição de iluminação, desordem a resolução, os quais são problemas constantes na área da segmentação.

Esta revisão com a análise dos métodos, experimentos realizados, avaliação de performance e perspectivas futuras, é útil para pesquisas que trabalham com imagens tomográficas e necessitam um método de segmentação 3D para uma melhor compreensão e discretização da imagem.

TABLE I
RESUMO DOS TRABALHOS ANALISADOS.

Autor	Título	Ano	Área	Método	Conhecimento a priori	Tipo de imagem
Huang	A TIGHTLY COUPLED REGION-SHAPE FRAMEWORK FOR 3D MEDICAL IMAGE SEGMENTATION	2006	Área Médica	3D Markov Random fields e Modelos deformáveis	Formato a priori dos objetos ou pontos sementes	Tomografia computadorizada
Monga	Defining and computing stable representations of volume shapes from discrete trace using volume primitives: Application to 3D image analysis in soil science	2006	Geologia	Crescimento de regiões com base em grafo	Não	Tomografia computadorizada
Saragaglia	Assessment of Airway Remodeling in Asthma: Volumetric Versus Surface Quantification Approaches	2006	Área médica	Modelos deformáveis, morfologia matemática e pareamento de contornos com base em energia	Não	Tomografia computadorizada
Davis	Assessment of 18F PET signals for automatic target volume definition in radiotherapy treatment planning	2006	Área médica	Threshold e crescimento de regiões	Informação de pontos iniciais pertencentes a região de interesse e ao plano de fundo	Positron Emission Tomography
Li	Automatic clinical image segmentation using pathological modeling, PCA and SVM	2006	Área médica	Principal component analysis, level set e Support vector machine	Treinamento	Tomografia computadorizada
Staal	Automatic rib segmentation and labeling in computed tomography scans using a general framework for detection, recognition and segmentation of objects in volumetric data	2006	Área médica	Threshold, crescimento de regiões e classificação	Treinamento	Tomografia computadorizada
Wieclawek	Live-Wire-Based 3D Segmentation Method	2007	Área médica	Live-wire, wavelet e Fuzzy C-Means	Escolha de um ponto semente para selecionar uma borda desejada	Tomografia computadorizada
Bulu	Comparison of 3D Segmentation Algorithms for Medical Imaging	2007	Área médica	Dois algoritmos de crescimento por regiões um com semente outro sem, Weibull E-SD Fields (WESDF) e Thresholding automático de múltiplos níveis utilizando o OTSU	Escolha de sementes para o crescimento de regiões	Tomografia computadorizada
Diciotti	3-D Segmentation Algorithm of Small Lung Nodules in Spiral CT Images	2008	Área médica	Crescimento de regiões e distância geodésica	Seleção de um voxel central para a delimitação do Volume de interesse	Tomografia computadorizada
Chen	ACCURATE AND FAST 3D COLON SEGMENTATION IN CT COLONOGRAPHY	2009	Área médica	Threshold e adaptative level set function	Informação de 4 sementes iniciais	Tomografia computadorizada

Aslan	SEGMENTATION OF TRABECULAR BONES FROM VERTEBRAL BODIES IN VOLUMETRIC CT SPINE IMAGES	2009	Área médica	Corte em grafos e crescimento de regiões	Localização da região vertebral	Tomografia computadorizada
Ding	3D Segmentation of Soft Organs by Flipping-Free Mesh Deformation	2009	Área médica	Modelos deformáveis e deformação de malhas	Não	Tomografia computadorizada
Chang	A New 3D Segmentation Algorithm Based On 3D PCNN For Lung CT Slices	2009	Área médica	Modelo de rede neural	Treinamento	Tomografia computadorizada
JIANG	Automatic 3D Segmentation of CT Images Based on Active Contour Models	2009	Área médica	Contornos ativos e threshold	Não	Tomografia computadorizada
Lai	Automatic 3D Segmentation of Lung Airway Tree: A Novel Adaptive Region Growing Approach	2009	Área médica	Crescimento de regiões adaptativo	Não	Tomografia computadorizada
Zhang	3D SEGMENTATION OF THE LIVER USING FREE-FORM DEFORMATION BASED ON BOOSTING AND DEFORMATION GRADIENTS	2009	Área médica	Modelos deformáveis	Conhecimento a priori de intensidade e formato	Tomografia computadorizada
Wang	Segmentation of Lung Nodules in Computed Tomography Images Using Dynamic Programming and Multidirection Fusion Techniques	2009	Área médica	Modelo de programação dinâmica	Treinamento	Tomografia computadorizada
Bert	An automatic method for colon segmentation in CT colonography	2009	Área médica	Threshold e crescimento por regiões 3D adaptativo	Conhecimento prévio dos valores de threshold adequados para o nicho aplicado	Tomografia computadorizada
Moussavi	3D segmentation of cell boundaries from whole cell cryogenic electron tomography volumes	2009	Histologia	Conditional random field associado um processo de inferência atualizado a cada iteração	Rotulação de uma fatia inicial	Cryon-electron tomography
Jun-Wei	A Novel Automatic Extraction Method of Lung Texture Tree from HRCT Images	2009	Área médica	Level set, modelo de contornos ativos guiado por função de energia	Não	Tomografia computadorizada
REN	A 3D Segmentation Method of Lung Parenchyma Based on CT Image Sequences	2010	Área médica	Thresholding adaptativo, Componente conexos, crescimento de regiões e operadores morfológicos	Não	Tomografia computadorizada
Aslan	A NOVEL, FAST, AND COMPLETE 3D SEGMENTATION OF VERTEBRAL BONES	2010	Área médica	Corte em grafos e crescimento de regiões	Não	Tomografia computadorizada
Gloger	A fully automatic three-step liver segmentation method on LDA-based probability maps for multiple contrast MR images	2010	Área médica	Threshold, Crescimento de regiões e mapas de probabilidade	Treinamento	Tomografia computadorizada

Badakhshanoory	A Model-Based Validation Scheme for Organ Segmentation in CT Scan Volumes	2011	Área médica	Principal component analysis e Construção de modelo	Não	Tomografia computadorizada
Cascio	Automatic detection of lung nodules in CT datasets based on stable 3D mass-spring models	2011	Área médica	Crescimento de regiões e Modelos deformáveis	Não	Tomografia computadorizada
Bhole	3D Segmentation in CT Imagery with Conditional Random Fields and Histograms of Oriented Gradients	2011	Área médica	Markov Random Fields, Conditional Random Fields e histogramas da orientação do gradiente	Iteração do usuário para o alinhamento inicial	Tomografia computadorizada
Müller	Model Based 3D Segmentation and OCT Image Undistortion of Percutaneous Implants	2011	Área da veterinária	Modelo com base em markov random fields	Formato dos implantes subcutâneos	Optical Coherence Tomography
Lu	Segmentation of the central-chest lymph nodes in 3D MDCT images	2011	Área médica	Deteção de borda e procura ótima em grafo com a análise de contornos ativos	Informação de pontos semente	Tomografia computadorizada
Chen	Three-Dimensional Segmentation of Fluid-Associated Abnormalities in Retinal OCT: Probability Constrained Graph-Search-Graph-Cut	2012	Área médica	Procura e corte em grafo, inicialização com uma pré-segmentação e classificação de voxels	Treinamento	Optical Coherence Tomography
Barbosa	B-Spline Explicit Active Surfaces: An Efficient Framework for Real-Time 3D Region-Based Segmentation	2012	Objetos em geral	Minimização de função de energia	Definição da origem do sistema de coordenadas	Tomografia computadorizada e Ultrassom
Liu	Fully Automatic 3D Segmentation of Iceball for Image-Guided Cryoablation	2012	Área médica	Corte em grafos	Informação a priori sobre formato dos objetos	Tomografia computadorizada e de ressonância magnética
Spampinato	Automatic 3D Segmentation of Mandible for Assessment of Facial Asymmetry	2012	Área médica	Componentes conexos e bit-plane slice coding	Definição de Região de interesse	Tomografia computadorizada
Uher	Automatic 3D Segmentation of Human Brain Images Using Data-mining Techniques	2012	Área médica	Crescimento de regiões	Treinamento	Tomografia computadorizada
Urschler	Forensic-Case Analysis: From 3D Imaging to Interactive Visualization	2012	Análise forense	Geodesic active countour e minimização de energia em um framework variacional	Conhecimento a priori sobre distribuição do valor de cinza e textura, restrição de formato	Tomografia computadorizada e de ressonância magnética
SHAOTING ZHANG	TOWARDS ROBUST AND EFFECTIVE SHAPE PRIOR MODELING: SPARSE SHAPE COMPOSITION	2012	Área médica	Segmentação com base em um conjunto de dados de formato	Treinamento	Tomografia computadorizada e de ressonância magnética

Loss	Automatic Segmentation and Quantification of Filamentous Structures in Electron Tomography	2012	Objetos em geral	Tensor voting e rastreamento de curvas	Não	Tomografia 3D com base em elétron
Hutter	Prior-Based Automatic Segmentation of the Carotid Artery Lumen in TOF MRA	2012	Área médica	Threshold adaptativo e level set	Conhecimentos a priori sobre padrões irregulares de fluxo sanguíneo	Angiografias por ressonância magnética
Delibasis	An Intelligent Tool for Anatomical Object Segmentation Using Deformable Surfaces	2012	Objetos em geral	Modelos deformáveis	Iteração inicial do usuário para inicialização das superfícies ativas	Tomografia computadorizada
Lloréns	Jaw tissues segmentation in dental 3D CT images using fuzzy-connectedness and morphological processing	2012	Área médica	Threshold e Fuzzy Connectedness Object Extraction	Não	Tomografia computadorizada
Rüegsegger	Statistical Modeling of the Eye for Multimodal Treatment Planning for External Beam Radiation Therapy of Intraocular Tumors	2012	Área médica	Modelo estatístico e active shape model	Treinamento	Tomografia computadorizada
Rusu	Automated tracing of filaments in 3D electron tomography reconstructions using Sculptor and Situs	2012	Área médica	Algoritmos genéticos e procura estocástica com base em templates	Pontos iniciais para o processo remoção de regiões sem filamentos	Cryon-electron tomography
Andrá	Digital rock physics benchmarks - Part I: Imaging and segmentation	2012	Geologia	3 - Abordagens (1 - threshold e watershed, 2 - OTSU e threshold manual, 3 - Filtro de suavização e threshold)	Não	Tomografia computadorizada
El-Zehiry	A SPLICE-GUIDED DATA DRIVEN INTERACTIVE EDITING	2013	Objetos geral	Grafo e framework de minimização de energia	Pré-segmentação, Dados e Pontos de entrada do usuário	Tomografia computadorizada e ressonância magnética
Qi	3D Segmentation of the Lung Based on the Neighbor Information and Curvature	2013	Área médica	Framework de minimização com base em Level set functions	Dados de vizinhança e informações com base em tons de cinza	Tomografia computadorizada
Werz	Particle tracking during Ostwald ripening using time-resolved laboratory X-ray microtomography	2013	Área de análise de materiais	Watershed e mapa de distância ao fundo	Não	Microtomografia computadorizada
Chaves	Integrating discretization and association rule-based classification for Alzheimer's disease diagnosis	2013	Área médica	Threshold com base em histograma e Association Rule mining	Treinamento	Single Photon Emission Computed Tomography (SPECT) e Positron Emission Tomography (PET)
Ontiverosa	Computer Tomography 3D edge detection comparative for metrology applications	2013	Área metrologia	Vários métodos	Variável conforme o método	Tomografia computadorizada
Shi	Greedy Algorithm Based Deformable Simplex Meshes Using Gradient Vector Flow as External Energy	2014	Área médica + geral	Modelos deformáveis	Não	Tomografia computadorizada

Paulano	3D segmentation and labeling of fractured bone from CT images	2014	Área médica	Crescimento de regiões	Sementes iniciais	Tomografia computadorizada
Santos	Automatic detection of small lung nodules in 3D CT data using Gaussian mixture models, Tsallis entropy and SVM	2014	Área médica	Support Vector Machine e threshold	Treinamento	Tomografia computadorizada
Czabaj	Numerical reconstruction of graphite/epoxy composite microstructure based on sub-micron resolution X-ray computed tomography	2014	Área médica	Template-matching e algoritmo de rastreamento	Seleção de templates	Microtomografia computadorizada
Rudyanto	Comparing algorithms for automated vessel segmentation in computed tomography scans of the lung: the VESSEL12 study	2014	Área médica	Plataforma de comparação de algoritmos de segmentação de vasos	Dependente do algoritmo analisado	Tomografia computadorizada
Badura	Soft computing approach to 3D lung nodule segmentation in CT	2014	Área médica	Thresholding por otsu e conectividade difusa	Informação de semente	Tomografia computadorizada
Jodoin	Left-Ventricle Segmentation of SPECT Images of Rats	2015	Área veterinária	Mudança de espaço e corte em grafo	Inserção do alvo no formato de T	Single Photon Emission Computed Tomography (SPECT)
Fabijanska	3D Segmentation of the Cerebrospinal Fluid from CT Brain Scans using Local Histogram Similarity Map	2015	Área médica	Similaridade entre histogramas por distância de Bhattacharyya e OTSU	Inserção de uma semente elíptica	Tomografia computadorizada
Mezlini	High resolution volume quantification of the knee joint space based on a semi-automatic segmentation of computed tomography images	2015	Área médica	Operadores morfológicos, threshold e contornos ativos	Seleção de um volume de interesse	Tomografia computadorizada com alta resolução periférica
Biesdorf	3D SEGMENTATION OF VESSELS BY INCREMENTAL IMPLICIT POLYNOMIAL FITTING AND CONVEX OPTIMIZATION	2015	Área médica	Ajuste de polinômio, otimização convexa e rastreamento ao longo da linha central	Pontos iniciais na linha central para a primeira fatia	Tomografia Computadorizada de Angiografia
Valente	Automatic 3D pulmonary nodule detection in CT images: A survey	2015	Área médica	Survey com vários métodos	Dependente do método sendo avaliado	Tomografia computadorizada
Kitrungsrotakul	Supervoxels based Graph Cut for Medical Organ Segmentation	2015	Área médica	Corte em grafo e supervoxel (clusterização)	Não	Tomografia computadorizada
Hemmat	Semi-automatic 3D segmentation of carotid lumen in contrast-enhanced computed tomography angiography images	2015	Área médica	Level set function	Seleção de uma região de interesse e de sementes	Tomografia Computadorizada de Angiografia
Gangsei	Automatic segmentation of Computed Tomography (CT) images of domestic pig skeleton using a 3D expansion of Dijkstra's algorithm	2015	Área da veterinária	Threshold e expansão 3D do algoritmo de Dijkstra	Conhecimento sobre orientação, dimensão dos ossos e pontos de referência conhecidos	Tomografia computadorizada
Javaid	A novel approach to CAD system for the detection of lung nodules in CT images	2016	Área médica	Thresholding, K-means e Support Vector Machine	Treinamento	Tomografia computadorizada
Gonçalves	Hessian based approaches for 3D lung nodule segmentation	2016	Área médica	Matrizes hessianas e threshold	Localização de um nódulo inicial	Tomografia computadorizada
Zeng	Liver vessel segmentation based on extreme learning machine	2016	Área médica	Extreme machine learning	Treinamento	Tomografia computadorizada

REFERENCES

- [1] R. C. Gonzalez and R. E. Woods, *Digital Image Processing (Second Edition)*. Upper Saddle River, NJ, USA: Prentice-Hall, Inc., 2006.
- [2] D. E. Ilea and P. F. Whelan, "Image segmentation based on the integration of colour-texture descriptors- a review," *Pattern Recognition*, vol. 44, no. 10-11, pp. 2479 – 2501, 2011, semi-Supervised Learning for Visual Content Analysis and Understanding.
- [3] B. Kitchenham, "Procedures for Performing Systematic Reviews," Joint Technical Report TR/SE-0401, Tech. Rep., 2004.
- [4] N. Health, M. R. C. (Australia), and N. Staff, *How to Review the Evidence: Systematic Identification and Review of the Scientific Literature*, ser. Handbook series on preparing clinical practice guidelines. National Health and Medical Research Council, 2000. [Online]. Available: <https://books.google.com.br/books?id=G8JmAAAACAAJ>
- [5] —, *How to Use the Evidence: Assessment and Application of Scientific Evidence*, ser. Handbook series on preparing clinical practice guidelines. National Health and Medical Research Council, 2000. [Online]. Available: <https://books.google.com.br/books?id=7ioKGQAACAAJ>
- [6] C. Collaboration, *Cochrane Reviewers' Handbook*, ser. Version 4.2.1. National Health and Medical Research Council, 2003.
- [7] U. of York. NHS Centre for Reviews and Dissemination, *Undertaking systematic reviews of research on effectiveness: CRD's guidance for those carrying out or commissioning reviews*, ser. CRD report. NHS Centre for Reviews and Dissemination, University of York, 2001. [Online]. Available: <https://books.google.com.br/books?id=2rhpAAAAMAAJ>
- [8] M. Javaid, M. Javid, M. Z. U. Rehman, and S. I. A. Shah, "A novel approach to {CAD} system for the detection of lung nodules in {CT} images," *Computer Methods and Programs in Biomedicine*, vol. 135, pp. 125 – 139, 2016.
- [9] W. Wiclawek and E. Pietka, "Live-wire-based 3d segmentation method," in *2007 29th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, Aug 2007, pp. 5645–5648.
- [10] M. S. Aslan, A. Ali, B. Arnold, R. Fahmi, A. A. Farag, and P. Xiang, "Segmentation of trabecular bones from vertebral bodies in volumetric ct spine images," in *2009 16th IEEE International Conference on Image Processing (ICIP)*, Nov 2009, pp. 3385–3388.
- [11] M. S. Aslan, A. Ali, H. Rara, B. Arnold, R. Fahmi, A. A. Farag, and P. Xiang, "A novel, fast, and complete 3d segmentation of vertebral bones," in *2010 IEEE International Conference on Acoustics, Speech and Signal Processing*, March 2010, pp. 654–657.
- [12] K. Lu and W. E. Higgins, "Segmentation of the central-chest lymph nodes in 3d {MDCT} images," *Computers in Biology and Medicine*, vol. 41, no. 9, pp. 780 – 789, 2011.
- [13] X. Liu, K. Tuncali, W. M. Wells, P. R. Morrison, and G. P. Zientara, "Fully automatic 3d segmentation of iceball for image-guided cryoablation," in *2012 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, Aug 2012, pp. 2327–2330.
- [14] Y. Boykov and V. Kolmogorov, "An experimental comparison of min-cut/max- flow algorithms for energy minimization in vision," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 26, no. 9, pp. 1124–1137, Sept 2004.
- [15] X. Chen, M. Niemeijer, L. Zhang, K. Lee, M. D. Abramoff, and M. Sonka, "Three-dimensional segmentation of fluid-associated abnormalities in retinal oct: Probability constrained graph-search-graph-cut," *IEEE Transactions on Medical Imaging*, vol. 31, no. 8, pp. 1521–1531, Aug 2012.
- [16] N. El-Zehiry, M. P. Jolly, and M. Sofka, "A splice-guided data driven interactive editing," in *2013 IEEE 10th International Symposium on Biomedical Imaging*, April 2013, pp. 1098–1101.
- [17] P. M. Jodoin, F. Pinheiro, A. Oudot, and A. Lalonde, "Left-ventricle segmentation of spect images of rats," *IEEE Transactions on Biomedical Engineering*, vol. 62, no. 9, pp. 2260–2268, Sept 2015.
- [18] T. Kitrungratsakul, Y.-W. Chen, X.-H. Han, and L. Lin, "Supervoxels based graph cut for medical organ segmentation," *IFAC-PapersOnLine*, vol. 48, no. 20, pp. 70 – 75, 2015, 9th {IFAC} Symposium on Biological and Medical Systems {BMS} 2015 Berlin, Germany, 31 August-2 September 2015.
- [19] L. E. Gangsei and J. Kongsro, "Automatic segmentation of computed tomography (ct) images of domestic pig skeleton using a 3d expansion of dijkstra's algorithm," *Computers and Electronics in Agriculture*, vol. 121, pp. 191 – 194, 2016. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0168169915003841>
- [20] J.-W. LIU, H.-Q. FENG, Y.-Y. ZHOU, and C.-F. LI, "A novel automatic extraction method of lung texture tree from {HRCT} images," *Acta Automatica Sinica*, vol. 35, no. 4, pp. 345 – 349, 2009. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S1874102908600816>
- [21] D. Chen, R. Fahmi, A. A. Farag, R. L. Falk, and G. W. Dryden, "Accurate and fast 3d colon segmentation in ct colonography," in *2009 IEEE International Symposium on Biomedical Imaging: From Nano to Macro*, June 2009, pp. 490–493.
- [22] D. Chen, A. A. Farag, R. L. Falk, and G. W. Dryden, "A variational framework for 3d colonic polyp visualization in virtual colonoscopy," in *2009 16th IEEE International Conference on Image Processing (ICIP)*, Nov 2009, pp. 2617–2620.
- [23] J. Hutter, H. G. Hofmann, R. Grimm, A. Greiser, M. Saake, J. Hornegger, A. Dörfler, and P. Schmitt, "Prior-based automatic segmentation of the carotid artery lumen in tof mra (pascal)," in *Proceedings of the 15th International Conference on Medical Image Computing and Computer-Assisted Intervention - Volume Part II*, ser. MICCAI'12. Berlin, Heidelberg: Springer-Verlag, 2012, pp. 511–518.
- [24] Y. Qi, K. Dong, L. Yin, and M. Li, "3d segmentation of the lung based on the neighbor information and curvature," in *Proceedings of the 2013 Seventh International Conference on Image and Graphics*, ser. ICIG '13. Washington, DC, USA: IEEE Computer Society, 2013, pp. 139–143. [Online]. Available: <http://dx.doi.org/10.1109/ICIG.2013.34>
- [25] H. Hemmati, A. Kamli-Asl, A. Talebpour, and S. Shirani, "Semi-automatic 3d segmentation of carotid lumen in contrast-enhanced computed tomography angiography images," *Physica Medica*, vol. 31, no. 8, pp. 1098 – 1104, 2015.
- [26] R. Huang, V. Pavlovic, and D. N. Metaxas, "A tightly coupled region-shape framework for 3d medical image segmentation," in *3rd IEEE International Symposium on Biomedical Imaging: Nano to Macro, 2006.*, April 2006, pp. 426–429.
- [27] O. Müller, S. Donner, T. Klinder, R. Dragon, I. Bartsch, F. Witte, A. Krüger, A. Heisterkamp, and B. Rosenhahn, "Model based 3d segmentation and oct image undistortion of percutaneous implants," in *Proceedings of the 14th International Conference on Medical Image Computing and Computer-assisted Intervention - Volume Part III*, ser. MICCAI'11. Berlin, Heidelberg: Springer-Verlag, 2011, pp. 454–462. [Online]. Available: <http://dl.acm.org/citation.cfm?id=2044839.2044900>
- [28] C. Bhole, N. Morsillo, and C. Pal, "3d segmentation in ct imagery with conditional random fields and histograms of oriented gradients," in *Proceedings of the Second International Conference on Machine Learning in Medical Imaging*, ser. MLMI'11. Berlin, Heidelberg: Springer-Verlag, 2011, pp. 326–334. [Online]. Available: <http://dl.acm.org/citation.cfm?id=2046063.2046103>
- [29] A. Saragaglia, C. Fetita, and F. Prêteux, "Assessment of airway remodeling in asthma: Volumetric versus surface quantification approaches," in *Proceedings of the 9th International Conference on Medical Image Computing and Computer-Assisted Intervention - Volume Part II*, ser. MICCAI'06. Berlin, Heidelberg: Springer-Verlag, 2006, pp. 413–420.
- [30] H. Zhang, L. Yang, D. J. Foran, J. L. Noshier, and P. J. Yim, "3d segmentation of the liver using free-form deformation based on boosting and deformation gradients," in *Proceedings of the Sixth IEEE International Conference on Symposium on Biomedical Imaging: From Nano to Macro*, ser. ISBI'09. Piscataway, NJ, USA: IEEE Press, 2009, pp. 494–497. [Online]. Available: <http://dl.acm.org/citation.cfm?id=1699872.1699997>
- [31] F. Ding, W. Yang, W. K. Leow, and S. K. Venkatesh, "3d segmentation of soft organs by flipping-free mesh deformation," in *Applications of Computer Vision (WACV), 2009 Workshop on*, Dec 2009, pp. 1–7.
- [32] D. Cascio, R. Magro, F. Fauci, M. Iacomi, and G. Raso, "Automatic detection of lung nodules in ct datasets based on stable 3d mass-spring models," *Comput. Biol. Med.*, vol. 42, no. 11, pp. 1098–1109, nov 2012. [Online]. Available: <http://dx.doi.org/10.1016/j.combiomed.2012.09.002>
- [33] K. K. Delibasis, A. Christodoulidis, and I. Maglogiannis, "An intelligent tool for anatomical object segmentation using deformable surfaces," in *Proceedings of the 7th Hellenic Conference on Artificial Intelligence: Theories and Applications*, ser. SETN'12. Berlin, Heidelberg: Springer-Verlag, 2012, pp. 206–213.
- [34] C. Shi, C. Guo, Y. Cheng, and J. Wang, "Greedy algorithm based deformable simplex meshes using gradient vector flow as external energy," in *2014 7th International Conference on Biomedical Engineering and Informatics*, Oct 2014, pp. 199–204.
- [35] H. Jiang and Q. Cheng, "Automatic 3d segmentation of ct images based on active contour models," in *Computer-Aided Design and Computer Graphics, 2009. CAD/Graphics '09. 11th IEEE International Conference on*, Aug 2009, pp. 540–543.

- [36] D. Barbosa, T. Diertenbeck, J. Schaerer, J. D'hooge, D. Friboulet, and O. Bernard, "B-spline explicit active surfaces: An efficient framework for real-time 3-d region-based segmentation," *IEEE Transactions on Image Processing*, vol. 21, no. 1, pp. 241–251, Jan 2012.
- [37] Q. Duan, E. D. Angelini, and A. F. Laine, "Real-time segmentation by active geometric functions," *Comput. Methods Prog. Biomed.*, vol. 98, no. 3, pp. 223–230, Jun. 2010. [Online]. Available: <http://dx.doi.org/10.1016/j.cmpb.2009.09.001>
- [38] M. Urschler, A. Bornik, E. Scheurer, K. Yen, H. Bischof, and D. Schmalstieg, "Forensic-case analysis: From 3d imaging to interactive visualization," *IEEE Comput. Graph. Appl.*, vol. 32, no. 4, pp. 79–87, Jul. 2012. [Online]. Available: <http://dx.doi.org/10.1109/MCG.2012.75>
- [39] H. Mezlini, R. Youssef, H. Bouhadoun, E. Budyn, J. D. Laredo, S. S. Ghalila, and C. Chappard, "High resolution volume quantification of the knee joint space based on a semi-automatic segmentation of computed tomography images," in *2015 International Conference on Systems, Signals and Image Processing (IWSSIP)*, Sept 2015, pp. 157–161.
- [40] F. Moussavi, G. Heitz, F. Amat, L. R. Comolli, D. Koller, and M. Horowitz, "3d segmentation of cell boundaries from whole cell cryogenic electron tomography volumes," *Journal of Structural Biology*, vol. 170, no. 1, pp. 134 – 145, 2010.
- [41] H. Badakhshanoory and P. Saedi, "A model-based validation scheme for organ segmentation in ct scan volumes," *IEEE Transactions on Biomedical Engineering*, vol. 58, no. 9, pp. 2681–2693, Sept 2011.
- [42] S. Zhang, "Towards robust and effective shape prior modeling: Sparse shape composition," Ph.D. dissertation, New Brunswick, NJ, USA, 2012, aAI3502515.
- [43] M. B. Rügsegger, M. B. Cuadra, A. Pica, C. A. Amstutz, T. Rudolph, D. Aebersold, and J. H. Kowal, "Statistical modeling of the eye for multimodal treatment planning for external beam radiation therapy of intraocular tumors," *International Journal of Radiation Oncology*Biophysics*Physics*, vol. 84, no. 4, pp. e541 – e547, 2012.
- [44] J. B. Davis, B. Reiner, M. Huser, C. Burger, G. Székely, and I. F. Ciernik, "Assessment of 18f {PET} signals for automatic target volume definition in radiotherapy treatment planning," *Radiotherapy and Oncology*, vol. 80, no. 1, pp. 43 – 50, 2006.
- [45] J. Staal, B. van Ginneken, and M. A. Viergever, "Automatic rib segmentation and labeling in computed tomography scans using a general framework for detection, recognition and segmentation of objects in volumetric data," *Medical Image Analysis*, vol. 11, no. 1, pp. 35 – 46, 2007. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S1361841506000788>
- [46] O. Monga, "Defining and computing stable representations of volume shapes from discrete trace using volume primitives: Application to 3d image analysis in soil science," *Image Vision Comput.*, vol. 25, no. 7, pp. 1134–1153, jul 2007. [Online]. Available: <http://dx.doi.org/10.1016/j.imavis.2006.06.012>
- [47] H. Bulu and A. Alpkocak, "Comparison of 3d segmentation algorithms for medical imaging," in *Twentieth IEEE International Symposium on Computer-Based Medical Systems (CBMS'07)*, June 2007, pp. 269–274.
- [48] S. Diciotti, G. Picozzi, M. Falchini, M. Mascacchi, N. Villari, and G. Valli, "3-d segmentation algorithm of small lung nodules in spiral ct images," *IEEE Transactions on Information Technology in Biomedicine*, vol. 12, no. 1, pp. 7–19, Jan 2008.
- [49] K. Lai, P. Zhao, Y. Huang, J. Liu, C. Wang, H. Feng, and C. Li, "Automatic 3d segmentation of lung airway tree: A novel adaptive region growing approach," in *2009 3rd International Conference on Bioinformatics and Biomedical Engineering*, June 2009, pp. 1–4.
- [50] A. Bert, I. Dmitriev, S. Agliozzo, N. Pietrosemoli, M. Mandelkern, T. Gallo, and D. Regge, "An automatic method for colon segmentation in {CT} colonography," *Computerized Medical Imaging and Graphics*, vol. 33, no. 4, pp. 325 – 331, 2009.
- [51] O. Gloger, J. Kühn, A. Stanski, H. Völzke, and R. Puls, "A fully automatic three-step liver segmentation method on lda-based probability maps for multiple contrast {MR} images," *Magnetic Resonance Imaging*, vol. 28, no. 6, pp. 882 – 897, 2010.
- [52] Y. h. Ren, X. w. Sun, and S. d. Nie, "A 3d segmentation method of lung parenchyma based on ct image sequences," in *2010 International Conference on Information, Networking and Automation (ICINA)*, vol. 2, Oct 2010, pp. V2–332–V2–336.
- [53] V. Uher and R. Burget, "Automatic 3d segmentation of human brain images using data-mining techniques," in *Telecommunications and Signal Processing (TSP), 2012 35th International Conference on*, July 2012, pp. 578–580.
- [54] H. Andrä, N. Combaret, J. Dvorkin, E. Glatt, J. Han, M. Kabel, Y. Keehm, F. Krzikalla, M. Lee, C. Madonna, M. Marsh, T. Mukerji, E. H. Saenger, R. Sain, N. Saxena, S. Ricker, A. Wiegmann, and X. Zhan, "Digital rock physics benchmarks-part i: Imaging and segmentation," *Computers & Geosciences*, vol. 50, pp. 25 – 32, 2013, benchmark problems, datasets and methodologies for the computational geosciences.
- [55] P. Badura and E. Pietka, "Soft computing approach to 3d lung nodule segmentation in {CT};" *Computers in Biology and Medicine*, vol. 53, pp. 230 – 243, 2014.
- [56] T. Werz, M. Baumann, U. Wolfram, and C. K. III, "Particle tracking during ostwald ripening using time-resolved laboratory x-ray microtomography," *Materials Characterization*, vol. 90, pp. 185 – 195, 2014.
- [57] F. Paulano, J. J. Jiménez, and R. Pulido, "3d segmentation and labeling of fractured bone from ct images," *Vis. Comput.*, vol. 30, no. 6-8, pp. 939–948, jun 2014. [Online]. Available: <http://dx.doi.org/10.1007/s00371-014-0963-0>
- [58] S. Li, T. Fevens, A. Krzyżak, and S. Li, "Automatic clinical image segmentation using pathological modeling, {PCA} and {SVM};" *Engineering Applications of Artificial Intelligence*, vol. 19, no. 4, pp. 403 – 410, 2006, recent Advances in Data Mining.
- [59] Q. Chang, J. Shi, and Z. Xiao, "A new 3d segmentation algorithm based on 3d pcnn for lung ct slices," in *2009 2nd International Conference on Biomedical Engineering and Informatics*, Oct 2009, pp. 1–5.
- [60] A. M. Santos, A. O. de Carvalho Filho, A. C. Silva, A. C. de Paiva, R. A. Nunes, and M. Gattass, "Automatic detection of small lung nodules in 3d {CT} data using gaussian mixture models, tsallis entropy and {SVM};" *Engineering Applications of Artificial Intelligence*, vol. 36, pp. 27 – 39, 2014.
- [61] Y. Z. Zeng, Y. Q. Zhao, M. Liao, B. J. Zou, X. F. Wang, and W. Wang, "Liver vessel segmentation based on extreme learning machine," *Physica Medica*, vol. 32, no. 5, pp. 709 – 716, 2016. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S1120179716300187>
- [62] R. Chaves, J. Ramírez, and J. Górriz, "Integrating discretization and association rule-based classification for alzheimer's disease diagnosis," *Expert Systems with Applications*, vol. 40, no. 5, pp. 1571 – 1578, 2013.
- [63] C. Spampinato, C. Pino, D. Giordano, and R. Leonardi, "Automatic 3d segmentation of mandible for assessment of facial asymmetry," in *Medical Measurements and Applications Proceedings (MeMeA), 2012 IEEE International Symposium on*, May 2012, pp. 1–4.
- [64] Q. Wang, E. Song, R. Jin, P. Han, X. Wang, Y. Zhou, and J. Zeng, "Segmentation of lung nodules in computed tomography images using dynamic programming and multidirection fusion techniques1," *Academic Radiology*, vol. 16, no. 6, pp. 678 – 688, 2009.
- [65] M. Rusu, Z. Starosolski, M. Wahle, A. Rigort, and W. Wriggers, "Automated tracing of filaments in 3d electron tomography reconstructions using sculptor and situs," *Journal of Structural Biology*, vol. 178, no. 2, pp. 121 – 128, 2012, special Issue: Electron Tomography.
- [66] R. Lloréns, V. Naranjo, F. López, and M. Alcañiz, "Jaw tissues segmentation in dental 3d {CT} images using fuzzy-connectedness and morphological processing," *Computer Methods and Programs in Biomedicine*, vol. 108, no. 2, pp. 832 – 843, 2012.
- [67] L. Gonçalves, J. Novo, and A. Campilho, "Hessian based approaches for 3d lung nodule segmentation," *Expert Systems with Applications*, vol. 61, pp. 1 – 15, 2016. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0957417416302494>
- [68] A. Fabijańska and J. Goćlawski, "3d segmentation of the cerebrospinal fluid from ct brain scans using local histogram similarity map," in *Cybernetics (CYBCONF), 2015 IEEE 2nd International Conference on*, June 2015, pp. 113–118.
- [69] A. Biesdorf, S. Wörz, H. von Tengg-Kobligh, K. Rohr, and C. Schnörr, "3d segmentation of vessels by incremental implicit polynomial fitting and convex optimization," in *2015 IEEE 12th International Symposium on Biomedical Imaging (ISBI)*, April 2015, pp. 1540–1543.
- [70] M. W. Czabaj, M. L. Riccio, and W. W. Whitacre, "Numerical reconstruction of graphite/epoxy composite microstructure based on sub-micron resolution x-ray computed tomography," *Composites Science and Technology*, vol. 105, pp. 174 – 182, 2014.
- [71] L. A. Loss, G. Bebis, H. Chang, M. Auer, P. Sarkar, and B. Parvin, "Automatic segmentation and quantification of filamentous structures in electron tomography," in *Proceedings of the ACM Conference on Bioinformatics, Computational Biology and Biomedicine*, ser. BCB '12. New York, NY, USA: ACM, 2012, pp. 170–177. [Online]. Available: <http://doi.acm.org/10.1145/2382936.2382958>
- [72] S. Ontiveros, J. Yagüe, R. Jiménez, and F. Brosted, "Computer tomography 3d edge detection comparative for metrology applications," *Procedia Engineering*, vol. 63, pp. 710 – 719, 2013, the Manufacturing Engineering Society International Conference, {MESIC} 2013.
- [73] R. D. Rudyanto and et al, "Comparing algorithms for automated vessel segmentation in computed tomography scans of the lung: the

- {VESSEL12} study,” *Medical Image Analysis*, vol. 18, no. 7, pp. 1217–1232, 2014.
- [74] I. R. S. Valente, P. C. Cortez, E. C. Neto, J. M. Soares, V. H. C. de Albuquerque, and J. a. M. R. Tavares, “Automatic 3d pulmonary nodule detection in ct images,” *Comput. Methods Prog. Biomed.*, vol. 124, no. C, pp. 91–107, feb 2016. [Online]. Available: <http://dx.doi.org/10.1016/j.cmpb.2015.10.006>
- [75] L. R. Dice, “Measures of the Amount of Ecologic Association Between Species,” *Ecology*, vol. 26, no. 3, pp. 297–302, Jul. 1945.
- [76] A. Ben-Hur, A. Elisseeff, and I. Guyon, “A stability based method for discovering structure in clustered data.” *Pacific Symposium on Biocomputing. Pacific Symposium on Biocomputing*, pp. 6–17, 2002.
- [77] A. A. Taha and A. Hanbury, “Metrics for evaluating 3d medical image segmentation: analysis, selection, and tool,” *BMC Medical Imaging*, vol. 15, no. 1, p. 29, 2015.

**ANEXO C – RELATÓRIO TÉCNICO - REVISÃO DA LITERATURA PARA
RECONHECIMENTO/CLASSIFICAÇÃO DE OBJETOS 3D**



INCoD

Instituto Nacional para Convergência Digital



UNIVERSIDADE FEDERAL
DE SANTA CATARINA

ISSN 2236-5281

Relatório Técnico

INCoD/LAPIX.01.2017.P.v01

Revisão da literatura para reconhecimento/classificação de objetos 3D

Autores:

Luis Eduardo R. De Carvalho

Aldo von Wangenheim

Versão 1.0

Status: Final

Distribuição: Externa

SETEMBRO - 2017

© 2017 **INCoD** – Instituto Nacional para Convergência Digital

Todos os direitos reservados e protegidos pela Lei 9.610 de 19/02/1998. Nenhuma parte deste documento, sem autorização prévia por escrito do Instituto, poderá ser reproduzida ou transmitida sejam quais forem os meios empregados: eletrônicos, mecânicos, fotográficos, gravação ou quaisquer outros.



Instituto Nacional para Convergência Digital

Universidade Federal de Santa Catarina - UFSC
Campus Universitário João David Ferreira Lima - Trindade
Departamento de Informática e Estatística - Sala 320
Florianópolis-SC - CEP 88040-970

Fone / FAX: +55 48 3721-9516 R.17

www.incod.ufsc.br

ISSN 2236-5281

Relatório Técnico do Instituto Nacional para Convergência Digital/
Departamento de Informática e Estatística, Centro Tecnológico,
Universidade Federal de Santa Catarina. -- v.1, n.1 (2017).--
Florianópolis: INE, UFSC, 2017 -

Semestral
Resumo em português

ISSN 2236-5281

1. Convergência digital. 2. Reconhecimento de padrões.
3. Reconhecimento de objetos 3D. 4. Revisão sistemática I.
Universidade Federal de Santa Catarina. Departamento de
Informática e Estatística.

Revisão da literatura para reconhecimento/classificação de objetos 3D

Carvalho, L.E. et al, e von Wangenheim, A

I. INTRODUÇÃO

A área de reconhecimento/classificação de objetos 3D é uma área que nos últimos anos teve um crescimento impulsionado pela maior disponibilidade de bases de dados com dados de objetos 3D e da popularização de sensores para captura de objetos em um ambiente. A aplicabilidade destes métodos vão desde o domínio da robótica, voltada para movimentação de robôs em ambientes e manipulação, por braços robóticos, de objetos até o domínio de segurança, onde utilizam-se as técnicas de reconhecimento/classificação de objetos 3D para detectar possíveis objetos proibidos em um avião. Visando ter um melhor conhecimento sobre o assunto, identificando métodos de classificação/reconhecimento, e, por consequência, possíveis formas de descrição de objetos, elaborou-se uma revisão sistemática da literatura. Este documento tem como propósito detalhar todos os procedimentos relacionados para a revisão sistemática da literatura, para o tópico de classificação/reconhecimento de objetos 3D, para o período de 2006-2016, e realizar uma análise do estado da arte sobre o assunto, com o foco nas possíveis aplicações destes métodos em outras áreas.

Este relatório técnico está dividido nas seguintes seções. A seção de metodologia, define qual foram os passos executados para a realização deste trabalho. A seção de descrição da pesquisa, detalha os parâmetros de pesquisa utilizados bem como critérios de seleção empregados na seleção dos trabalhos a serem analisados. A seção de trabalhos realizados, categoriza e detalha os trabalhos selecionados para análise, com foco na forma de descrição e reconhecimento/classificação dos objetos. A análise e discussão sobre os trabalhos selecionados é realizada na seção de discussão sobre os trabalhos analisados. Por fim,

II. METODOLOGIA

A metodologia utilizada neste trabalho teve como base a metodologia proposta por Kitchenham[1] para conduzir o processo de pesquisa. Esta metodologia, para análise sistemática da literatura, tem como objetivo, como descrito pela própria autora, avaliar e interpretar todas as pesquisas relevantes disponíveis à uma particular pergunta de pesquisa, tópico ou fenômeno de interesse. Três guias utilizados para pesquisa na área da saúde serviram como base para a metodologia de Kitchenham [2] [3] [4] [5]. A principal característica ou propriedade desta metodologia é a de manter a procura reproduzível, quando todos os passos descritos forem repetidos e as mesmas palavras-chave forem utilizadas. Depois, para a

análise e categorização dos resultados, utilizaram-se como referência a categorização dos tipos de representação de objetos 3D apresentados por Alaoui [6] e Atmosukarto [7], estendendo a forma de agrupamento a medida que novos métodos de representação apareceram.

III. DESCRIÇÃO DA PESQUISA

Inicialmente, com base em Kitchenham[1], definiu-se a tópico de pesquisa, **Métodos de reconhecimento/classificação de objetos 3D**. Com base neste tópico, definiram-se palavras chave e utilizou-se três bases de artigos científicos (ScienceDirect, ACM e IEEE) para a pesquisa. As palavras-chave utilizadas em cada base e os filtros de pesquisa foram:

ScienceDirect: *2017 < pub-date and pub-date > 2005 and TITLE-ABSTR-KEY("3D object classification") or TITLE-ABSTR-KEY("3D object recognition")*

ACM: *"3D object recognition" "3D object classification" published after 2006 until 2016*

IEEE: *("3D object recognition") OR "3D object classification" and refined by Year: 2006-2016*

Obteve-se um total de 446 artigos que foram analisados os *abstracts*. Nesta análise utilizaram-se os seguintes **critérios de exclusão** afim de selecionar os trabalhos a serem analisados:

- Artigos escritos em outras línguas diferentes da inglesa;
- Artigos repetidos e que não descrevessem nos *abstracts* uma técnica para reconhecimento/classificação de objetos 3D.
- Artigos que apenas fazem uso de softwares para a parte de reconhecimento/classificação de objetos e/ou não possuem foco na técnica de reconhecimento/classificação de objetos 3D apenas citando o seu uso.

Com a aplicação do critério de exclusão, obteve-se um total de 277 artigos que foram lidos na íntegra. Destes 277, 10 trabalhos são referentes a livros, apresentação de bases de dados 3D e revisões de métodos e tipos de classificação e, portanto, foram separados do processo de categorização segundo o tipo de representação de objetos 3D sendo utilizado.

IV. TRABALHOS ANALISADOS

A. Representação com base em características

A abordagem proposta por Pichler [8] visa a solução do problema de consumo de tempo nas linhas de montagem de carros, mais especificamente na parte de pintura dos carros, com a utilização de um braço robótico. No esquema proposto, um dos passos para obter uma melhor visualização das peças do carro, sombreadas nos sensores, é o reconhecimento de

objetos 3D. Nesta etapa tem-se como base de representação dos objetos o método de spin images e o sistema utiliza um esquema de reconhecimento com base em modelos 3D, no qual busca-se a correspondência entre spin images dos pontos do modelo e da cena. Mais especificamente, buscando reduzir o tempo necessário pelo método de reconhecimento, utilizou-se um técnica para reconhecimento de pontos chave nas representações por spin images com a finalidade de achar os pontos mais representativos dos modelo 3D e dos objetos na cena. Desta forma tendo calculado os pontos chaves dos modelos 3D eles são indexados em uma base de dados dos modelos. Depois desse processo inicial de representação dos modelos por pontos chave, o próximo passo é, para cada cena de entrada, descrever os objetos em pontos chave e buscar na base de dados do modelo os pontos chaves que tenham a maior semelhança com os pontos chaves da cena. Esse passo é realizado com a seleção randômica de pontos 3D da cena e os vizinhos mais próximos são avaliados na base de dados de modelos. Se a distância entre um ponto da cena e do modelo for menor que um threshold o modelo correspondente ganha um voto e o modelo com o maior número de votos é selecionado. Em uma etapa de verificação testam-se todos os modelos selecionados em relação à cena através da correspondência entre pontos da cena rotulados e os respectivos pontos nos modelos 3D utilizando um algoritmo iterativo de Closest Point. A distância de Hausdorff entre os pontos da cena e do modelo foi utilizado como medida da qualidade do resultado de reconhecimento. O sistema de reconhecimento foi testado em um conjunto de 25 objetos referente a 4 objetos industriais pendurados em várias posições. Na aplicação em questão, o propósito principal do reconhecimento de objetos é rotular, de forma robusta, pontos 3D na cena com os números referentes aos modelos existentes na base de dados e no passo subsequente fazer a correspondência dos pontos da cena e modelo afim de obter a orientação e posição dos modelos CAD em relação ao mundo, os quais são de extrema importância para uma aplicação robótica.

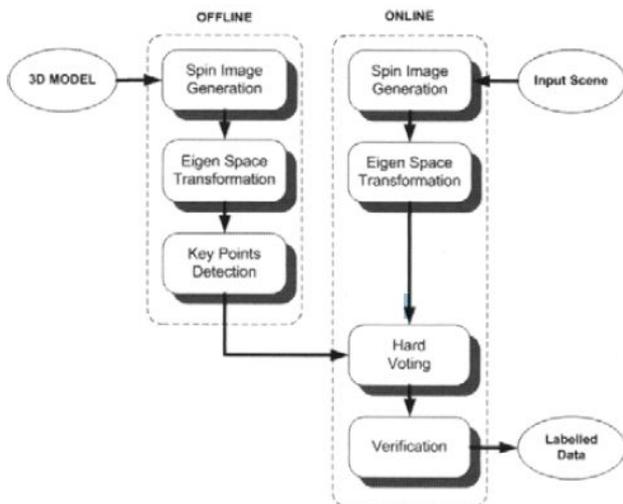


Fig. 1. Representação do método proposto por Pichler. Figura extraída de [8].

Wei-Yang Lin [9] propõe um *framework* para *summation invariants*, juntamente com quatro classes importantes de *summation invariants*. As características das invariantes propostas são a robustez na presença de ruído e erro de quantização. Estas invariantes são utilizadas para definir uma representação de formato para várias aplicações. Uma delas é o reconhecimento de objetos 3D. Nesta aplicação, deriva-se as *summation invariants* das superfícies sob ambas as transformações, euclidiana e afim, e propõe-se um algoritmo para reconhecimento de faces 3D com base nessas invariantes. O algoritmo proposto utiliza imagens de profundidade normalizada, extraindo das mesmas as *summation invariants* semi-locais e usa os resultados obtidos como características invariantes. Em cada pixel, uma *summation invariant* semi-local é computada de uma janela local. Computam-se invariantes de curva tanto a partir de uma janela horizontal quanto a partir de uma janela vertical. Ao invés de usar a imagem inteira, o autor extrai as características invariantes de uma região retangular centrada na ponta do nariz. As *summation invariants* computadas da região retangular foram denominadas como *summation image*. Para todos os dados de profundidade, computa-se as *summation image* da região inicial e procura-se as suas imagens média, denominada *mean of summation images*. Então, para cada dado de profundidade normalizado, procura-se uma nova região a qual tenha o mínimo sum of squared differences (SSD) com a *mean of summation images* e computa-se uma *summation invariant* da nova região. Com a finalidade de reduzir o espaço do vetor de característica, utiliza-se o principal component analysis (PCA). Existem muitas formas de combinar múltiplos classificadores para alcançar uma melhor performance de reconhecimento. No trabalho apresentado, o autor utilizou a regra da soma para combinar a pontuação de correspondência produzido pelas diferentes características invariantes. Os experimentos com o algoritmo proposto foram realizados na base de dados the Face Recognition Grand Challenge (FRGC)[10] e os resultados avaliados em termos de Equal Error Rate (EER) e da curva Receiver Operating Characteristics (ROC), comparando o método proposto com o algoritmo base do FRGC 3D.

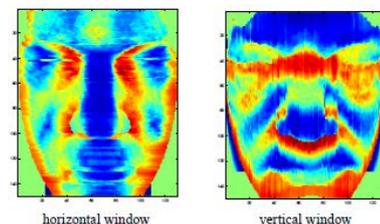


Fig. 2. Representação do método proposto por Wei-Yang Lin. Figura extraída de [9].

Shivaswamy [11] propõe uma extensão do Support Vector Machine (SVM) de forma a tornar o classificador invariante a permutação de subelementos de cada entrada, demonstrando a aplicabilidade do método em experimentos de reconhecimento de caracteres, reconhecimento de objetos 3D e em várias bases de dados UCI. Este método tem como base os bem conhecidos limites que relacionam a probabilidade de erro

com a razão do raio dos dados e a margem de separação. Este critério de aprendizagem é alcançado através da intercalação do algoritmo clássico de Kuhn-Munkres [12] no processo de aprendizagem discriminativo de margens largas do SVM. O método de Kuhn-Munkres serve para solucionar o problema central da técnica proposta, o Linear Assignment Problem (LAP). Logo, a extensão proposta denominada Permutation Invariant SVM, pode ser descrita pelos seguintes passos: inicialmente, dado um conjunto de treinamento e um número máximo de iterações, encontram-se o centroide, o raio da base de dados e o hiperplano com a maior margem. Depois, utiliza-se o algoritmo de Kuhn-Munkres para encontrar a permutação para cada padrão de exemplo que o traga para perto do centroide da esfera, enquanto assegurando que a sua margem do limite de decisão apenas aumenta. Desse modo, movem-se os dados, para assegurar que o raio pode ser apenas diminuído e a margem pode ser apenas aumentada. Então, recomputa-se o centroide, o hiperplano e os novos raio e margem nos dados permutados e repete-se o processo previamente descrito. Este processo é repetido até alcançar o número máximo de iterações. Para demonstrar a eficácia do método, em termos de classificação de objetos 3D, elaboraram-se 3 experimentos para a classificação de três pares de imagens da base de dados Amsterdam Library of Object Images [13]. O primeiro experimento utilizou imagens de aviões e dardos, o segundo utilizou imagens de dois vasos com aparências quase idênticas e o terceiro utilizou duas bonecas. Os resultados obtidos para cada experimento foram comparados com outros métodos clássicos de SVM com *kernels* gaussianos e *kernels* lineares.

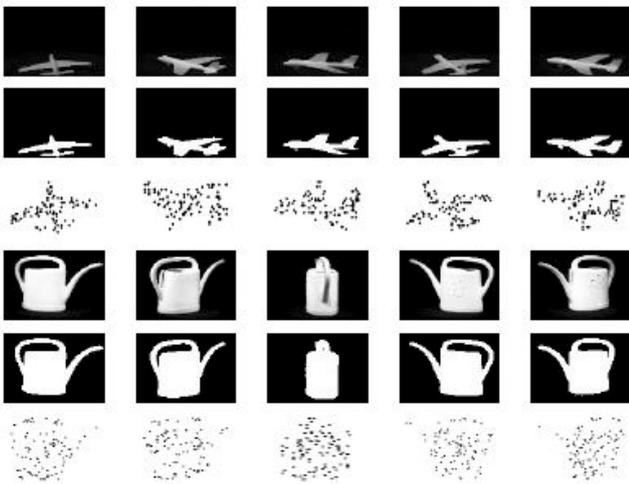


Fig. 3. Representação do método proposto por Shivaswamy. Figura extraída de [11].

Um método que explora a similaridade de modelos no reconhecimento de objetos 3D com base em modelos é proposto por Yi Tan [14]. Em cenários que consistem em grandes bases de dados de modelos 3D, uma forma de indexação e de correspondência rápida devem ser realizadas sem alinhamento sequencial do modelo. Nestes cenários, a competição entre características de formato de modelos similares pode ocasionar um desafio para o reconhecimento. Para solucionar tal problema, propõe-se uma métrica que

quantitativamente mede a similaridade de modelos. Para cada modelo, utiliza-se uma medida de similaridade para definir o Model Centric Class (MCC), o qual contém um grupo de modelos e transformações similares entre o modelo e seus membros da classe. A informação de similaridade incorporada ao MCC é utilizada para impulsionar a geração de hipóteses de correspondência, de forma que o modelo correto ganhe mais oportunidades de ser hipotetizado e identificado. Para calcular a medida de similaridade entre dois pares de modelos utiliza-se a medida de distância entre eles. Dois conjuntos de dados são utilizados para a medição, a nuvem de pontos do modelo e as assinaturas de formato extraídas das posições amostradas. Antes da medição, os dois modelos são alinhados, Via Iterative Closest Point (ICP)[15], para se obter as suas poses relativas. O alinhamento de pose é iniciado através da inserção de um dos modelos no sistema de indexação, o que gera um alinhamento grosseiro entre este modelo e um modelo da base de dados. A similaridade é calculada pela soma ponderada das distâncias ponto-a-ponto e da diferença entre as assinaturas de formato. Para a geração de hipóteses correspondentes utiliza-se o spin images[16], aumentado com posições 3D, para representar os modelos e os objetos da cena. No processo de procura de características, Q melhores correspondências de características do modelo são obtidas para cada característica dos objetos da cena. Um método dirigido pelos dados[17] é utilizado para gerar hipóteses de correspondência entre cena e modelo. Existe um problema neste processo, o qual ocorre quando duas melhores correspondências das características do modelo em um par de características não pertencem ao mesmo modelo, fazendo com que nenhuma hipótese de correspondência seja gerada. Neste caso utiliza-se o MMC para impulsionar o processo geração de hipóteses. Para os experimentos com o método proposto foi gerada uma base de dados com modelos 3D de 366 veículos, os quais foram processados para gerar um nuvem de pontos, adicionado a mais dados de três conjuntos de dados de LADAR coletados e os resultados foram demonstrados em termos de curvas de precisão e lembrança (ROC).

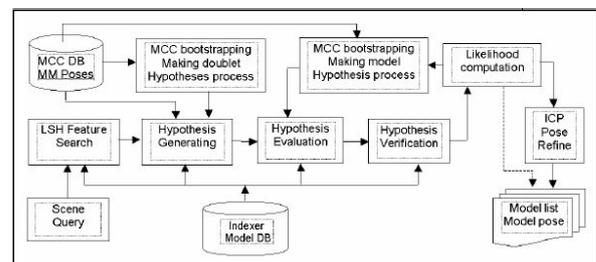


Fig. 4. Representação do método proposto por Yi Tan. Figura extraída de [14].

Um método para o reconhecimento e estimativa de pose de objetos 3D combinando características fotométricas e geométricas em uma sequência de cenas é proposto por Sukhan Lee [18]. Afim de utilizar cenas consecutivas, o método de filtro de partícula é utilizado e todas as partículas que representam uma possível pose do objeto são geradas por cada uma das características. Estas partículas são espalhadas

onde se considera a existência do objeto e a probabilidade de cada partícula é obtida através de um método de teste de correspondência com cada característica da cena. Então, o conjunto de partículas derivadas do SIFT(característica fotométrica) e do 3D lines(característica geométrica) são fundidas fornecendo a pose do objeto estimado. Esse processo de fusão é a combinação do conjunto de partículas do SIFT e do 3D line de forma a determinar a nova probabilidade das partículas em um determinado instante de tempo. Visando eficiência computacional, o sistema de reconhecimento proposto utiliza um processo hierárquico, o que significa dizer que as características são amostradas em fases ao em vez de utilizar todas as características. Quando a característica é amostrada pela primeira vez, elas são utilizadas randomicamente. Depois, utilizando a probabilidade da partícula calculada no nível 1, o espaço a ser amostrado e a contagem relativa podem ser especificados. Por fim, estima-se a posição da próxima melhor visão com base nos resultados do reconhecimento. Os experimentos com o método proposto focaram no reconhecimento de objetos dentro do refrigerador com uma câmera estereó, um robô e uma configuração eye-on-hand. Todo o passo a passo do método foi mostrado passo a passo nos experimentos.

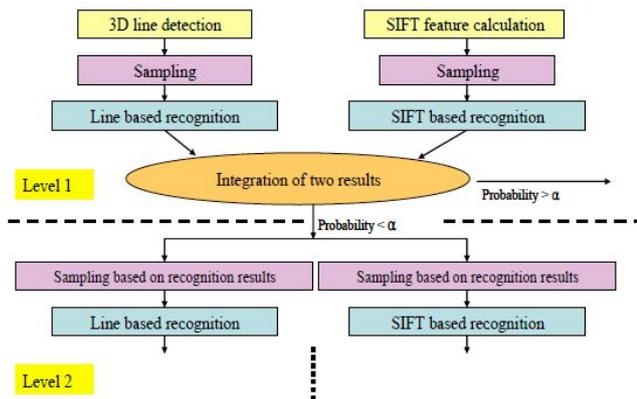


Fig. 5. Representação do método proposto por Sukhan Lee. Figura extraída de [18].

Stasse [19] apresenta um método de reconhecimento de objetos 3D com base em Spin-Images (SI) [16] para um robô humanoide com um sistema de visão estereoscópico. O método proposto inicia com cálculo do SI e seus passos relacionados, preenchimento do SI através de um método de interpolação bilinear e o cálculo da normal. Depois, computa-se multi-resolução do SI pelo cálculo do SI mais preciso através da verificação de todos os pontos. Para cada ponto do SI, quatro densidades correspondendo a cada quadrante são armazenados. Para imagens com baixa resolução, a densidade é computada utilizando a posição do ponto em relação ao quadrante considerado e a equação apresentada pelo autor. Outra característica necessária pela aplicação é a de identificar a resolução do objeto de interesse. Esta etapa depende das capacidades do sistema estereoscópico sendo utilizado, da distância entre o robô e o objeto e do possível esquema de subamostragem utilizado durante o processamento da imagem. Por fim, parte-se para o processo de procura com base em uma malha de pontos 3D. Inicialmente, a cena é dividida em

sub-blocos, onde o tamanho de cada sub-bloco é dado pela caixa delimitadora do objeto sendo procurado. Em cada sub-bloco os seguintes passos são aplicados: Seleção da melhor resolução de acordo com o tamanho médio da borda, extração da principal transformação rígida que projeta o modelo na cena e conferência se o modelo está na cena utilizando a transformação rígida computada. O ultimo passo, fornece um coeficiente de correlação principal e a posição mais a orientação na cena do objeto visto. O método descrito foi testado em diferentes situações para verificar sua eficiência. Na primeira situação, testou-se o SI do modelo de um coelho contra o SI de um dinossauro, na segunda, adicionando ruído e calculando o coeficiente de correlação e na terceira, simulando uma possível oclusão do modelo 3D. Por fim, foram realizados experimentos com um simulador, para identificar o modelo dentro de uma cena virtual, e com dados reais, verificando o coeficiente de relação obtido ao tentar encontra uma caixa de biscoitos.

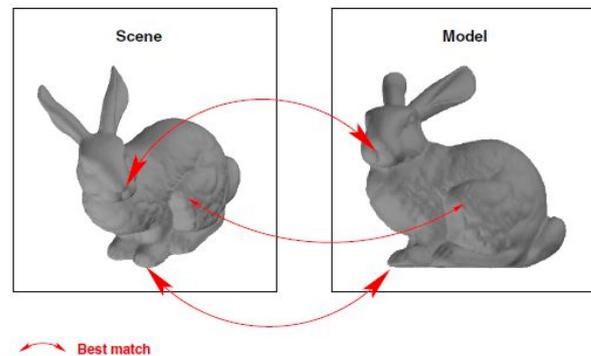


Fig. 6. Representação do método proposto por Stasse. Figura extraída de [19].

Chen Zhe [20] propõe uma extensão da *geometric hashing method* para o reconhecimento de objetos 3D sob transformação de perspectiva. Neste extensão, aspectos 3D do objeto e estruturas geométricas restritas são utilizadas para construir a tabela *hash*. Desta forma, as invariantes geométricas das estruturas restritas podem fornecer a função *hashing* e os aspectos 3D do objeto fornecem informações da pose do objeto. O procedimento para construção da tabela *hash* do modelo 3D do objeto é descrito pelos seguintes passos: geração do modelo 3D do objeto utilizando um software CAD; Configuração da estrutura de dados do modelo 3D na forma de uma lista de 3 níveis de pontos, linhas e planos para cada modelo, numerando os pontos, as linhas e planos e gravando suas inter-relações; Computação da representação de aspectos 3D do objeto para produzir a tabela de aspectos 3D; Para cada aspecto 3D na biblioteca de modelos, analisa-se a relação entre os pontos, linhas e planos do objeto e então extrai-se a estrutura geométrica restrita e computa-se a invariante geométrica da estrutura para formar um par (structure-id, invariant-id), e armazena-se na tabela *hash*. Para o reconhecimento os seguintes passos são realizados: extração das características de linha do objeto da imagem 2D; Procura de estruturas geométricas restritas das linhas obtidas e cálculo das invariantes de perspectiva da estrutura; Uso das invariantes

para indexar a tabela *hash*. Seleção dos modelos candidatos de acordo com o número de votos para cada modelo; Dependendo da relação de correspondência entre as estruturas geométricas 3D restritas e as características da imagem, computa-se a matriz de perspectiva de projeção; Utilizando a matriz calculada, transforma-se os aspectos 3D em espaços de imagem 2D. Então calcula-se o grau de correspondência de cada característica de linha no espaço de imagem 2D; De acordo com a consistência da correspondência, identifica-se o objeto. Se as características não correspondem, repete-se os dois últimos passos até que todos os modelos candidatos sejam verificados; Por fim, utiliza-se a matriz de projeção de perspectiva para calcular a pose do objeto. Para os experimentos foram utilizados quatro objetos 3D iguais aos objetos utilizados em [21], e os resultados demonstrados em termos da tabela com os modelos candidatos obtendo o maior número de votos.

Geometric hash table structure

Model	3D aspect	Perspective invariant
Model_1	Aspect_1	(Invariant_id , Structure_id)
	Aspect_2	...

	Aspect_n1	...
...
Model_N	Aspect_1	...
	Aspect_2	...

	Aspect_nN	...

Fig. 7. Representação do método proposto por Chen Zhe. Figura extraída de [20].

Raptis [22] propõe um sistema de reconhecimento de caricaturas de objetos/faces. Um caricatura é reconhecida através de um procedimento de fusão de probabilidades. A inovação introduzida é que a caricatura 2D do objeto/face que são obtidas em 3D, são fundidas em termos dos seus contornos. Adicionalmente, estas características são diretamente conectadas a todos os objetos e faces armazenadas na base de dados. Um rosto / objeto é assim considerado como a saída de uma detalhada análise Bayesiana probabilística dos contornos das visões. As características utilizadas são os pixels de borda do objeto/face que são extraídos das informações de baixo nível das bordas. As faces foram separadas do plano de fundo utilizando o C-means para dois *clusters* (plano de fundo e principal) para a primeira visão e então aplicou-se o classificador por Nearest Neighbour. Ambos os tipos de padrão são modelados como distribuições, dado que elas são vagas devido as condições imperfeitas de iluminação e diferentes posturas faciais. Para todas as características de contorno, hipóteses parciais, expressadas como condições probabilísticas Gaussianas, foram examinadas em tempo real, em termos de suas plausibilidades com relação a qual objeto elas estão provavelmente conectadas. Chega-se a uma distribuição final alocando um certo grau de confiança a um conjunto de objetos/faces disponíveis. Nos experimentos, as características dos objetos foram extraídos de imagens de objetos da base de dados University Object Image Library (C.O.I.L.- 120)[23] e para as faces utilizou-se a base de dados University of

Manchester face database (UMIST) [UMIST] e os resultados demonstrados com base na performance de classificação do método apresentado.

Um método para modelagem 3D de objetos e reconhecimento, o qual é robusto a variação de escala, iluminação e variações de ponto de visão é proposto por Delponte [24]. Este método possui duas partes, modelagem do objeto e reconhecimento. Na fase de modelagem, inicialmente, pontos chave locais são extraídos, por meio do descritor SIFT, de toda a sequência de imagens de treinamento de um objeto e explora-se a coerência temporal através do rastreamento desses pontos chave, obtendo assim uma lista de trajetórias, ou *trains* de pontos chave. Depois, esses *trains* de pontos chave são representados em uma forma compactada denominada característica virtual e essas representações utilizadas para construir um vocabulário. O vocabulário é o modelo do objeto, dado que é a informação do objeto que será utilizada para o reconhecimento. Na parte de reconhecimento, utiliza-se um *kernel* de intersecção de histogramas com base no vocabulário montado para treinar um Support Vector Machine (SVM) binário, visando o reconhecimento do objeto. Os experimentos com o método proposto foram realizados em um conjunto com 11 objetos diferentes e os resultados foram avaliados em termos de taxa de acerto em ambientes experimentais com variações de escala e oclusão.



Fig. 8. Representação do método proposto por Delponte. Figura extraída de [24].

Um método para reconhecimento de objetos 3D que explora a coerência espaço-temporal de uma sequência de imagens para capturar as características mais relevantes do objeto é apresentado por Arnaud [25]. Com a combinação de uma abordagem baseada em visões com o uso de características locais, obtém-se descritores que modelam como as características evoluem com o tempo. No método apresentado, cada objeto é representado por uma sequência de imagens adquiridas em um ambiente controlado. Pontos de interesse local (SIFT) são extraídos e rastreados na sequência de imagens com um algoritmo de filtragem (Kalman filter). Todas as trajetórias (*trains of features*) que são estáveis na sequência são utilizadas para descrever a aparência do objeto sob diferentes condições de visualização. Similarmente com a ideia de *bag of keypoints*, derivada do conceito de *bag of words*, cada *train of features* constitui uma palavra no vocabulário. Para cada objeto de interesse constrói-se um vocabulário e estes são utilizados para treinar um classificador, Support Vector machine (SVM), para

reconhecer o objeto. Para os experimentos realizados utilizou-se uma base de dados com 11 objetos diferentes representados por uma sequência de 200 quadros por objeto em diferentes pontos de vista. Os resultados foram avaliados com base na porcentagem acertos.



Fig. 9. Representação do método proposto por Arnaud. Figura extraída de [25].

O método apresentado por Kushal [26], propõe um método simples para construir automaticamente modelos de objetos 3D consistindo em um conjunto denso de de fragmentos de superfícies pequenas e de descritores dos padrões de textura de alguns pares estéreo. Para a descrição de características de baixo nível da imagem utilizou-se o detector *affine region* proposto por Mikolajczyk e Schmid [27]. O detector é inicializado com o detector de pontos de interesse Harris-Laplacian e com o operador de Difference of Gaussian (DoG). Os dois detectores acham tipos de pontos complementares. A saída do processo de detecção/retificação é um conjunto de fragmentos da imagem com formato de paralelogramo junto com os correspondentes mapeamentos de transformações de retificação, estes dentro de um quadrado 2×2 centrados na origem. Cada região detectada é representada pela matriz de transformação que mapeia o fragmento da textura retificada de volta a sua posição na imagem. Utilizam-se as câmeras estereo calibradas para determinar as estruturas 3D do objeto e construir o modelo. Possíveis correspondências primárias entre as regiões encontradas em cada par estereo são primeiro filtradas utilizando restrições de consistência geométrica e fotométrica, e então aumentadas com correspondências secundárias adicionais, para uma densa cobertura do objeto. A localização 3D e o formato dos fragmentos é determinada utilizando procedimentos estereo padrão para gerar modelos parciais, os quais são combinados posteriormente para formar um modelo completo do objeto. Um esquema similar é seguido durante a fase de reconhecimento. Primeiro, os fragmentos primário do modelo são correspondidos com as regiões encontradas na imagem de teste. Estes fragmentos primários são então utilizados como guias para o processo de correspondência de fragmentos secundários próximos. O objeto é reconhecido com base no número fragmentos correspondidos. O método proposto foi avaliado em uma base de dados consistindo de 9

objetos e 80 imagens de teste, comparando a taxa de detecção e a Equal Error Rate com outros dois métodos.

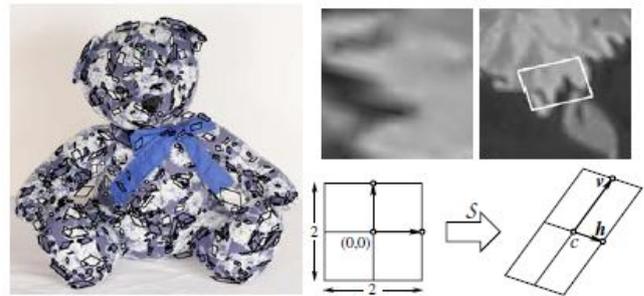


Fig. 10. Representação do método proposto por Kushal. Figura extraída de [26].

Arana-Daniel [28] apresenta uma teoria detalhada do Clifford Support Vector Machines (CSVM) e sua aplicação na classificação de objetos 3D, provenientes de nuvens de pontos, em múltiplas classes. A CSVM é introduzida como uma generalização da Support Vector Machine utilizando a álgebra geométrica de Clifford [29]. O autor explica todo o fundamento e teoria do CSVM e mostra um experimento para classificação de objetos em múltiplas classes com dois tipos de dados de treinamento, artificial e real (obtido por um sistema de visão estereo). Comparou-se também os resultados obtidos pelo CSVM contra o SVM e MLP. Nos experimentos procurou-se utilizar apenas um CSVM com um quaternião como entrada e um quaternião como saída, o que permite ter 16 classes. Basicamente empacotou-se em quaternião de característica um ponto 3D (na superfície do objeto) e a magnitude da distância entre este ponto e o ponto que está no eixo principal do objeto na mesma curva de nível. Para cada objeto foi treinado o CSVM utilizando um conjunto de várias características de quaternião obtidos de diferentes curvas de nível, o que significa que cada objeto é representado por várias características de quaternião. Devido a esta forma de treinamento, a ordem que cada características de quaternião é mostrada é importante. Começa-se amostrando dados da base até o topo do objeto e mostra-se os dados de treinamento e teste desta forma. Processa-se a saída usando um contador que grava qual classe ativa mais para cada conjunto de treinamento ou teste, afim de decidir a qual classe o objeto pertence. Os resultados iniciais foram computados primeiro para os dados sintéticos e para os dados sintéticos normalizados. Depois, utilizou-se os dados reais, comparando os resultados do CSVM com os resultados do MLP e do SVM em termos do número de amostras de teste classificadas corretamente.

Ramalingam [30] apresenta um paradigma fuzzy de classificação de superfícies, o qual é uma extensão das técnicas convencionais com base no sinal das curvaturas média e gaussiana. Curvaturas Fuzzy da superfície são modeladas por funções de adesão fuzzy. Estas superfícies permitem múltiplas descrições, modelando assim a incerteza em estimativas de curvas, e a classificação dos pontos da superfície é adiada até que pistas apropriadas sejam obtidas [31]. No entanto, a abordagem apresentada nos trabalhos referenciados é um tanto primitiva e as características extraídas das curvaturas média

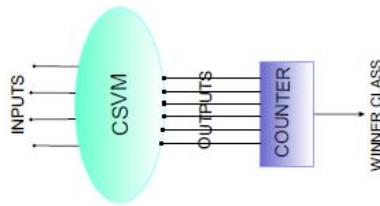


Fig. 11. Representação do método proposto por Arana-Daniel. Figura extraída de [28].

e gaussiana são diferentes da abordagem proposta. Tipicamente as características morfológicas das partes segmentadas, como descrito em [32], são utilizadas para a classificação da superfície como descrito em [31]. Na técnica apresentada, utilizam-se regras com base em fuzzy para classificar as superfícies, tornando possível tratar as duas abordagens como diferentes. Neste trabalho, a técnica proposta em [33] é empregada para a computação dos descritores de formato. Besl e Jain [34] propuseram uma técnica que utiliza o sinal das curvas média e gaussiana para classificar as superfícies em oito tipos básicos, os quais são posteriormente referenciados neste trabalho como superfícies *crisp*. No entanto, a abordagem produz resultados espúrios de segmentação enquanto classificando os *pixels* nas bordas entre tipos de superfícies. Esse problema é solucionado permitindo definições imprecisas para os sinais de curvatura média e gaussiana em termos da adesão nos conjuntos fuzzy. Permitindo a imprecisão na definição dos sinais resulta em classificações de superfícies adicionais, as quais são referenciadas como superfícies transitórias. Neste *framework* um *pixel* pode pertencer em até quatro tipos de superfícies com diferentes graus de confiança. O grau de confiança na classificação do *pixel* é dado pela função de adesão fuzzy em cada tipo de superfície e é derivada dos valores de adesão para os rótulos dos sinais da curvatura média e gaussiana. Deste modo, tem-se dois níveis de incerteza inter-relacionados: rótulos fuzzy dos sinais das curvaturas média e gaussiana e classificação fuzzy da superfície. Em ambos os casos os valores de adesão estão no intervalo de 0 a 1, onde 1 significa confiança total na classificação. Em 12 mostra-se a arquitetura geral para o método proposto de segmentação assim como para a abordagem tradicional. A entrada em ambos os sistemas é uma imagem de profundidade e a saída do sistema é um conjunto de imagens decompostas cada uma representando a distribuição dos *pixels* de um tipo de superfície. A computação da superfície é básica para os dois métodos, ambos requerem um processo de fusão para suavizar qualquer resultado espúrio e reduzir o efeito da *oversegmentation*. A operação de fusão é normalmente realizada com base em um determinado critério de suavização dos segmentos da superfície, resultando em um segmento de maior tamanho. A operação atribui a um *pixel* um tipo de superfície que é dominante em sua vizinhança. Esse processo é similar a convolução de uma máscara Gaussiana sob a imagem toda. A diferença principal entre o método tradicional e a abordagem proposta reside na forma como as curvaturas média e gaussiana são definidas. O método de classificação dos *pixels* proposto tem como base os sinais das curvaturas média e gaussiana em termos de funções de

adesão fuzzy. Deste modo, a arquitetura fuzzy consiste de dois módulos gerais denominados, parametrização fuzzy e defuzzificação. A parametrização fuzzy consiste em um fuzzificador e uma base de regras, as quais tem como base um sistema de mapeamento *multiple-input-multiple-output* (MIMO), enquanto a defuzzificação consiste em uma segmentação e fusão fuzzy com base em um processo de voto. Com base neste método de segmentação, apresenta-se uma arquitetura para o reconhecimento de objetos. Em sistemas convencionais, o reconhecimento com base em superfícies deriva características dos segmentos de superfícies satisfazendo algum critério de suavização. Dado que o conjunto de características rotuladas com base na segmentação fuzzy da superfície não necessariamente satisfaz tal condição, elas podem ser tratadas como distribuição da superfície para o propósito de extração de características globais. Neste trabalho, uma técnica de reconhecimento com base em *moments* Fuzzy, descrita e testada em [35], foi empregada. Vetores de características normalizados dos *moments* são gerados para cada instância de um objeto. A técnica é testada em duas bases de modelos de objetos, uma base de modelos de cinco objetos sintéticos de xadrez formado por cinco instâncias de cada e uma base de onze objetos de profundidade formada por cinco instâncias de cada objeto. Três experimentos foram conduzidos, avaliação da taxa de reconhecimento com instâncias vistas como objetos de teste; avaliação da taxa de reconhecimento com instâncias não vistas; e avaliação da taxa de rejeição, comparando os sistema fuzzy proposto com um sistema convencional.

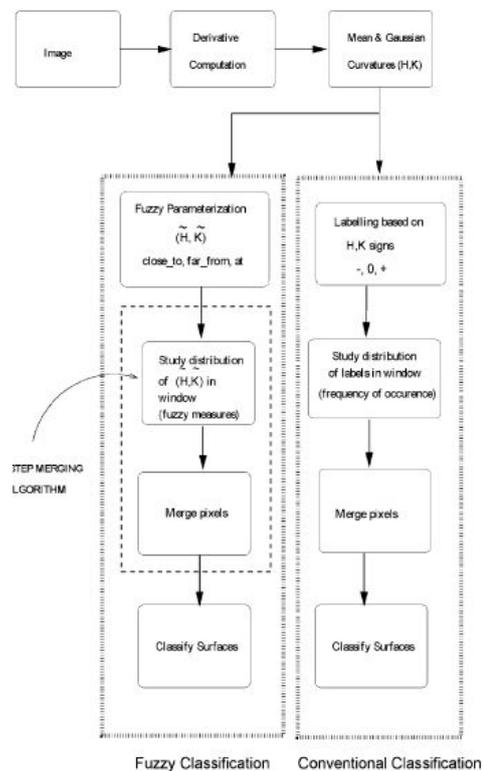


Fig. 12. Representação do método proposto por Ramalingam. Figura extraída de [30].

Uma abordagem para transformar objetos 3D em string de

características, as quais representam a distribuição dos voxels sob o *grid* de voxels, é apresentado por Assfalg [36]. Nesta abordagem, inicialmente tem-se o mapeamento dos objetos 3D em um *grid* de voxels. Depois, realiza-se o processo de computação de strings de características, o qual é obtido pela iteração através do *grid* de voxels, uma vez para cada dimensão, x,y e z, criando uma string de característica para cada dimensão. Essa *string* reflete a distribuição de voxels preenchidos ao longo dos três eixos do sistema de coordenadas. Por fim, após repetir esse procedimento para todas as dimensões, um objeto tridimensional é descrito por um conjunto de três strings. Escolheu-se utilizar uma medida básica de similaridade nas strings de características denominada Spectrum Kernel com a finalidade de determinar a similaridade entre duas strings de entrada. O Spectrum Kernel conta os pares de *substrings* idênticas em duas strings de entrada como sendo a medida de similaridade para as duas strings. Logo, dado dois objetos, computa-se spectrum kernel entre as strings representantes do objeto obtendo assim três valores de similaridade referentes aos três eixos dos dois objetos. Esses valores de similaridade podem ser unificados em um único valor representando o valor de similaridade entre os objetos. Foram propostas três formas de fundir estes valores, os quais foram comparados elegendo a melhor forma de fusão. Para os experimentos utilizando o método proposto foram utilizadas três bases de dados: Protein Data Bank (PDB) [37], a NTU 3D Model Benchmark Dataset [38] e a Princeton Shape Benchmark Dataset [39] e comparados os resultados obtidos com outros métodos de descrição de objetos (Shell model, Spherical Harmonics, Eigenvalue Model e Grid D2) em termos de acurácia obtida no processo de classificação, utilizando o Nearest Neighbor para realizar a classificação.

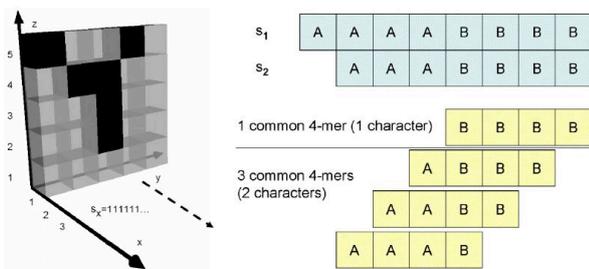


Fig. 13. Representação do método proposto por Assfalg. Figura extraída de [36].

Leitner [40] propõe uma técnica de aprendizagem não supervisionada para o reconhecimento de objetos partindo de um conjunto de imagens de treinamento não rotuladas e desordenadas. O sistema como um todo é dividido em aprendizagem e reconhecimento. Ambas as etapas iniciam com uma extração de características utilizando Maximally Stable Extremal Regions (MSER) [41], seguido por uma caracterização da aparência da região, por meio do SIFT [42], e do contorno da região, por meio do Modified Fourier Descriptor (MFD) [43]. Na etapa de treinamento, uma clusterização de imagens com base em árvore KD[44] identifica subconjuntos iniciais de imagem contendo o mesmo objeto, evitando uma correspondência exaustiva de todo o conjunto de treinamento.

Dentro destes subconjuntos, pares de imagens são correspondidos utilizando dois descritores de característica diferentes [45] afim de obter uma correspondência local. Um algoritmo guloso de otimização determina um ordenamento de imagens adequado para o rastreamento das características de treinamento para a geração do modelo de características de múltiplas visões. Um modelo probabilístico local da posição mútua das características completa o modelo do objeto. Assim como na aprendizagem, o reconhecimento inicia com a detecção com o MSER e a caracterização das regiões utilizando o SIFT e o MFD. A parte principal do reconhecimento é a correspondência dos descritores de características da imagem com os modelos de características de múltiplas visões, o qual é realizado de forma similar ao processo de correspondência de características executado nos subconjuntos de treinamento. Esse processo iterativamente otimiza o conjunto de correspondências utilizando o custo de correspondência dos descritores e o custo da posição mútua das características. Os testes do *framework* proposto foram realizados com base em um conjunto de dez objetos com um conjunto de treinamento com 360 imagens (36 para cada objeto). Os resultados obtidos foram comparados com uma abordagem SIFT supervisionada em termos de curva ROC.

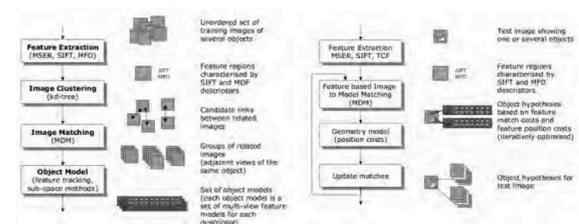


Fig. 14. Representação do método proposto por Leitner. Figura extraída de [40].

Kootstra [46] apresenta um sistema que adota a estratégia que os sistemas naturais exploram os objetos para o reconhecimento de objetos 3D utilizando robôs. A exploração dos objetos permite ao sistema aprender os objetos de diferentes pontos de vista e simplifica a segmentação do objeto do seu plano de fundo. Esta exploração é realizada via visão ativa, a qual, no trabalho do autor, consiste em utilizar robôs que exploram os objetos circulando ao redor deles, observando-os a cada 10 graus. Pela mudança ativa de visão, os robôs juntam informações que são utilizadas de dois modos: para detectar pontos chave estáveis e classificar eles, como objeto ou plano de fundo, e explorar os objetos afim de juntar mais evidências para resolver situações de ambiguidade. Para a detecção e descrição dos pontos chave utiliza-se o SIFT [47]. Um dos problemas do SIFT é que o tempo computacional aumenta com o número de pontos chave armazenados na base de dados. Para reduzir este efeito, utilizou-se uma rede growing-when-required (GWR) [48] para clusterizar de forma eficiente pontos chave similares. O método apresentado para a correspondência dos pontos chave observados na base de dados é um pouco diferente do apresentado nos trabalhos anteriores pelas seguintes razões: Primeiro, foca-se somente na correspondência de pontos chave individualmente, e portanto

não se utiliza a correspondência geométrica de conjuntos de pontos chave como utilizado em [47]. Segundo, na abordagem proposta, utiliza-se uma busca exaustiva através da base de dados para encontrar o vizinho mais próximo, dado que se esta interessado na melhor correspondência possível. Terceiro, usa-se um valor de limiar na distância do vizinho mais próximo, ao invés da taxa do vizinho mais próximo e do segundo vizinho mais próximo, para determinar a correspondência. Por fim, utiliza-se um modelo probabilístico levemente diferente, para o reconhecimento, que o usado em [44]. Para os experimentos, utilizaram-se 7 objetos colocados em ambientes desordenados para a base de imagens. No primeiro experimento, comparou-se a performance da abordagem ativa proposta com uma abordagem passiva em termos de taxa de reconhecimento em relação ao número de pontos de vista. No segundo experimento, comparou-se a performance da rede GWR com o SIFT padrão e com o SIFT utilizando apenas 36% dos pontos chave em termos de taxa de reconhecimento.

Okada [49] descreve o projeto e implementação de um sistema de reconhecimento de objetos 3D com base em conhecimento e integração de múltiplas pistas utilizando a técnica de filtro de partículas. Para reconhecer os objetos, define-se o conhecimento das seguintes pistas visuais: informação do formato 3D do objeto, histograma de cor da superfície de um objeto e bordas retas visíveis na superfície do objeto. A informação 3D do objeto é gerada por dois passos: Primeiro, pontos de características 2D são gerados utilizando o método de extração de características KLT [50], no qual os pontos de característica são localizados pelo cálculo do autovalor mínimo de cada matriz de gradiente 2×2 . Então, a correspondência estéreo com base em correlação é aplicada para calcular a disparidade dos pontos de características e obter a distância 3D dos pontos. Essas informações de pontos são utilizadas para medir a probabilidade do formato de entrada em relação ao de referência. O histograma de cor da superfície utiliza a distribuição com base no espaço HSV e tem sua probabilidade medida com base no coeficiente de Bhattacharyya [51] da distribuição de cores dos histogramas de referência e de entrada. As bordas retas visíveis do objeto são extraídas por meio da utilização do detector de bordas Canny, para extrair bordas confiáveis de uma imagem de entrada, e através do seguinte processo de extração de bordas retas: Geração de uma aresta entre pontos de início e fim; Cálculo da distância entre a borda e o ponto mais longe; e divisão da aresta neste ponto se a distância está abaixo de um valor de limiar. As bordas associadas com o modelo de referência são projetadas no plano da imagem 2D para medir a similaridade entre as bordas da imagem de entrada e as bordas do modelo de referência, e por consequência obter a probabilidade das bordas. As três pistas visuais possuem um parâmetro de peso, definido pelo usuário. Estas pistas visuais são integradas em um algoritmo de filtro de partículas[52], o qual é muito utilizado no rastreamento de objetos devido as suas características robustas. Para os experimentos com o sistema de reconhecimento proposto, utilizou-se imagens adquiridas por um sistema de câmeras estéreo equipado no robô HRP2JSK, mostrando os resultados por meio de imagens.

Delponte [53] explora o link entre reconhecimento de

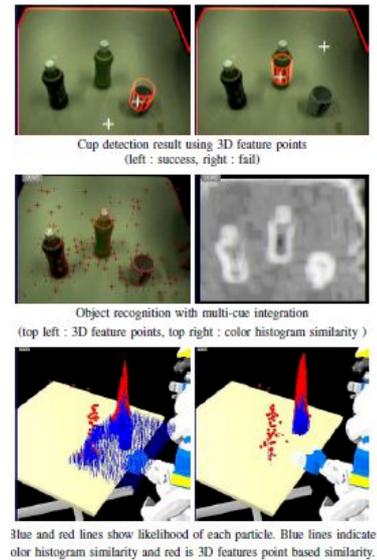


Fig. 15. Representação do método proposto por Okada. Figura extraída de [49].

objetos com base em visão temporalmente densa e uma representação esparsa da imagem com pontos chave locais. Inicialmente, olha-se para modelos dos objetos 3D começando com pistas visuais. Dado um objeto de interesse, na fase de treinamento, captura-se sua aparência 2D em pontos de vista temporalmente densos diferentes, utilizando o seguinte processo: primeiro detectam-se os cantos em um espaço-escala seguindo Lindeberg [54], também estima-se o canto principal de forma similar a [42]. Depois, rastreiam-se esse cantos através de uma sequencia de imagens com um filtro dinâmico não linear, unscented Kalman filter [55], que permite lidar com não linearidades no sistema. No final deste procedimento, pequenas trajetórias são descartadas. Cada ponto chave das características de trajetórias é representado pelo SIFT[42], aplicando-se um procedimento de limpeza para remover trajetórias ruidosas. Após o processo de limpeza, a característica invariante ao tempo é descrita por um descritor de aparência espacial, o qual é a média de todos os vetores SIFT de sua trajetória, e um descritor temporal, que contém informações sobre quando a característica apareceu primeiro na sequencia e quando foi observada pela ultima vez. Para aumentar a estabilidade do descritor de aparência, descarta-se os pontos chave que estão fora de uma hiper-bola centrada no centroide do descritor de aparência, a qual seu raio é proporcional ao desvio padrão dos pontos chave. No final da fase de treinamento obtém-se um modelo de um objeto como um conjunto de características invariantes ao tempo, ou seja, características obtidas das trajetórias dos pontos chave maiores que um número fixo de *frames*. Além disso, como não é feita nenhuma suposição nem do movimento da câmera, nem da distância entre a câmera e objetos observados, escolhe-se descrições invariantes a escala por meio do SIFT. Em tempo de execução, dado uma tomada de vídeo, extrai-se as características invariantes ao tempo e então corresponde-se as descrições da sequencia de teste contra os vários modelos

de treinamento por meio de dois estágios de comparação. O primeiro estágio compara a aparência, dado que o descritor de aparência é uma média do SIFT compara-se a similaridade entre as características por meio de uma intersecção de histogramas. No segundo estágio utiliza-se a descrição temporal de cada característica para procurar por grupos de características que se assemelham em um dado intervalo de pontos de vista nos modelos de treinamento e teste. Alguns experimentos foram realizados com a bordagem proposta em diferentes sequencias de vídeos e os resultados foram demonstrados em termos de porcentagem de acerto entre os vídeos de treinamento e teste.

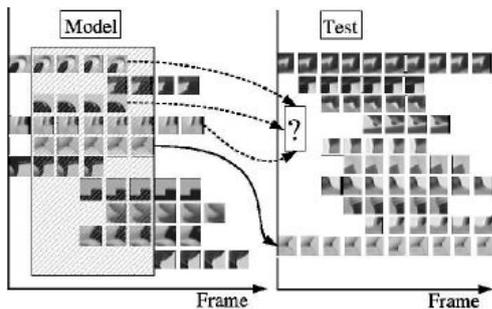


Fig. 16. Representação do método proposto por Delponte. Figura extraída de [53].

Jun-Hai Zhai[56] propõe um método de reconhecimento de objetos 3D com base em visão, o qual consiste de três passos. No primeiro, emprega-se a transformação wavelet [57] para decompor imagens de visões do objeto em sub-imagens de diferentes frequências. No segundo passo, para cada uma dessas sub-imagens, as características são extraídas utilizando uma abordagem de Singular Value Decomposition (SVD). Como uma imagem pode ser vista como uma matriz, o SVD pode ser utilizado para extrair características da imagem [58], [59], [60]. Estas características extraídas são combinadas para construir um vetor de características da imagem original. Por fim, no terceiro passo, o vetor construído é alimentado em um Support Vector Machine (SVM) para classificar os objetos. Os experimentos foram conduzidos na base de dados Columbia Object Image Library [23] e os resultados foram demonstrados em termos de taxa de classificação.

Salgian [61] analisa a performance de três descritores de regiões no contexto de reconhecimento de objetos 3D (SIFT [47], PCA-SIFT [62] e keyed context patches [63]). O conjunto de dados utilizado contém imagens de 6 objetos com planos de fundo limpos e com planos de fundo desordenados, tiradas ao redor de toda a esfera de visão, analisando as performances individuais e fundidas dos descritores. Para cada um desses descritores, tentou-se classificar cada imagem de teste através da procura da imagem na galeria de imagens que fornece a melhor correspondência. O autor fez uma breve descrição dos 3 descritores, explica como fez para realizar a combinação dos ranques dos métodos e por fim parte para os experimentos. Nestes, o autor avaliou a taxa de reconhecimento em imagens sem desordem e em imagens com desordem no plano de fundo para os três descritores e os resultados obtidos com três combinações dos descritores.

Xinju li [64] propõe um método para reconhecer imagens de profundidade utilizando aprendizagem supervisionada, a qual mede a similaridade entre imagens de profundidade utilizando o seus conjuntos de características. O método proposto funciona da seguinte forma: para cada imagem de profundidade, primeiro seleciona-se um conjunto de pontos, os quais são característicos e carregam informações salientes de geometria, como proposto em [65]. Cada ponto saliente é combinado com o descritor de superfície definido no remendo da superfície local perto do ponto. Após detectar os pontos salientes e computar suas assinaturas locais, através de duas definições de assinatura spin images e a assinatura proposta em [65], representam-se as imagens de profundidade como um conjunto de descritores de superfície não-ordenados. A função pyramid match kernel proposta por [66] é utilizada para medir a similaridade entre os conjuntos de características não-ordenadas. Por fim, dado um conjunto de n classes de imagens rotuladas, aprendem-se n classificadores utilizando a similaridade aos pares entre estas imagens, onde cada classificador separa uma classe de imagens das outras. O classificador utilizado foi o Support vector machine. Uma imagem de entrada pode ser submetida a esses classificadores e reconhecida com base na classe mais similar. Os experimentos com o método proposto foram realizados em duas bases, FRGC face database [10], [67] e na base de imagens de profundidade disponível em [68]. Os resultados obtidos foram avaliados em termos de taxa de reconhecimento, tanto para os experimentos de classificação de objetos quanto os experimentos para o reconhecimento de faces.

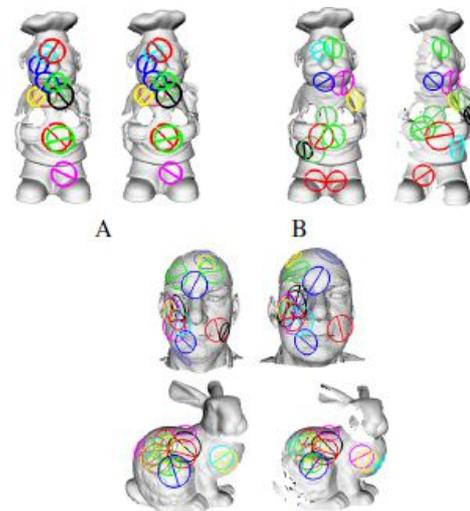


Fig. 17. Representação do método proposto por Xinju Li. Figura extraída de [64].

Assfalg [69] apresenta um método, referenciado como Spin Image Signatures (SIS), o qual se desenvolve sob a abordagem do Spin Images (SI) com adaptações para suportar recuperação eficaz por conteúdo. De acordo com o método, um conjunto de SIs é derivado para cada modelo, para obter uma descrição independente da visão de seu formato 3D e uma assinatura é avaliada para cada SI no conjunto. Como uma assinatura é computada para cada vértice no modelo

de malha, ou seja, o número de SISs é igual ao número de vértices na malha, gera-se um número intratável de descritores. A solução para esse problema é a utilização de uma técnica de clusterização dos SISs, clusterização por fuzzy [70], para obter uma representação mais compacta. A computação do número ideal de *cluster* é realizada com base em [71]. Para se obter a similaridade entre dois modelos 3D, o conjunto de centros de *cluster* para os dois modelos são comparados. A melhor função de correspondência entre centros de *cluster* é definida pela permutação que minimiza o soma das distâncias entre os centros de *cluster* correspondentes. A solução para a função definida é obtida de acordo com uma busca gulosa, através do escaneamento de todos os centros de *cluster* no conjunto de centros de *cluster* associando a cada centro de *cluster* o centro de *cluster* mais similar ainda não associado. Os experimentos foram realizados em duas bases de dados, Princeton Shape Benchmark [39] e Core-art (formada por uma miscelânea de modelos disponíveis na web, alguns são escaneamentos em alta qualidade de modelos 3D do De Espona 3-D Models Encyclopedia). Comparando os resultados do método com outros métodos (Light Field [38], the Curvature Histograms [72], the Shape Function [73], the Geometric Moments [74], and Spin Images na sua formulação original [16]) em termos de curva de precisão e lembrança.

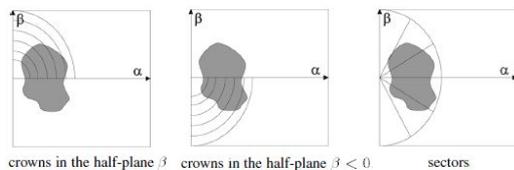


Fig. 18. Representação do método proposto por Assfalg. Figura extraída de [69].

Uma técnica para classificação de objetos 3D utilizando uma Global Geodesic Function (GGF), para intrinsecamente descrever a superfície de um objeto, é proposta por Aouada [75]. A variabilidade estatística desta função de formato é analisada utilizando uma medida teórica de informação conhecida como Jensen-Shannon Divergence (JSD)[76]. Para comparar objetos de forma eficiente, caracteriza-se cada classe de objeto por dois parâmetros na etapa de aprendizagem. O algoritmo de classificação pode ser descrito pelos seguintes passos: inicialmente define-se um número de classes de objetos e, na fase de aprendizagem, associa-se para cada classe uma característica de resolução da classe e um valor de limiar. Então, ordena-se todas as classes de forma crescente com base no valor de resolução. Constrói-se uma super-classe através da mescla de classes que compartilham o mesmo parâmetro de característica de resolução de classe. Inicia-se da menor resolução e inicializa-se uma variável de controle L com o valor 1. Computa-se o GGF do objeto na resolução de característica de classe na posição L e obtém-se os seus parâmetros de resolução. Compara-se os parâmetros de resolução do objeto com aqueles da classe na posição L e se a similaridade é estabelecida, termina-se a procura se a classe na posição L é uma classe ou aplica-se um valor de limiar se a classe na posição L é uma super-classe. Caso contrário, se a similaridade

não é estabelecida, passa-se para a próxima posição, i.e., $L = L + 1$ e repete-se a operação da comparação da similaridade com a próxima classe até obter uma similaridade estabelecida. Os testes com a técnica de classificação proposta foram realizados nas bases de dados de Princeton [39] e MCGill [77] e os resultados demonstrados em termos do número de decisões corretas sob o número de testes executados.

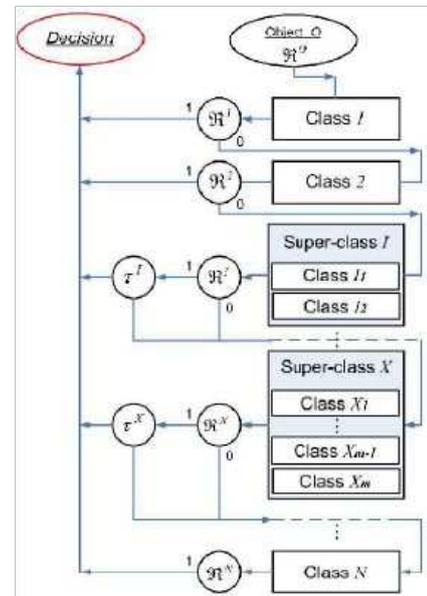


Fig. 19. Representação do método proposto por Aouada. Figura extraída de [75].

Lee-Yeng Ong [78] apresenta um *framework* teórico para derivar invariantes de escala a translação para o 3-D Legendre moments, através da utilização de métodos direto e indireto. O método indireto gera as 3-D Legendre invariantes a partir das invariantes de momentos geométricos 3-D existentes. O método direto, por outro lado, elimina os fatores de deslocamento e escala dos polinômios de Legendre para gerar as invariantes de escala e translação. Para a computação do método indireto os seguintes passos são executados: definição das invariantes de translação através da derivação dos momentos geométricos 3D com respeito ao centroide. Então, os centroides são fatorados para expressar central moments em termos dos moments originais [79]. Depois, para obter as invariantes translation scale (TS), as invariantes de translação são aplicadas em invariantes de escala. Por fim, as invariantes TS do Legendre moments são obtidas através da extensão da equação de relação do 2-D Legendre-geometric para 3D. Para o método direto os mesmos procedimentos são aplicados, entretanto, as invariantes TS são derivadas diretamente dos Legendre polinômios. Para verificar as invariantes TS para os métodos direto e indireto, um experimento com 3 imagens binárias com resolução 60x60x60, deslocadas e escalonadas, foi realizado, utilizando a porcentagem de *moment spread* do centroide para medir a performance dos descritores invariantes.

Em [80], Kordelas apresenta um método para reconhecimento de objetos 3D em uma cena 2.5D com oclusão, desordem e ruído. O método proposto tem como base a extração de

mapas de distância que caracterizam a topologia local de uma superfície. Os descritores do modelo são extraídos a partir de objetos 3D triangulados, enquanto os descritores da cena são extraídos de uma superfície 2.5D triangulada. Os mapas de distância são criados utilizando um algoritmo de intersecção raio-triângulo, onde as distâncias entre um setor circular de pontos, que expressa a origem dos raios orientados longe da superfície, e a superfície triangulada são computadas. Estas distâncias por setor, são organizadas em ordem ascendente, e um histograma 2D é criado, o qual expressa o número de setor de pontos dentro de cada distância de bin. A partir do histograma 2D, um vetor descritor é formado como a porcentagem normalizada do número de cada ponto no bin para o número total de pontos no setor. Ao fazê-lo, o número de descritores extraídos é independente ao número de vértices que um objeto ou cena contém. Deste modo, o número de descritores para um objeto 3D, em diferentes níveis de detalhamento, permanece o mesmo, em contraste a outros métodos apresentado [16], [81]. Desta forma, uma comparação dos descritores mais rápida e simples é alcançada. Antes do procedimento de reconhecimento, os descritores são criados e armazenados em uma biblioteca de modelos. Depois, a dissimilaridade dos descritores do modelo e cena são computados. Os descritores os quais a dissimilaridade computada é abaixo de um limiar, determinam o objeto ao qual o descritor corresponde. Deste modo uma correspondência entre pontos da cena e do objeto modelo é estabelecida. Os pares de todos os pontos correspondidos resulta em um conjunto considerável de correspondências de pontos entre a superfície do objeto e uma superfície de cena. Este conjunto é utilizado para alinhar a superfície do objeto da biblioteca com a superfície da cena. Para avaliar a performance do sistema de reconhecimento proposto, uma biblioteca de 8 modelos sintéticos foi criada. Durante os experimentos, a performance da abordagem proposta foi testada para cenas com ruído, oclusão e desordem e foi comparado com o método spin image [16]. O sucesso de reconhecimento foi verificado pela computação das taxas de verdadeiros positivos, falsos positivos e falsos negativos.

O trabalho proposto por Hui Chen [82] apresenta um descritor local para representação de superfícies e reconhecimento de objetos 3D. Esse descritor local apresentado, é caracterizado por seu centróide, o tipo de superfície local e um histograma 2D. O método apresentado possui dois estágios. O primeiro estágio é a construção do modelo e o segundo é parte de reconhecimento. Inicialmente, tem-se a definição de pontos de características definidos em áreas de grande variação de formato e medidos através do índice de formato, os quais são calculados através de curvaturas principais. Depois, para cada ponto de característica, tem-se o cálculo do Local Surface Patches (LSP) definido por seu ponto de característica, o tipo de superfície e um histograma dos valores dos índices de formato vezes o produto escalar da normal da superfície no ponto de característica e seus vizinhos. Devido a incerteza da localização do ponto de característica, também calcula-se o LSP dos vizinhos do ponto de referência, utilizando a média e o desvio padrão dos valores do índice de formato da região para indexar a tabela hash de informações sobre o

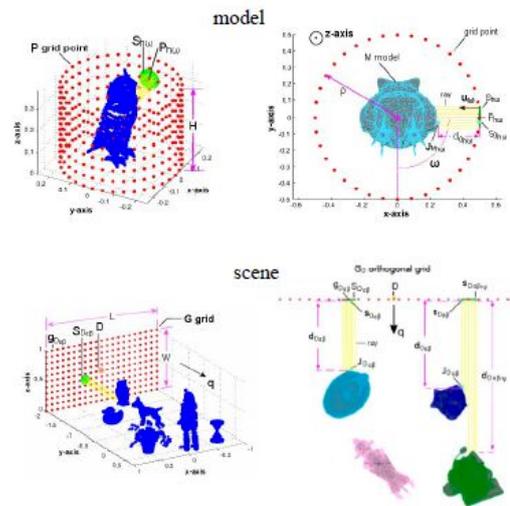


Fig. 20. Representação do método proposto por Kordelas. Figura extraída de [80].

LSPs do modelo. Para cada objeto de modelo, repete-se esse processo para construir a base de dados dos modelos. Para a parte de reconhecimento, repete-se o processo de extrair os pontos de característica, calcular os LSP, a média e o desvio padrão da região para cada LSP e coloca-se um voto para a tabela *hash* se a dissimilaridade do histograma de um LSP modelo e um LSP de teste cai dentro de um limiar e se o tipo de superfície é o mesmo. O modelo com maior número de votos é indicado como o tipo de objeto de entrada. Para a parte de teste e validação, 20 objetos foram modelados com esta técnica e foram colocados contra 20 objetos de teste em duas poses. Outros experimentos foram realizados em bases de dados maiores para avaliar a robustez e a invariância a rotação da representação. Testes comparativos contra outras duas formas de representação também foram realizados (Spin Image e o Spherical Spin Image). A verificação dos resultados foi realizada utilizando o algoritmo Iterative Closest Point (ICP) para alinhar os modelos da cena com os modelos mais parecidos.

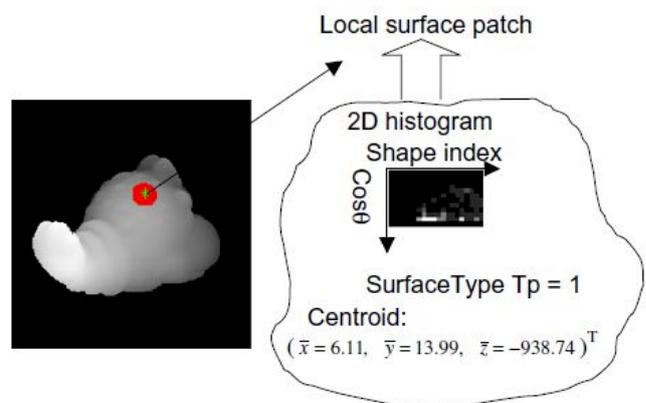


Fig. 21. Representação do método proposto por Hui Chen. Figura extraída de [82].

Weiwei Xing [83], [84] propõe um sistema para reconstrução e reconhecimento de objetos 3D. Esse sistema é composto de 3 subsistemas: extração de descrição estrutural com base em *Superquadric*, reconstrução de objetos com múltiplas partes e reconhecimento de objetos 3D. O *Superquadric*, utilizado pelo autor anteriormente na classificação de objetos 3D por fusões de partes [85], [86], é uma família de formatos paramétricos que pode descrever uma vasta variedade de formatos primitivos 3D com parâmetros compactos. O poder de modelagem do *Superquadric* é aumentado através da aplicação de operações de deformação aos *Superquadric* básicos. A descrição estrutural de objetos 3D com base em *Superquadric* é implementada em dois níveis: nível de características geométricas, descrevendo os atributos de formato de cada parte do objeto, e nível de características topológicas, refletindo a conexão global das estruturas entre as partes do objeto. O sistema de reconstrução com múltiplas partes reconstrói modelos 3D com diferentes números de partes e diferentes formatos, mostrando visualmente, em tempo real e de forma iterativa, o modelo 3D. Para implementar o reconhecimento de objetos 3D, constrói-se uma biblioteca de modelos 3D com um número de modelos para cada classe de acordo com a atual diversidade da classe. Os modelos dentro de uma classe de objetos são similares, mas exibem algumas diferenças entre eles, logo uma classe de objetos é um conjunto no qual cada elemento é um modelo 3D denotado por um vetor de característica. Para o reconhecimento entre os modelos 3D da biblioteca e um conjunto de objetos desconhecidos representados por descrições de características estruturais apresenta-se um método formado por dois estágios. O primeiro estágio é a procura em árvore, na qual as restrições forçadas pelas características extraídas e as regras de procura em árvore são definidas para encontrar rapidamente possíveis interpretações entre o modelo 3D e o objeto desconhecido. Com base nas interpretações e nas características estruturais de formato presentes, uma medida de similaridade é utilizada para obter resultados de correspondência com classificações de similaridade, atribuindo o objeto a classe de objetos com maior correspondência. O teste realizado com o sistema descrito utilizou uma biblioteca de modelos 3D privada (não descrita no artigo) e os resultados obtidos foram demonstrados com base em uma execução do sistema classificando um objeto desconhecido.

O trabalho apresentado por Sungho Kim [87], apresenta um *framework* com base em grafos para o reconhecimento de objetos, lugares e partes em ambientes fechados. Para tanto, a cena é interpretada como um modelo de grafo não direcionado composto de nodos de lugar, nodos de objetos e nodos de partes com ligações não direcionadas. Com a introdução da ideia de reconhecimento colaborativo de lugares e objetos, denominado contexto hierárquico, unificado com o contexto espacial, monta-se um Hierarchical Graphical Model (HGM). Neste modelo, os nodos de objeto e partes contém rótulos e informações de pose importantes para a inferência robusta de objetos em uma estrutura de grafo hierárquica que relaciona o lugar, objetos e partes. No caso, a camada de lugar recebe evidências das características da imagem e dos objetos. A camada de objeto recebe informações de partes,

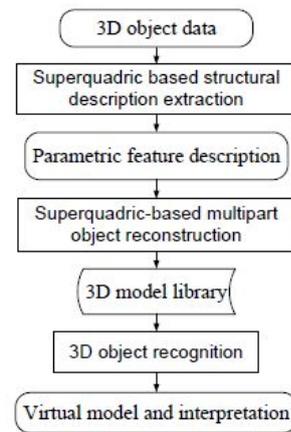


Fig. 22. Representação do método proposto por Weiwei Xing. Figura extraída de [83].

da cena e de objetos vizinhos. A camada de partes recebe evidências das características, dos objetos e de partes vizinhas. Essas evidências são utilizadas para o cálculo de probabilidade condicional e posterior inferência para o reconhecimento. Para a representação das características visuais de forma a serem adequadas para o reconhecimento com base em contexto utiliza-se uma generalização do SIFT estendendo o detector de características com o detector de cantos de Harris. Para aprender os parâmetros do modelo, utiliza-se uma aprendizagem por partes [88], a qual envolve dividir o modelo de grafo não direcionado em partes, treinando-as independentemente e combinando os pesos aprendidos. Após a aprendizagem do modelo tem-se a parte da inferência. Como a inferência inclui estimativas variáveis de estrutura com a distribuição marginal de cada nodo, aproxima-se a pseudo-probabilidade da distribuição marginal utilizando multimodal sequential Monte Carlo com atualização dos pesos através da propagação de crença. Hipótese multimodal baseada em dados e poda com base em contexto fornecem a inferência correta. Os resultados obtidos com a aplicação do *framework* proposto, utilizando 112 objetos em 12 lugares, foram avaliados em termos de taxa de detecção e taxa de falsos alarmes.

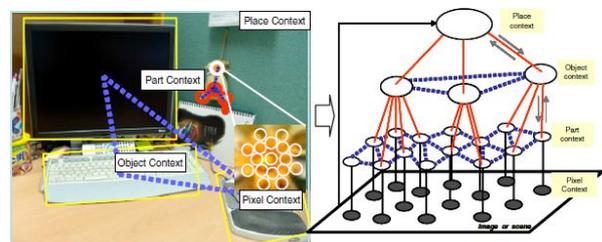


Fig. 23. Representação do método proposto por Sungho Kim. Figura extraída de [87].

Tim K. Lee [89] apresenta um método em dois estágios para o reconhecimento de objetos 3D com base em imagens 2D. Inicialmente, a imagem é descrita em termos de Curvature Scale Space (CSS). A CSS tem como base uma imagem binária 2D para representar o formato de uma curva fechada

parametrizada em múltiplas escalas. A imagem binária é construída pela convolução de uma curva fechada por uma série de gaussianas. Utilizando essa representação aplica-se um método de correlação de fase para formar um novo tipo de imagem CSS. Com base nesta nova representação monta-se um vetor de características através da soma separada de cada tamanho de arco e escala, concatenando em um novo vetor denominado *marginal-sum vector*. Em adição a essa representação, propõe-se realizar a parte de correspondência em um *eigenspace* de vetores de imagem reduzidos, como por exemplo, formados pelo método de Singular Value Decomposition (SVD). A parte de correspondência é projetada para uma base de dados de formatos com informações de categorização, como por exemplo, a base de dados 3D do laboratório de Michael Tarr's [90], onde os contornos foram extraídos com uma técnica simples de remoção de bordas. A parte de reconhecimento em dois estágios inicia com a tentativa de identificação da categoria do objeto de teste, o que pode ser traduzido para um o problema de encontrar o *eigenspace* que melhor descreve o objeto de teste. Para tanto, projeta-se o vetor de características do objeto de teste em cada *eigenspace* das categorias em turnos. O *eigenspace* da categoria que fornecer a reconstrução mais próxima do vetor de características do objeto de teste é definida como a melhor categoria. Uma vez que a melhor categoria foi identificada, pode-se determinar a melhor correspondência do objeto teste com objetos na categoria através de um cálculo de distância, distância euclidiana, entre os pontos dos objetos no *eigenspace*. Os testes foram realizados na base de dados 3D do laboratório de Michael Tarr's, a qual consiste de 209 objetos capturados de 14 pontos de vista diferentes. Os resultados foram avaliados em termos de taxa de reconhecimento.

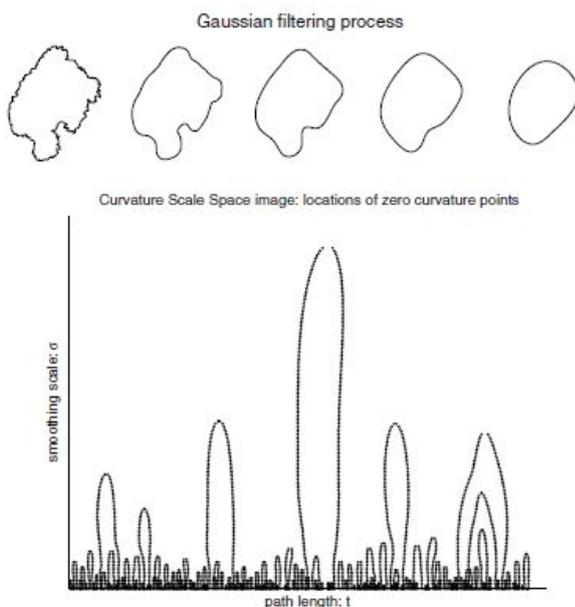


Fig. 24. Representação do método proposto por Tim K. Lee. Figura extraída de [89].

Kietzmann [91] apresenta uma nova variação do generalized relevance learning vector quantization (GRLVQ) utilizado

para reconhecimento de objetos. A variação proposta é uma versão incremental do GRLVQ, o iGRLVQ, a qual permite a automática seleção do número de protótipos (*codebook size*) para cada classe. Esta seleção automática é realizada através do início com apenas um protótipo para cada objeto e a subsequente adição de novas representações para as instâncias. Devido a esse recrutamento sob demanda, modelos comparativamente esparsos são criados. Outro efeito da variação proposta é a quantidade de conhecimento sobre o modelo necessária, a qual tem que ser fornecida manualmente através do ajuste dos parâmetros, é reduzida dado que o tamanho do *codebook* é selecionado automaticamente durante o processo. Integrando o mecanismo de aprendizagem incremental com base em protótipos com um conjunto genérico de características visuais, acoplado com a habilidade de aprendizagem de características em tempo real, monta-se uma arquitetura efetiva para o reconhecimento de objetos. Com base nesta arquitetura, o sistema de reconhecimento foi denominado Feature and Incremental Learning of Objects Utility (FILOU), o qual, com base no treinamento realizado classifica os objetos alvo. Para testar a performance de reconhecimento do FILOU, vários experimentos foram realizados utilizando a base de dados COIL100 [23], avaliando os resultados em termos de acurácia, efeitos da redução da dimensão e relevância da aprendizagem.

Xu Sheng [92] propõe um método de reconhecimento de objetos 3D, o qual usa apenas algumas características extraídas de cada imagem 2D de objetos 3D. Color moments [93] e características de textura (Gray Level Co-occurrence Matrix) [94] são utilizadas para distinguir objetos 3D com formatos similares e diferentes cores e texturas. Quando os objetos 3D são projetados em imagens 2D, a distorção dos objetos é um problema inevitável. Para contornar este problema, utiliza-se o Hu's moment invariants [95], o qual demonstrou ser invariável sob rotação, translação e escalonamento dos objetos em imagens 2D, e o affine moment invariants [96], o qual possui propriedades de invariância sob deformações afim, quando as visões dos objetos variam. Então compõe-se estas características em um vetor de características unidimensional de 23 componentes para cada imagem dos objetos 3D, e apresenta-se esses vetores para uma rede neural por Back Propagation para aprendizagem. O método proposto foi testado com 40 objetos 3D selecionados da base de dados Columbia Object Image Library (COIL-100) [23], com e sem a presença de ruído. Os resultados dos testes foram demonstrados em termos de taxa de reconhecimento.

Uma abordagem para reconhecimento de objetos 3D representados por nuvem de pontos é apresentado por Yu Zhong [97]. Esta abordagem utiliza o método de representação de formato denominado Intrinsic Shape Signatures (ISS), o qual consiste em um *frame* de referência intrínseco, permitindo a extração de características invariantes a visão e um rápido registro da pose, e um vetor de característica discriminativo codificando as características do formato 3D. Dois ISS são correspondidos através da comparação dos vetores de característica do formato. Dado que os vetores de característica consistem de uma contagem ponderada dos pontos nos *bins* de particionamento do volume, utiliza-se um cálculo estatístico para

computar a distância entre os dois vetores de característica. A estimativa da pose é realizada diretamente utilizando o *frame* de referência intrínseco. Para corresponder duas nuvens de pontos, primeiro utiliza-se o ISS para representar a nuvem de pontos. Depois, se estabelece a correspondência entre as ISSs das nuvens de pontos. Para um ISS de uma nuvem de pontos de consulta, computa-se a sua distância para cada ISS para cada modelo de nuvem de pontos. Se a distância for menor que um valor de limiar, adiciona-se a pontuação de correspondência e o par de pontos na lista de resultados. Ordena-se os resultados em uma ordem crescente de valores e, começando do início da lista, salva-se o resultado se os dois pontos base para o resultado não foram utilizados ainda, e marca-se os dois pontos base como utilizados. Caso contrário, remove-se o resultado da lista de forma que cada ponto base conte no máximo uma vez para as correspondências válidas. Realiza-se esse processo para todos os pontos na lista e o resultado é um conjunto de correspondências entre ISS com propriedades de formato similares. Este conjunto é então utilizado para medir a similaridade do formato e a transformação da pose entre duas nuvens de pontos. A medida de similaridade entre duas nuvens de pontos é a porcentagem de ISSs, com pares de descritores correspondentes, que concordam com a transformação de pose global entre as duas nuvens de pontos. Os experimentos para avaliar a performance do algoritmo proposto foram realizados utilizando a Princeton Shape Benchmark (PSB) [39] para o reconhecimento de diferentes tipos de veículos. A performance do ISS foi comparada com outros métodos (Spin image [16] e o 3D Shape Context (3DSC) [98]) em termos de taxa de reconhecimento \times ranque dos modelos recuperados na presença de ruído e desordem na cena.

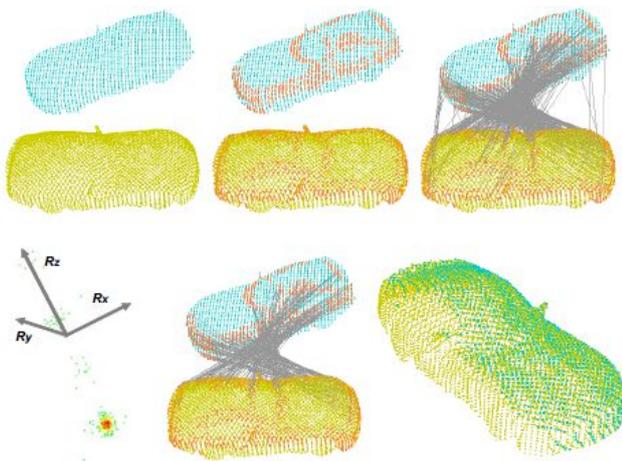


Fig. 25. Representação do método proposto por Yu Zhong. Figura extraída de [97].

O trabalho apresentado por Sungho Kim [99] [100], propõe o um método de representação de objetos 3D escalável e um método de aprendizagem para o reconhecimento de objetos do cotidiano. O método proposto tem como base o Common-Frame Constellation Model (CFCM). Este método fornece vantagens em termos de eficiência computacional e remoção de redundância através do compartilhamento dos parâmetros de visão do objeto. Fixando o ID do objeto e o ponto de vista,

cada parte pode compartilhar parâmetros de visão ou o *frame* do objeto. O termo Common Frame significa as coordenadas de referência do objeto. Este esquema de representação com base em partes do objeto é referenciado como CFCM porque cada parte compartilha parâmetros do objeto (ID do objeto, pose). O CFCM é a forma de compartilhamento de parâmetros de visão buscando a escalabilidade à forma de representação de objeto proposto. Outras duas forma de compartilhamento são: compartilhamento de aparência local e compartilhamento de múltiplas visões, que condizem, respectivamente, ao compartilhamento de partes similares de um mesmo objeto ou diferentes objetos e o compartilhamento de uma visão que junta múltiplas visões. Motivado pelas três formas de compartilhamento, a forma escalável de representação de objetos 3D proposta utiliza uma biblioteca de aparências locais. A característica local utilizada é composta por um vetor de aparência e informações locais de pose. A aparência de uma parte individual pode ser qualquer coisa (SIFT, momentos, ou PCA). O vetor de aparências é gerado através da clusterização de um conjunto de características locais extraídas das imagens de treinamento. Cada imagem de treinamento é representada por um CFCM, onde uma parte tem um índice para a biblioteca de aparências e a outra parte contém informações relativas ao *frame* de referência. CFCMs pertencentes ao espaço de transformação de similaridade são fundidos em um único CFCM. Cada CFCM contém informações da pose e índices para a biblioteca de aparências assim como cada elemento na biblioteca de aparências, compartilhada por todos os objetos, contém um link para as partes nos CFCMs. Esse fato pode ser utilizado para gerar hipóteses durante a etapa de reconhecimento. A parte de detecção de características locais é realizada pelo Generalized-robust invariant feature (G-RIF) e a clusterização das aparências é realizada através da combinação de métodos *bottom-up* e *top-down*. Com a biblioteca de aparências e os CFCMs construídos, o próximo passo é o reconhecimento dos objetos. Nesta etapa a cena é representada também por um conjunto de CFCMs e as hipóteses entre pares correspondentes são construídas com base em Hough transform. A hipótese gerada é aceita ou rejeitadas com base no tamanho do *bin* com um valor de limiar ótimo. O método proposto foi avaliado em na base de dados COIL-100 [23] e em um conjunto de dados capturados em um ambiente fechado desordenado do prédio da engenharia elétrica. Os resultados obtidos foram avaliados em termos de taxa de reconhecimento e comparados com o método base do estudo, sem a parte de compartilhamentos.

Huy Tho Ho [101] apresenta um método para extração de características locais salientes de modelos 3D utilizando a curvatura da superfície destes modelos. A curvatura da superfície é calculada utilizando o algoritmo multi-escalar proposto, o qual tem como base uma medida de curvatura local, invariante a rotação e translação, conhecida como *curvedness*. Esta medida, é um número positivo que captura a quantidade de curvatura em uma região local ao redor de um ponto. Diferentes valores de *curvedness* de um ponto são calculadas em múltiplas escalas através do ajuste de uma superfície de diferentes tamanhos a sua vizinhança. Um conjunto de pontos de características salientes confiáveis é formado pela procura

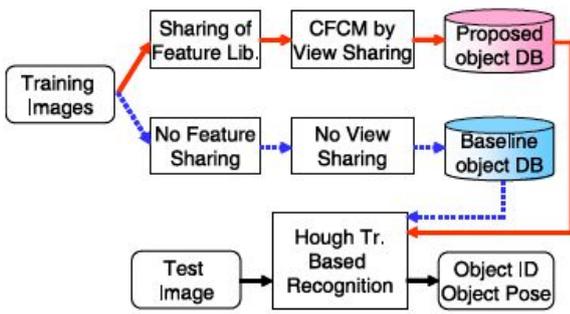


Fig. 26. Representação do método proposto por Sungho Kim. Figura extraída de [100].

do conjunto de extremos da representação espaço-escala da superfície de entrada. Também propõe-se um método para avaliação de confiança em cada ponto chave com base no desvio dos valores de *curvedness* da vizinhança. Este método de avaliação de confiança tem como propósito reduzir os pontos que não possuem um desvio maior no valor de *curvedness*, em relação a outros valores na vizinhança imediata. Os testes com o método proposto foram realizados em uma variedade de modelos 3D com diferentes níveis de ruído, com a finalidade de demonstrar a efetividade e robustez do método em termos de número de características com diferentes valores de limiar para um determinado nível de confiança.

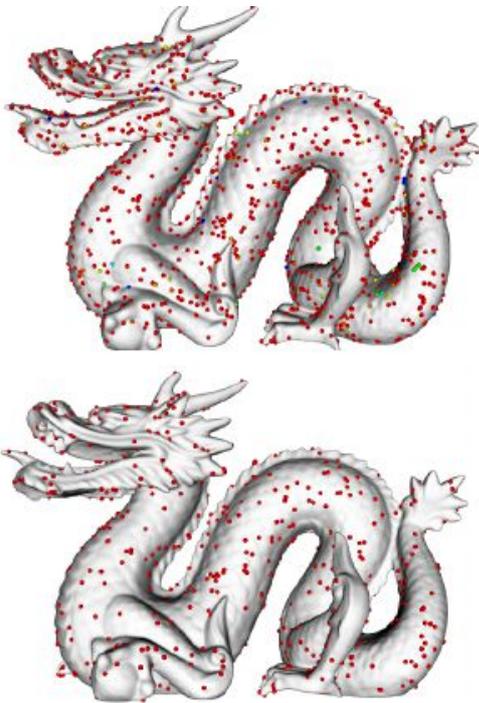


Fig. 27. Representação do método proposto por Huy Tho Ho. Figura extraída de [101].

Xiaolan Li [102] apresenta um *framework* para identificação de formato 3D parcial em partes de CAD 3D. O método proposto é dividido em dois estágios. O primeiro estágio, realizado offline, tem como propósito construir o dicionário

de palavras visuais, presente no *framework Bag of Words*, com base em uma base de dados 3D. Primeiro, as características locais são extraídas de cada modelo na base de dados, utilizando o *Spin Images*. Depois, um método de clusterização/classificação (K-means) é aplicado na coleção de características para construir dicionário de palavras visuais. O segundo estágio é a comparação em tempo real. Para os dados consultados, são extraídas as características locais e procura-se no dicionário pela palavra visual mais próxima. Depois, representa-se os dados consultados com um vetor de características, no qual cada elemento corresponde a uma palavra visual no dicionário e o valor corresponde a frequência de aparecimento da palavra nos dados de consulta. Por fim, uma certa métrica de similaridade, *KL divergence*, é escolhida para comparar a diferença entre os dados consultados e os modelos na base de dados. Uma lista de classificação de recuperação é a saída do *framework*. Para os testes com o *framework* proposto foi utilizado o Purdue Engineering Benchmark [103], o qual contém 801 modelos CAD 3D classificados em 42 classes, e os resultados foram avaliados em termos das curvas de precisão e lembrança com base nos valores medidos pela métrica utilizada.

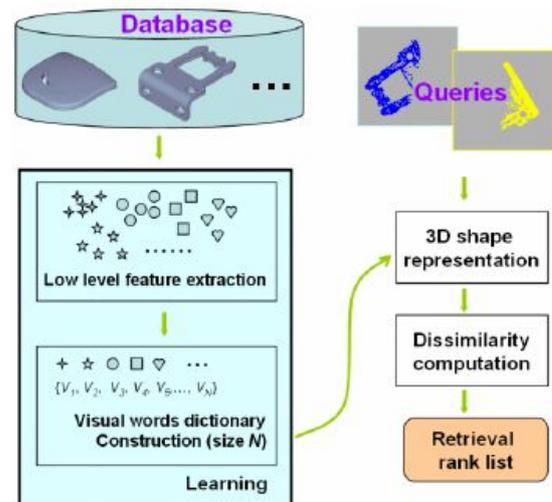


Fig. 28. Representação do método proposto por Xiaolan Li. Figura extraída de [102].

Salgian [104] mostra que através da combinação de descritores locais de imagem em nível de característica, pode-se melhorar de forma significativa a performance de reconhecimento. Os dois descritores utilizados em combinação foram o SIFT[47] e o Keyed Context Patches[63], ambos os descritores já tiveram suas performances demonstradas separadamente. O esquema de correspondência utilizado por ambos os descritores é similar, de forma que as etapas realizadas passam por: extração de características, utilizando os descritores citados, do modelo e da cena e posterior correspondência, geração de hipóteses e seleção da hipótese com maior pontuação de evidências. Para combinação dos descritores, extraem-se as características com o SIFT e o Keyed context patches da imagem de teste e pesquisa-os no banco de dados de treinamento correspondente. Cada característica SIFT

e Keyed context patch contribuiu com uma série de hipóteses que foram verificadas quanto à consistência utilizando as restrições definidas, independentemente do tipo de descritor que a gerou. Da mesma forma que no caso dos descritores sozinhos, o último passo é selecionar o grupo com maior pontuação de evidência. Para os testes no reconhecimento de objetos, utilizou-se 6 objetos facilmente distinguíveis entre eles e comparou-se o erro no reconhecimento do SIFT e do Keyed context patch sozinho e combinados, demonstrando os resultados em termos de taxa de reconhecimento e de erro. No teste de categorização de objetos utilizou-se a base de dados ETH-80 [105] e, igualmente ao teste descrito anteriormente, compararam-se os descritores separados e juntos em termos de performance de reconhecimento para categorização de objetos desconhecidos.

Donghui Wang [106] apresenta um método para reconhecimento de objetos 3D a partir de uma imagem de profundidade sob uma pose arbitrária por meio de um correlacionamento rápido de esfera. Primeiro, todas as visões extended Gaussian image (EGI) sob diferentes pontos de vista são extraídas e combinadas em uma esfera gaussiana para formar uma descrição de características para cada objeto. Depois, a imagem de profundidade examinada, em uma pose arbitrária, é representada como uma característica phase encoded Fourier transform (PFT) pela codificação da fase com transformação de Fourier e a característica PFT é mapeada em um hemisfério gaussiano através da transformação das coordenadas e escalonamento da intensidade. Então, o algoritmo de correlação esférica com base em funções harmônicas esféricas é utilizado para corresponder e medir a similaridade entre o PFT mapeado e a visões EGIs combinadas. Tanto os EGIs de visão combinada quanto a visão EGI de PFT podem ser consideradas como duas funções de característica na esfera da unidade S^2 . Logo a tarefa de reconhecimento de objetos 3D de uma imagem de profundidade pode ser convertida para uma correlação esférica entre duas funções esféricas. Pela detecção do pico máximo global no resultado de correlacionamento, pode-se decidir se a imagem de profundidade de entrada corresponde com um objeto 3D na base de dados. O pico máximo da correlação esférica pode produzir tanto o resultado de reconhecimento quanto a estimativa de pose. Para testar a performance do método proposto dois experimentos foram projetados. Um experimento testa se a característica é robusta a ruído e o outro testa a capacidade discriminatória da característica.

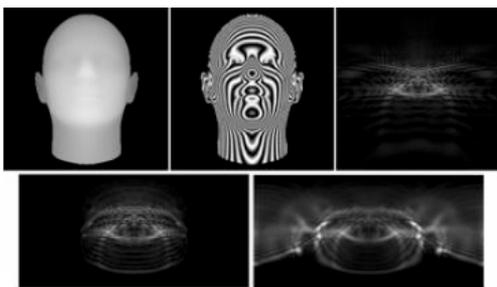


Fig. 29. Representação do método proposto por Donghui Wang. Figura extraída de [106].

Mavrinac [107] apresenta uma abordagem para o reconhecimento de objetos 3D em poses arbitrárias, em relação ao equipamento de visão, fornecendo apenas um conjunto limitado de amostras de visão para treinamento. Esta abordagem envolve a computação de um mapa de disparidade utilizando câmeras estéreo. Com o mapa computado, um conjunto de características (Compactness, First Hu Moment e histograma da distribuição geral da intensidade na imagem) do mapa de disparidade é extraído. Por fim, a classificação destas características é realizada via uma mapa associativo fuzzy para um objeto treinado. Uma base de dados fuzzy associativa similar a encontrada em [108], adaptada para múltiplas imagens de treinamento, é descrita para a parte de classificação. A abordagem básica é armazenar uma tabela de funções de adesão, onde cada classe (tipo de objeto a ser reconhecido) possui uma função de adesão para cada valor de invariante, as quais são construídas a partir dos valores de invariantes "fuzzyficadas" extraídas dos mapas de disparidade do objeto de vários pontos de vista diferentes (conjunto de treinamento). O reconhecimento pode então ser realizado pela comparação dos valores de entrada das invariantes com as funções de adesão em cada classe de objeto, determinando qual classe combina melhor. Os testes foram conduzidos utilizando 3 objetos diferentes com mapas de disparidade extraídos de diferentes pontos de vista e os resultados dos testes foram demonstrados com base na taxa de reconhecimento para os três tipos de objetos.

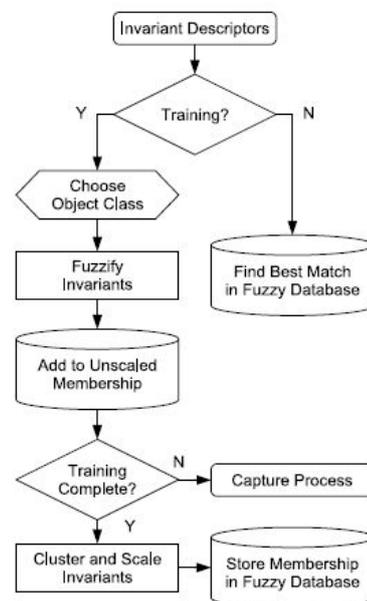


Fig. 30. Representação do método proposto por Mavrinac. Figura extraída de [107].

O método proposto por Elizabeth Gonzáles [109] descreve um método para reconhecimento de objetos 3D com base em clusterização de descritores de Fourier. A abordagem pode ser dividida em duas partes. A primeira parte consiste no cálculo e armazenamento da base de dados de clusters de descritores de Fourier correspondendo a silhueta do objeto visto de múltiplos pontos de vista. A base de dados é estruturada em clusters

utilizando 3 descritores de fourier, permitindo dividir o espaço de silhuetas em zonas onde as silhuetas são grosseiramente similares o que reduz o custo de reconhecimento. A segunda parte é focada na parte de reconhecimento do objeto, no estudo em questão o objeto está sendo visualizado por uma câmera em um ambiente real. As principais etapas deste estágio são: O cálculo dos descritores de Fourier, o processo de classificação (discriminação), a seleção do candidato, cálculo da pose e o uso do algoritmo de Next best view (NBV). Na etapa do cálculo descritores de Fourier, inicialmente executa-se uma etapa de pré-processamento que consiste em filtragem, *thresholding* e extração de contornos. Depois, os pontos de contorno são selecionados em sequência de números complexos e os descritores de Fourier são calculados. Na etapa de classificação, é definido um subconjunto de *clusters* nos quais a silhueta do objeto da cena é combinada. Como próxima etapa tem-se a seleção do candidato, a qual, em essência, compara a silhueta da cena com a silhueta em um subespaço através de uma minimização de erros quadráticos aplicada no módulo dos descritores de Fourier, enquanto que a fase de estimativa da pose é resolvida através do complexo descritores de Fourier. Quando há mais de uma silhueta candidata a solução é ambígua e torna-se necessário aplicar um algoritmo de NBV. Este algoritmo consiste na procura de pontos de visão discriminativos que correspondem ao primeiro objeto. Se na lista de pontos de vista discriminativos não se obteve sucesso, o processo parte para o próximo objeto e assim por diante. A decisão de rejeição ou sucesso vem da avaliação de uma função de custo definido por 3 fatores chaves: a área de trabalho do robô, o grau de discriminação do próximo ponto de visão e o movimento que o robô deve fazer para sair da posição atual para a próxima posição. Se um ponto de visão discriminativo não passa nos requerimentos impostos pela função de custo do NVB, ele será rejeitado. Para a validação foram executados testes experimentais com 28 objetos que foram previamente modelados com 80 e 320 visões. Os resultados dos testes com o sistema ativo de reconhecimento foram comparados tendo como base o número de silhuetas do modelo, o *threshold* para a redução dos descritores de Fourier e o erro médio quadrático entre as silhuetas. Inicialmente os testes foram realizados sem a inserção de ruído e depois com a inserção de ruído e para provar a eficiência do NBV foram realizados testes com a utilização e sem a utilização do método para a resolução de ambiguidade.

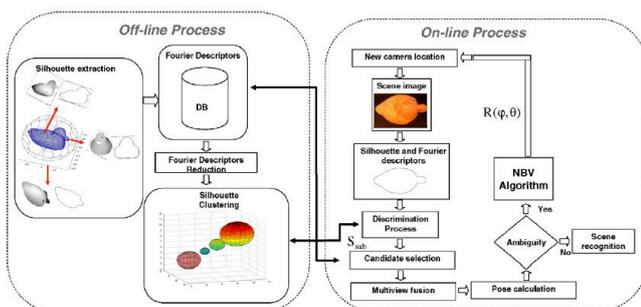


Fig. 31. Representação do método proposto por Elizabeth González. Figura extraída de [109].

O método apresentado por Li-Li Wan [110] mostra um método de classificação com base em Fuzzy KNN e Regras Bayesianas para determinar se um objeto 3D pertence ou não a classe Humano. Inicialmente, características visuais, por meio de Zernike moments descriptor, e estruturais, por meio de grafos de adjacência com base em decomposição de formatos, são extraídas dos objetos de treinamento. As características estruturais são utilizadas para montar a probabilidade a priori utilizada na parte de decisão Bayesiana. As características visuais são utilizadas no fuzzy KNN para atribuir a um objeto desconhecido uma probabilidade de pertencer a uma certa classe e essa probabilidade serve como entrada para decisão bayesiana. Depois, extraí-se o grafo de adjacência e as características visuais de uma visão selecionada do objeto 3D de consulta com a finalidade de determinar se o objeto pertence ou não a classe humano. As características visuais do objeto de consulta são utilizadas no KNN fuzzy para atribuir a probabilidade do objeto de consulta pertencer a classe, sendo esta uma das entradas da decisão bayesiana, e as características estruturais são utilizadas como outra entrada da decisão bayesiana. Estas duas entradas associadas a probabilidade a priori servem para determinar se o objeto de entrada pertence ou não a classe humano. Para os testes realizados foi utilizada a base de dados Princeton shape benchmark [39] e os resultados foram avaliados em termos de acurácia de classificação.

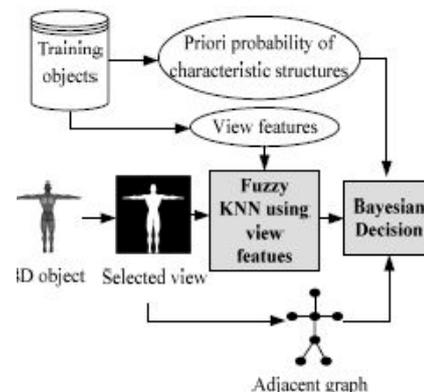


Fig. 32. Representação do método proposto por Li-Li Wan. Figura extraída de [110].

Hui chen [111] propõe um método que combina a incorporação de característica para a recuperação rápida de descritores de superfície, medidas de similaridade para correspondência e uma técnica de aprendizagem com base em Support Vector Machine (SVM) para ranquear as hipóteses na parte de reconhecimento de objetos 3D. O sistema proposto pode ser descrito pelos seguintes passos: primeiro, dado um modelo do objeto, extraí-se os pontos de característica que são definidos ou como mínimo local ou como o máximo local dos valores de índice de formato. Depois, calcula-se o descritor Local Surface Patch (LSP) para os pontos de característica e seus vizinhos. A representação por LSP inclui um ponto de característica, seu tipo de superfície, o centro do remendo e o histograma 2D dos valores de índice versos o produto cruzado da normal da superfície no ponto de característica

e seus vizinhos. Com base no tipo de superfície de uma LSP, um LSP é classificado em três tipos e para cada tipo aplica-se um algoritmo de incorporação de característica para incorporar o vetor original de característica em um espaço dimensional menor. Dado uma imagem de teste, repete-se o mesmo processo de mapeamento dos LSPs em seus respectivos espaços dimensionais de menor dimensionalidade com base no tipo de superfície. Através da procura do vizinho mais próximo do vetores de característica incorporados, acha-se, potencialmente, o LSP correspondente entre o par modelo-teste. As correspondências iniciais são filtradas e agrupadas para remover falsas correspondências utilizando restrições geométricas. Com base em um conjunto de correspondências, um conjunto de características é computada para medir a similaridade entre o par modelo-teste. Então, as hipóteses são ranqueadas utilizando o ranqueamento do algoritmo de aprendizagem do SVM para gerar uma pequena lista de modelos candidatos a verificação. Os parâmetros do classificador SVM são aprendidos de um subconjunto da base de dados. Para a verificação, aplica-se a correspondência entre superfícies através do Iterative Closest Point (ICP), no qual a transformação inicial é obtida dos LSPs correspondentes. O *framework* proposto foi utilizado em experimentos extensivos em duas bases de dados de orelhas 3D (University of Notre Dame (UND) [112] e University of California at Riverside (UCR) [113]) e os resultados mostrados em termos de taxa de reconhecimento, indexação e tempo utilizado.

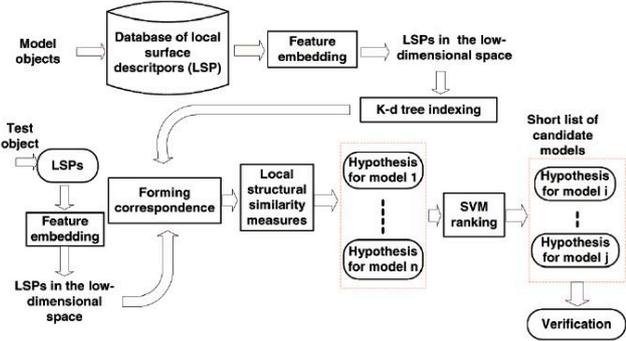


Fig. 33. Representação do método proposto por Hui Chen. Figura extraída de [111].

Um *framework* para o reconhecimento de objetos e estimativa de pose em ambientes reais é apresentado por Hyunjun Kim [114]. Neste *framework*, propõe-se a seleção de evidências e a coleta com base no teorema de Bayes, onde as evidências são características tais como Scale Invariant Feature Transform (SIFT), linha, cor e assim por diante. A maior vantagem do método é estimar a probabilidade mais facilmente pelo uso da regra bayesiana e selecionar um conjunto ótimo de evidências automaticamente. O *framework* pode ser descrito pelos seguintes passos: inicialmente utiliza-se uma câmera estéreo para coletar informações 3D com as imagens 2D. No monitoramento do local, a detecção de mudanças ambientais com base em informações de entrada 2D e 3D e a geração de Regions of Interest (ROI) para o foco de atenção e informação do ambiente na ROI são trans-

feridos para a seleção de evidências. A seleção de evidência seleciona a melhor característica para o ambiente com base nas informações ambientais utilizando o teorema de Bayes. As características são extraídas e correspondidas utilizando características selecionadas na parte de seleção de evidências, e interpretações são geradas. Cada interpretação gerada pela correspondência de características são fundidas em um filtro de partículas. O comportamento do robô é determinado para a minimização da incerteza do próximo ciclo de reconhecimento do objeto no processo de coleta de evidências, utilizando os resultados de fusão pelo filtro de partículas. O funcionamento do filtro de partículas é apresentado em outro trabalho elaborado pelo autor [115]. Os resultados do *framework* proposto foram demonstrados através de um teste experimental em um ambiente com múltiplos objetos desordenados em uma mesa em frente a posição inicial do robô.

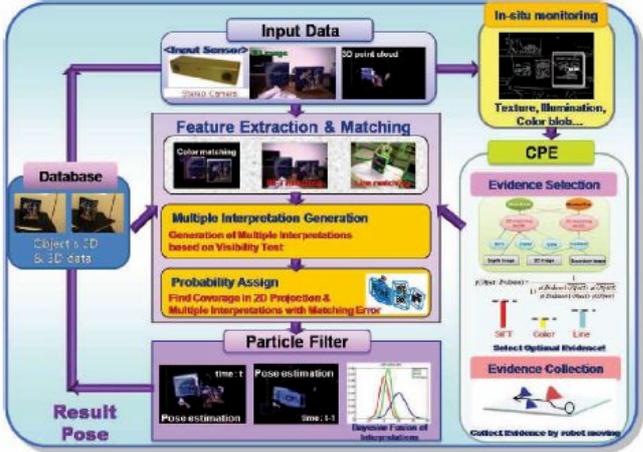


Fig. 34. Representação do método proposto por Hyunjun Kim. Figura extraída de [114].

Um sistema para o reconhecimento de objetos com base em algoritmos implementados em GPU e utilizando características SIFT é proposto por Gee-Sern Hsu [116]. O sistema proposto pode ser descrito pelos seguintes passos: Extração de características SIFT e posterior reconhecimento utilizando tais características. Aplicação de homografia para obter as orientações fora do plano do objeto; Por fim, utilização da visão estéreo para obter a profundidade do objeto. Para a extração das características SIFT, quatro passos fundamentais são executados: detecção extrema escala-espaco; localização de pontos chave; atribuição da orientação dos pontos chave; e determinação dos descritores de pontos chave; Para o reconhecimento de objetos através de características SIFT, primeiro as características são extraídas de um conjunto de imagens de treinamento, como anteriormente descrito, e armazenadas em uma base de dados. Uma imagem de teste, presumivelmente desconhecida dos objetos, é então correspondida através da comparação individual de cada característica da imagem de teste com a as características armazenadas na base de dados, encontrando candidatos de correspondências de características utilizando um algoritmo de nearest-neighbor. A comparação das características utiliza um teste da taxa de distância para

descartar possíveis correspondências erradas. Se a taxa de distância do nearest-neighbor para o segundo nearest-neighbor for maior que 0.8 a correspondência é considerada correta [47]. Depois, aplica-se a homografia, a qual é definida em um espaço 2D como o mapeamento entre um ponto visto em uma câmera para o mesmo ponto visto em uma segunda câmera, fornecendo assim um método para a composição de um objeto 2D ou 3D em uma imagem com a pose correta. Por fim, define-se as informações de profundidade do objeto com visão estereó e estas informações são utilizadas para controlar um braço de robô. Para demonstrar os potenciais de aplicação do método proposto, implementou-se o fluxo de trabalho completo para o controle do braço robótico em GPU, testando o ganho obtido em relação a sua implementação em CPU e se o braço do robô poderia tocar com precisão os objetos-alvo.

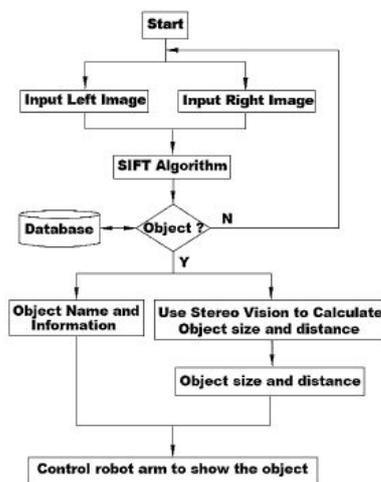


Fig. 35. Representação do método proposto por Gee-Sern Hsu. Figura extraída de [116].

Osman [117] apresenta uma análise de performance de dois *moments*, denominados Hu's moments [95] e Zernike's moments [118], para o reconhecimento de objetos. O estudo realizado mostra que com uma adaptação a uma técnica de múltiplas visões [119], os dois *moments* são suficientes para modelar objetos 3D. Na parte de reconhecimento, propõe-se o uso de um classificador neuro-fuzzy denominado Multiple Adaptive Network based Fuzzy Inference System (MANFIS) [120] para correspondência e classificação entre objetos. O *pipeline* utilizado pode ser descrito pelos seguintes passos principais: Aquisição de imagens do objeto de vários pontos de visão, extração dos Hu's e Zernike's moments, utilização desses *moments* como entrada na rede neuro-fuzzy MANFIS para treinamento e posterior classificação dos objetos. Dois tipos de objetos foram escolhidos para analisar a performance do sistema proposto. Cada tipo de objeto consiste de 11 objetos 3D. O primeiro tipo contém objetos de formatos simples, tais como cilindro, caixa, esfera, etc. O segundo tipo contém objetos de formato livre. Os resultados foram demonstrados em termos de acurácia de classificação para os dois *moments* utilizados separadamente para os dois tipos de objetos.

Himmelsbach [121] descreve um sistema de percepção com base em LIDAR para mobilidade de robôs. O sis-

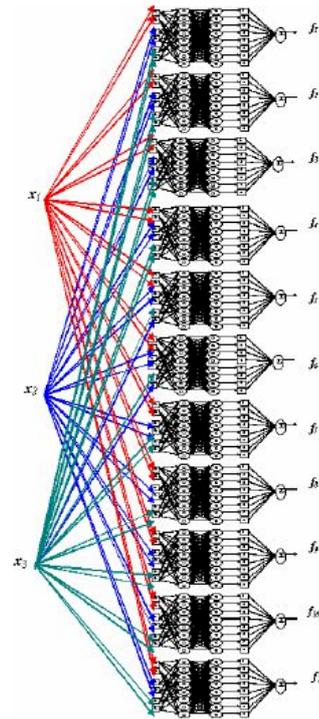


Fig. 36. Representação do método proposto por Osman. Figura extraída de [117].

tema de percepção é decomposto em três passos principais: segmentação, classificação e rastreamento. A segmentação é realizada em um grid de ocupação, fornecendo componentes conexas das células do grid não pertencentes a superfície chão, no caso objetos. Com uma operação eficiente, determina-se todas as medições dos pontos 3D LIDAR correspondendo aos objetos segmentados. Na etapa de classificação, extraem-se características de um dos objetos na nuvem de pontos identificados no passo anterior, capturando a distribuição local das propriedades espaciais e de refletividade extraídas sobre um volume de suporte com tamanho fixo ao redor de cada ponto. Para obter uma descrição compacta de um objeto para classificação, constrói-se histogramas sobre as características computadas para cada ponto do objeto. Em um *framework* de aprendizagem supervisionada, um classificador Support Vector Machine (SVM) é treinado para discriminar as classes de interesse, com base em exemplos de nuvens de pontos rotulados manualmente. O *framework* para classificação de objetos proposto foi testado na tarefa de discriminar objetos pertencentes a classe carros de passageiros dos outros tipos de objetos. O resultado desta tarefa foi avaliado em termos de acurácia. Para validar o sistema de percepção como um todo, o *framework* foi integrado em um sistema para rastreamento de objetos em um cenário de comboio, onde o MuCAR-3 deve seguir de forma autônoma o caminho pego pelo veículo liderando o comboio.

Gibbins [122] propõe e avalia o uso de métricas locais, tais como Zernike moments, curvatura e representações de cor e Spin utilizadas no reconhecimento de objetos 3D, para rotulamento/classificação das estruturas de terrenos. Inicial-

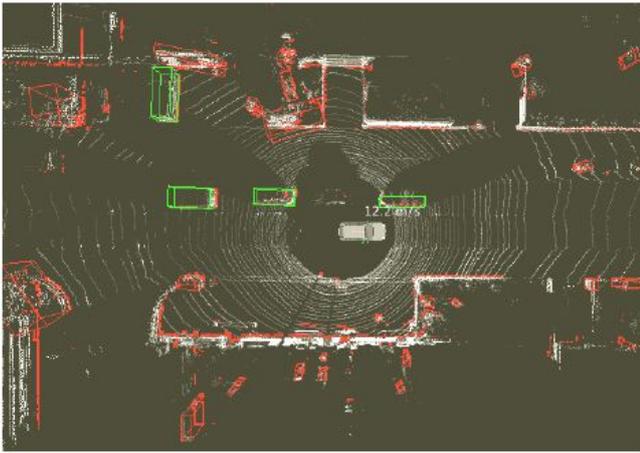


Fig. 37. Representação do método proposto por Himmelsbach. Figura extraída de [121].

mente, o autor apresenta diversos tipos de características locais a serem utilizadas e avaliadas no processo de classificação do terreno. São elas: curvaturas principais e índice de formato; 3D moments invariantes a translação, escala e rotação, proposto em [95]; Zernike moments [123], [124]; Características Spin image [16]; Medidas com base nos dados de cor e em tons de cinza de cada ponto da imagem [125], [126]; Cinco características da superfície local, imagem e Digital Terrain Elevation Modelo (DTED) propostas em [126]; E quatro características com base na inclinação do solo, na variância da altura e duas com base na variância em relação a um ajuste planar local dos dados. Para avaliar os sete tipos de características propostas utilizou-se amostras de dados reais coletados usando um escâner LADAR de baixo custo e carga ótica. Esses dados foram convertidos em uma nuvem de pontos com informações de cores associadas. Estas amostras de dados foram manualmente classificadas em 4 tipos com base em observações do terreno. Para cada tipo de característica, exceto a proposta em [126], as estimativas de características foram computadas sob vizinhanças esféricas de raios 1,2 e 4 metros. Para minimizar qualquer correlação entre os dados de teste e de treinamento, os dados de terreno foram separados em dois conjuntos independentes para treinamento e teste. Nos experimentos realizados, um classificador por Vector quantisation (VQ) foi empregado em conjunto com linear discriminant analysis (LDA) para redução das características. Os resultados foram demonstrados em termos de resultado de classificação para cada tipo de característica utilizada individualmente e com algumas combinações de características.

Um método de reconhecimento de objetos 3D com base em visões para sequências de vídeos é apresentado por Noceti [127]. O método proposto representa implicitamente a informação 3D de um vídeo através de características invariantes no tempo, ou seja, características que são distintivas em espaço e suaves e estáveis no tempo. Para tanto, são extraídas características locais invariantes a escala das imagens e estas são rastreadas ao longo da sequência do vídeo por meios do filtro de Kalman, formando a trajetória das características. Para uma dada trajetória, obtém-se uma

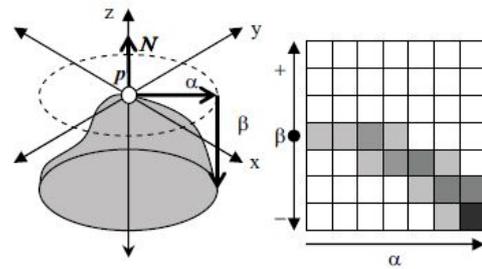


Fig. 38. Representação do método proposto por Gibbins. Figura extraída de [122].

característica invariante ao tempo através das médias das descrições para cada *keypoint* pertencente a mesma trajetória. Obtém-se assim um modelo espaço-temporal composto de todas as características invariantes ao tempo observadas no vídeo. Na fase de modelagem adquire-se uma sequência do vídeo de um objeto de interesse no tempo, observando-o em um ambiente relativamente controlado de diferentes pontos de vista significativos para propósito de reconhecimento. Para a modelagem dessas características invariantes ao tempo utilizou-se Harris corners no espaço de escala descritos utilizando o SIFT. Na fase de reconhecimento, realizado em dois estágios, explora-se a coerência espacial e temporal dentro de cada modelo. Considerando uma cena e um objeto o algoritmo consiste em ver se existe o objeto na cena. No primeiro estágio, utiliza-se uma correspondência de vizinho mais próximos entre modelo e teste, fornecendo uma hipótese inicial da presença do objeto na cena. Para cada característica invariante ao tempo presente no objeto hipótese, utiliza-se uma intersecção de histogramas entre descritores espaciais de aparência para conferir se a cena contém características similares, seta-se um limiar mínimo para obter a coleção de correspondências robustas entre as características. No segundo estágio, procura-se confirmar ou rejeitar a hipótese do estágio anterior utilizando restrições espaço-temporais e realizando um procedimento de correspondência para trás, do modelo de teste para o modelo de treinamento, o qual possui um limiar de similaridade informado para decidir se o objeto está ou não presente na cena. Para os experimentos com o método proposto, foi considerado uma base de dados com 20 objetos. Cada objeto é modelado com uma sequência do vídeo (treinamento) com uma câmera parada enquanto o objeto rotaciona em uma mesa giratória em um ambiente controlado. Para os testes os vídeos gravados não foram adquiridos em tais circunstâncias com a finalidade de mostrar uma seleção variada de possíveis diferentes condições. A performance dos experimentos foram medidos em termos de taxa de precisão e lembrança. O método proposto também foi comparado com outros dois métodos (bags-of-keypoints associado com Nearest Neighbors e bags-of-keypoints associado com Support Vector Machine).

O trabalho apresentado por Atmosukarto [7] demonstra a forma de representação de formatos 3D e uma metodologia para classificação desenvolvida para o uso em estudos de dismorfologia craniofacial. Este trabalho tem como base outro trabalho do mesmo autor apresentado em 2008 [128]

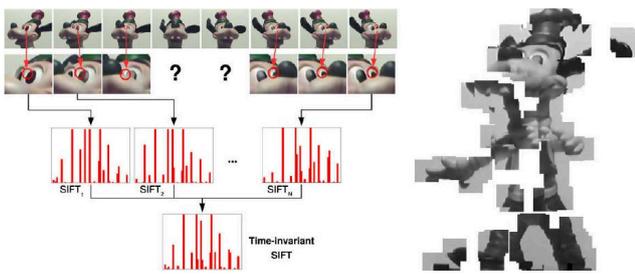


Fig. 39. Representação do método proposto por Nocetti. Figura extraída de [127].

e a metodologia proposta inicia com o reescalonamento dos objetos para ajustá-los a uma *bounding box* de tamanho fixo. O próximo passo consiste de duas fases que são: extração de características em um nível mais baixo e agregação de características de um nível médio. A extração de características de baixo nível é realizada com a aplicação de operadores de baixo nível (*absolute Gaussian curvature*, *BesJain curvature categorization*, *azimuth-elevation of surface normal vectors*) em cada ponto da superfície da malha, sendo que ao final da aplicação desta etapa cada ponto terá um valor ou alguns valores representando as características de baixo nível. A segunda fase realiza uma agregação de características do nível médio e computa um vetor de valores para uma dada vizinhança de cada ponto da malha da superfície. Estas características de nível médio são utilizadas para determinar pontos de saliência de um objeto, os quais são aprendidos especificamente para cada aplicação. No propósito do estudo em questão são utilizados os pontos de saliência que servem melhor para a classificação de distúrbios craniofaciais. Para encontrar os pontos de saliência no objeto 3D em questão, foi selecionada uma abordagem de aprendizagem. Um classificador de pontos salientes foi treinado em um conjunto de pontos de treinamento fornecido por especialistas na aplicação. Histogramas das características de baixo nível são utilizadas para o treinamento do classificador, fazendo com que o classificador aprenda os pontos de saliência no objeto 3D. O classificador utilizado para a aprendizagem dos pontos de saliência foi o Support Vector Machine (SVM) e foram utilizados para os testes, uma base de dados com cabeças humanas contendo distúrbios craniofaciais e, em outro teste, foram colocadas cabeças de animais e humanas para serem testadas quanto a aprendizagem dos pontos de saliência. Para a classificação a maioria dos métodos requer o uso de um descritor 3D ou uma assinatura para descrever o formato e as propriedades dos objetos. A assinatura do método proposto tem como base os pontos de saliência 3D mapeados em planos 2D através de uma transformação longitude-latitude. A classificação dos objetos 3D é então realizada através do treinamento de um classificador utilizando o mapa 2D dos objetos classificando em cabeças de animais, humanas ou objetos específicos dependendo da base de dados utilizada. Para a comparação da acurácia da metodologia proposta testou-se a acurácia da mesma contra alguns descritores de objetos 3D estado da arte presentes na literatura (*light field descriptor (LFD)*, *ray-based spherical harmonics (SPH)*, *shape*

distribution of distance between random points (D2) e *absolute angle distance histogram (AAD)*). Para a comparação da performance foram utilizadas 4 estatísticas comumente utilizadas (nearest-neighbor classification accuracy, first-tier classification accuracy, second-tier classification accuracy e F-measure). Foram realizados testes em duas base de dados de distúrbios craniofaciais e em duas base de dados de objetos 3D.

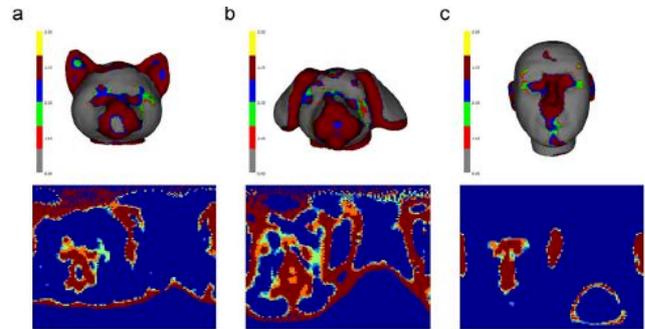


Fig. 40. Representação do método proposto por Atmosukarto. Figura extraída de [7].

Um *framework* que mantém a ambiguidade na correspondência de características para aumentar a performance de um sistema de reconhecimento de objetos 3D é apresentado por Hsiao[129]. Neste *framework* os autores introduzem a ambiguidade no processo de correspondência pela quantização das características dos modelos e pela associação de cada descritor quantizado com, potencialmente, múltiplas localizações no modelo. Quando uma característica da imagem é correspondida com uma característica quantizada, ela fica associada com todas as possíveis localizações daquela característica. Durante o teste das hipóteses, a correspondência mais provável fornecendo a pose atual pode então ser determinada. O *framework* permite escolher a hipótese mais provável dado o que foi visto, e combina ambas, características ambíguas e únicas, em um *framework* unificado. No intuito de validar a performance do *framework* no reconhecimento de objetos com base em características, dois conjuntos de experimentos foram realizados. O primeiro conjunto avalia a habilidade do algoritmo em reconhecer objetos em imagens, enquanto que o segundo conjunto avalia a acurácia do algoritmo em recuperar a pose 3D e a orientação dos objetos nas imagens. Os algoritmos base utilizados na avaliação são os propostos por Gordon e Lowe[130], EPnP [131] e Collen [132]. Todos estes sistemas utilizam modelos 3D esparsos de objetos com características SIFT para o reconhecimento e compartilham uma metodologia parecida, os três tem como objetivo estimar uma transformação de um modelo 3D com referência ao *frame* da câmera para cada instancia de objeto de classe na imagem. Em cada um desses métodos foram incorporados as características *simulated affine* e a quantização, via *view-constrained RANSAC*, separadamente e depois juntos, para mostra os ganhos de performance em cenas complexas. A base de dados utilizada para os experimentos foi um conjunto de 10 objetos domésticos com 62 imagens por objeto [133]. Os resultados com esses três algoritmos base associado com o

framework proposto foram demonstrados em termos de taxas de precisão e lembrança.

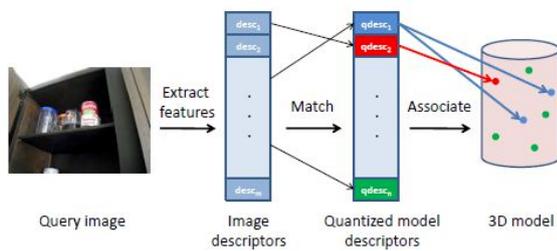


Fig. 41. Representação do método proposto por Hsiao. Figura extraída de [129].

Chih-Hong Kao [134] propõe um sistema para o reconhecimento de navios com o propósito de pesquisar e desenvolver uma efetiva captura do contorno do navio no mar. A arquitetura do sistema proposto, inicia com a extração da característica de contorno utilizando o método Gradient Vector Flow (GVF) [135]. O detector de bordas Canny é utilizado para realizar a detecção da borda. O contorno que se obtém utilizando o GVF não está sempre fechado. Logo deve-se fazer um processo de fechamento enquanto se obtém o contorno da borda. No processo de fechamento, utiliza-se o método de linha Bresenham's para conectar dois pontos separados. Os autovalores geométricos são calculados após obter-se o contorno detalhado da imagem. Utiliza-se a complexidade do objeto e a proporção entre o eixo mais longo e mais curto do objeto como critério para autovalores de reconhecimento de imagem. Todos esses autovalores geométricos são utilizados para uma correspondência grosseira. Utiliza-se o o descriptor Fourier para realizar a correspondência detalhada. Após obter os autovalores de geometria do objeto, pode-se proceder para comparação da imagem com a base de dados. Esta etapa pode ser separada em dois passos principais: correspondência grosseira e a fina. O propósito desses dois passos é aprimorar a eficiência no reconhecimento. A correspondência grosseira utiliza principalmente a complexidade da borda e a proporção dois eixos maior e menor para fazer a comparação. A correspondência fina utiliza os descritores de Fourier do objeto a ser identificado para fazer a comparação. Os experimentos com o método proposto foram realizados com imagens de modelos de navios e os resultados dos experimentos foram demonstrados em termos de valores médios de todos os *root mean square* para a base de dados de navios.

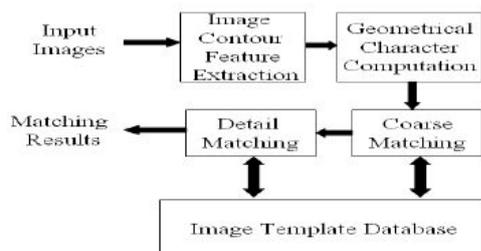


Fig. 42. Representação do método proposto por Chih-Hong Kao. Figura extraída de [134].

Drost [136] propõe um método que cria uma descrição global do modelo com base em característica de pares de pontos orientados e corresponde este modelo localmente utilizando um esquema de voto rápido. A descrição global do modelo consiste de todos as característica de pares de pontos do modelo e representa um mapeamento do par de ponto no espaço de característica para o modelo, onde características similares do modelo são agrupadas. As características de pares de ponto descrevem a posição relativa e orientação de dois pontos orientados, o que é similar ao *surflet-pair feature* apresentado em [137]. Inicialmente, assume-se que tanto cena quanto modelo são representados como um conjunto finito de pontos orientados, onde uma normal está associada com cada ponto. Tal representação pode ser facilmente calculada de uma malha ou de uma nuvem de pontos. Na fase offline do método, a descrição global do modelo é criada. Na fase online, um conjunto de pontos de referência na cena é selecionado. Todos os outros pontos na cena são emparelhados com os pontos de referência para criar características de pares de pontos. Estas características são correspondidas com as características do modelo contidas no modelo de descrição global e um conjunto de possíveis correspondências é recuperado. Cada correspondência em potencial vota para uma pose de objeto através de um eficiente esquema de voto, o qual é similar ao Generalized Hough Transform e eficiente dado que as coordenadas locais possuem apenas três graus de liberdade. Uma vez que a coordenada local ótima é encontrada, a pose global do objeto pode ser recuperada. Para o esquema de votação, utiliza-se um vetor bi-dimensional, no qual o número de linha é igual ao número de pontos amostrados do modelo e o número de colunas corresponde ao número de passos de amostragem do ângulo de rotação. Esse vetor acumulador representa o espaço discreto das coordenadas locais para um ponto de referência fixo. Para a parte de votação propriamente dita, um ponto de referência da cena é emparelhado com outro ponto da cena e a superfície do modelo é procurada por pares de ponto que possuem uma distância e orientação da normal similares. Para cada possível posição dos pares de pontos da cena na superfície do modelo, o ângulo de rotação é calculado e um voto é dado para a coordenada local. Após todos os pontos serem processados, o pico no vetor acumulador corresponde a coordenada ótima local, da qual um movimento global rígido pode ser calculado. Por razões de estabilidade, todos os picos que ganharem uma certa quantidade de votos relativos ao pico máximo são utilizados. Sob este esquema de votação adiciona-se uma etapa adicional para filtrar poses incorretas e aumentar a acurácia do resultado final. Esta etapa adicional *clusteriza* as poses recuperadas de forma que todas as poses em um *cluster* não possuam uma diferença em translação e rotação maior que um limiar definido. A pontuação do *cluster* é a soma das pontuações das poses contidas nele. Após encontrar o *cluster* com maior pontuação, a pose resultante é calculada pela média das poses contidas no *cluster*. Dado que múltiplas instâncias do objeto podem estar na cena, vários *clusters* podem ser retornados pelo método. O método proposto foi avaliado em conjuntos de dados reais e sintéticos. O conjunto de dados reais utilizados foi fornecido por [81], [138] e os resultados foram demonstrados em termos de taxa de reconhecimento na

presença de oclusão e desordem.

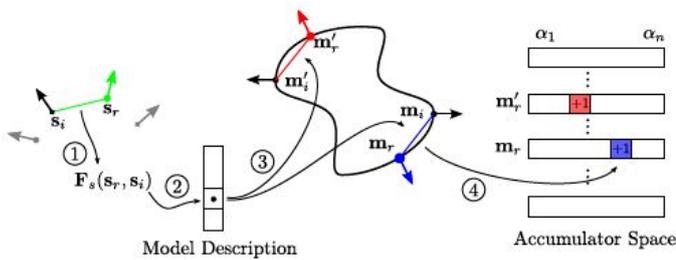


Fig. 43. Representação do método proposto por Drost. Figura extraída de [136].

Yangye Wang [139] apresenta um método com base em *edge moment* de sub-áreas para o reconhecimento de objetos 3D de *wireframe*. No reconhecimento de objetos 3D, o *wireframe* do objeto 3D é dividido em várias sub-áreas em formato de leque com o centroide como o centro. Através do uso do método de cálculo de *edge moments* melhorado, extensão do método proposto em [140], os *edge moments* de baixa ordem para cada sub-área foram computados. Então esses *edge moments* da baixa ordem são combinados em um novo conjunto de vetor de características de *edge moments* para ser a característica do objeto 3D. Este vetor de característica é apresentado para uma wavelet neural network para reconhecimento. Para os experimentos, 10 categorias de objetos 3D foram escolhidas para o reconhecimento. Cada categoria tem 20 amostras de treinamento e 20 amostras de teste. As amostras de treinamento e de teste são todos *wireframe* de objetos 3D. Para comparar a performance do método proposto, 5 amostras de cilindros diferentes e 5 amostras de cubos foram selecionados das amostras de treinamento, respectivamente, descrevendo e mostrando os resultados obtidos em cada etapa para estas amostras selecionadas.

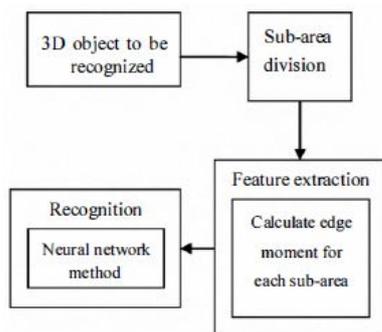


Fig. 44. Representação do método proposto por Yangye Wang. Figura extraída de [139].

Naikal [141] propõe um sistema de reconhecimento de objetos distribuído para detecção, compressão e reconhecimento de objetos 3D utilizando uma rede câmeras inteligentes sem fio. Os autores apresentam primeiro uma base de dados pública de reconhecimento de objetos com múltiplas visões denominada Berkeley Multiview Wireless (BMW) database. Esta base, contém a aparência 3D de 20 construções de referência, amostradas por 5 câmeras de baixa resolução de múltiplos

pontos de vista. Depois, eles revisam e avaliam métodos para extrair características e comprimir suas representações esparsas. Por fim, propõem um método de reconhecimento com múltiplas visões para classificar o objeto observado pelas câmeras. Dado o foco do estudo deste relatório, iremos analisar diretamente o método de reconhecimento. Este método, tem como entrada histogramas de múltiplas visões e como saída um rótulo classificando o objeto 3D. Dado um conjunto grande de características da imagem, como por exemplo características SIFT, pode-se construir uma árvore de vocabulário utilizando o k-means hierárquico, onde o k representa o fator de ramos da árvore [142]. No nível mais alto da árvore todos os descritores de característica são particionados em k *clusters*, com a média de cada *cluster* representando o centro do *cluster*. Em cada nível mais baixo, aplica-se o k-means dentro de cada *cluster* anterior com a finalidade de particionar o espaço em k *clusters*. O processo continua até existirem k^H *clusters* no nível *H-th*. Com a árvore de vocabulário construída, os descritores de característica em cada imagem de treinamento são propagadas para baixo na árvore. Então, um histograma ponderado term-frequency inverse-document-frequency (tf-idf) pode ser definido para cada imagem de treinamento da seguinte forma: Primeiro, atribua um peso com base na entropia para cada nodo folha de característica quantificado na árvore de vocabulário. Com esse peso computado desta maneira, todos os elementos do histograma de treinamento e teste são multiplicados com esta função de peso para alcançar o esquema de ponderação tf-idf. Para cada categoria de objetos, múltiplos histogramas ponderados são gerados, para todas as imagens de treinamento daquele objeto, e agrupados em um conjunto para formar o conjunto de treinamento. Durante a fase de teste, os descritores de característica são extraídos para cada imagem de uma única visão consultada e propagados para baixo na árvore de vocabulário da mesma forma para obter um único histograma ponderado. Então, atribui-se uma pontuação de relevância a imagem consultada com base na diferença 11-normalizada entre o histograma ponderado da imagem consultada e o conjunto de treinamento. Quando múltiplos histogramas do objeto consultado estão disponíveis, utiliza-se um outro método para realizar a classificação que leve em consideração a informação das múltiplas visões. Neste caso, a mediana da pontuação de relevância de uma única visão é utilizado para determinar a pontuação média de relevância das múltiplas visões. Por fim, o rótulo da categoria do objeto é atribuído com base na categoria de objeto que alcança um valor mínimo de pontuação de relevância. Nos experimentos, com base na base de dados BMW, foram selecionados para a comparação três descritores de característica para representar a aparência do objeto (SIFT [42], SURF [143], e CHoG [144]). Dois cenários de teste foram projetados para avaliar o esquema de reconhecimento distribuído. No primeiro cenário, imagens capturadas concorrentemente por múltiplas câmeras em um ponto de visão são utilizadas para determinar a categoria do objeto. A performance foi avaliada com o uso de 1, 2 e 3 câmeras. No segundo cenário, imagens capturadas de um dos três ponto de visão são randomicamente escolhidas da mesma categoria de teste para o reconhecimento. Os resultados para os dois cenários foram demonstrados em termos de taxa de

reconhecimento.

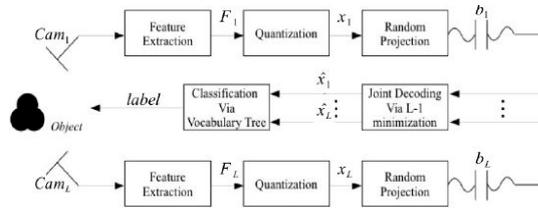


Fig. 45. Representação do método proposto por Naikal. Figura extraída de [141].

Um descritor para nuvem de pontos 3D que codifica ponto de visão e geometria é apresentado por Rusu [145]. O descritor proposto, denominado Viewpoint Feature Histogram (VFH), é composto por duas componentes, onde a segunda componente é inspirada no descritor Fast Point Feature Histogram (FPFH)[137]. A primeira componente do VFH é uma componente de ponto de visão, a qual é computada através da coleta de um histograma dos ângulos que a direção do ponto de visão faz com cada normal. A segunda componente do VFH mede os relativos pan, tilt e yaw angles como descritos em [137], mas agora medidos entre a direção de ponto de visão no ponto central e cada uma das normais da superfície. Este descritor é empregado em um sistema para o reconhecimento do objeto e de sua pose. A arquitetura do sistema emprega os seguintes passos: Aquisição de imagens pelo par estéreo e posterior cálculo do mapa de profundidade. Cálculo das normais na cena; Identificação e segmentação de planos e clusterização, em um espaço euclidiano, dos pontos remanescentes na nuvem de pontos pertencentes a objetos não planares; Cálculo do VFH sobre objetos grandes o suficiente. Por fim uso do Fast approximate K-NN para classificar o objeto e sua visão. O fast approximate K-Nearest Neighbors (K-NN), implementado na FLANN library [146], foi utilizado devido ao seu treinamento online e da disponibilidade de uma memória grande. Para avaliar o descritor e a arquitetura do sistema proposto, utilizou-se um conjunto de dados compostos de mais de 60 objetos de cozinha. O resultados dos experimentos realizados foram comparados com os resultados obtidos com o Spin Images em termos de reconhecimento de objeto e estimativa de posição.

O trabalho apresentado por Mian [138] estende um trabalho apresentado anteriormente pelo autor [147] e propõe um algoritmo de detecção de pontos chave em múltiplas escalas para a extração de características locais invariantes a escala. Em relação ao trabalho anteriormente apresentado as seguintes diferenças podem ser ressaltadas: enquanto o trabalho anterior utiliza uma decisão binária para selecionar ou rejeitar um ponto chave, o trabalho atual apresenta uma medida de qualidade para ranquear os pontos chave. No trabalho anterior os pontos chave e as características, extraídas a partir desses pontos chave, foram extraídas em uma escala fixa. No entanto, no trabalho atual propõe-se uma técnica para automaticamente selecionar uma escala apropriada em cada ponto chave e se extrai características invariantes a escala. Explorando a primeira diferença, a medida de similaridade apresentada para ranquear os pontos chave tem como base a curvatura principal da superfície local dentro de uma vizinhança do ponto chave.

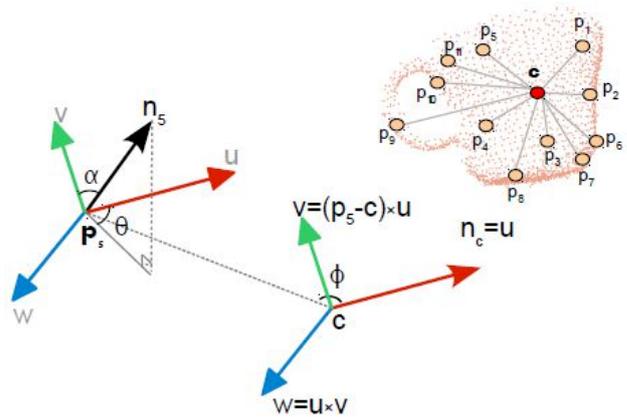


Fig. 46. Representação do método proposto por Rusu. Figura extraída de [145].

A curvatura principal é calculada com base em encaixe de superfície [148] e a qualidade do ponto chave é computada com base em uma fórmula descrita. A segunda diferença está na seleção automática de escala. A escala correta para um ponto chave é determinada como o tamanho da vizinhança para a qual a proporção entre os eixos principais da vizinhança atinge um máximo local. Quando o ponto atinge um máximo local e é, portanto, selecionado como ponto chave na respectiva escala, isto implica que as características subsequentes serão extraídas nesta escala. Para a parte de correspondência utiliza-se os valores de profundidade da superfície previamente calculada para formar um vetor de características e emprega-se um algoritmo de correspondência que *clusteriza* possíveis transformações entre o objeto consultado e os modelos da base de dados para a recuperação de objetos 3D a partir de cenas desordenadas. Para a redução da dimensionalidade da representação das superfícies locais, utiliza-se o PCA. Os experimentos de recuperação de objetos 3D foram realizados utilizando a base de dados pública disponível em [81] e os resultados foram apresentados de forma qualitativa, quantitativa e comparados com outras duas técnicas (tensor matching[81] e spin-images [16]) em termos de taxa de reconhecimento por quantidade de oclusão.

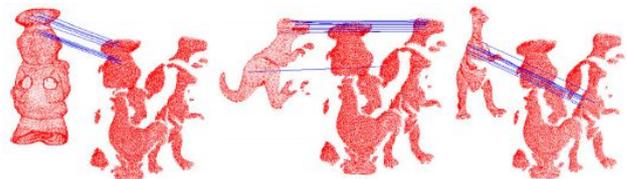


Fig. 47. Representação do método proposto por Mian. Figura extraída de [138].

Tombari [149] propõe um método para fornecer um Local Reference Frame (LRF) único com a finalidade de aprimorar os descritores de formato 3D. Mais especificamente, o autor propõe um descritor que tem como base a formulação do 3D Shape Context (3DSC)[98], mas que diferentemente do método original, utiliza LRF repetíveis e não-ambíguos com o intuito de alcançar uma descrição única em cada ponto de

característica. O método proposto pode não só fornecer um eixo normal repetível, mas também fornece um par único de direções situadas no plano tangente. A principal vantagem de tal abordagem, denominada Unique Shape Context (USC), é que se necessita computar, para cada característica do modelo, apenas um descritor sobre a rotação indicada pelos vetores repetitivos no plano tangente. Isto, não só fornece uma melhor escalabilidade através de uma base de dados, mas também simplifica o processo de correspondência, dado que cada característica da cena necessita ser correspondido com apenas uma instância de descritor para todas as características do modelo. Para gerar um único e não ambíguo LRF utiliza-se a abordagem proposta em [150]. Para gerar o descritor proposto, primeiramente, computa-se o LRF sob um região ao redor de um ponto de característica. Então, o volume esférico ao redor do ponto de característica é subdividido de forma única por meio de uma rede esférica orientada com três direções repetíveis fornecidas pelo LRF. Cada *bin* da rede acumula uma soma ponderada dos pontos da superfície de forma análoga a realizada pelo 3DSC. O método proposto foi validado por meio de dois experimentos, um comparando o USC com o 3DSC em um cenário típico de correspondência de características e o outro no reconhecimento de objetos em cenas bagunçadas. As bases de dados utilizadas foram a Stanford 3D Scanning Repository [151] e uma base de dados de objetos adquiridos no próprio laboratório do autor. Os resultados obtidos pelos dois métodos foram comparados em termos das curvas de precisão e lembrança.

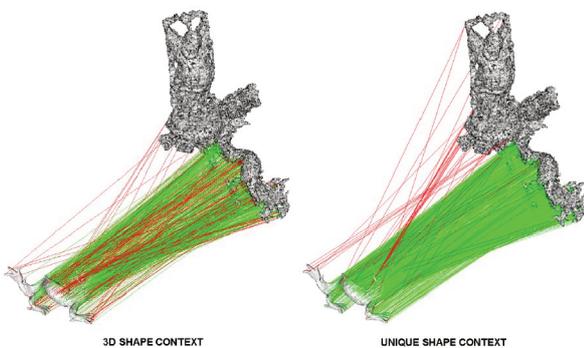


Fig. 48. Representação do método proposto por Tombari. Figura extraída de [149].

Um método invariante a rotação com base na configuração das características locais para a detecção, reconhecimento e classificação de objetos 3D é proposto por Knoop [152]. O método proposto é uma generalização do Implicit Shape Model (ISM) para o caso 3D, com um refinado esquema de votação com base na transformada de Hough. O processo de classificação utilizando transformada de Hough possui três partes principais: na primeira parte, um modelo de classe é aprendido do conjunto de formatos de treinamento. Na segunda parte, os modelos de classe aprendidos são utilizados para gerar hipóteses da provável classe de uma amostra pesquisada. Na última parte, procura-se no espaço de hipóteses, a hipótese com maior probabilidade de ser a classe correta. Os votos para cada hipótese são acumulados em um

espaço 4D específico. A classe com maior probabilidade e sua localização e escala deve ser obtida através da procura do máximo entre todas as classes. Como o esquema de votação apresentado utiliza um esquema de votação com direção fixa, o método torna-se sensível a orientação do objeto e a robustez da detecção da orientação das características. Logo, torna-se necessário analisar outros esquemas de votação (Point voting, circle voting e sphere voting). Para a avaliação dos métodos para classificação invariante a rotação utilizaram-se bases de dados com diferentes complexidades (Tosca [153] e SHREC'09 [154]), inicialmente comparando os diferentes esquemas de votação em termos de taxa de falso positivos e verdadeiros positivos e depois comparando com a utilização de outros descritores invariantes a rotação (Heat Kernel Signatures (HKS), Spin Images (SI), Shape Distribution (ShapeD2), HoughOct, 2.5D SURF, Spherical Harmonics (SH)), também utilizando as taxas de falsos positivos e verdadeiros positivos para a comparação.

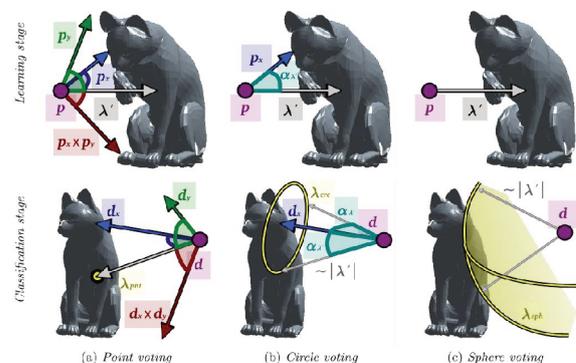


Fig. 49. Representação do método proposto por Knopp. Figura extraída de [152].

Seatovic [155] apresenta um sistema para o tratamento automático de plantas. O sistema combina um sensor de triangulação com laser infravermelho com uma câmera com alta resolução para gerar imagens 3D da erva em uma plantação. No processo de segmentação, remendos de superfícies contínuas são separadas umas das outras. Estes remendos de superfícies 3D são comparados com diferentes critérios na base de dados de plantas contendo parâmetros de superfície como formato, estado da superfície, entre outros. Se o objeto é reconhecido como um erva, as suas coordenadas são computadas e as folhas são pulverizadas com um herbicida. A análise da superfície no espaço pode impulsionar a performance da segmentação sob condições onde sistemas de reconhecimento 2D não obtém sucesso. O sistema como um todo esta descrito em [156], [157]. Cada nova medição do sistema gera um novo vetor de pontos 3D os quais são processados pelos seguintes passos: Filtragem para a remoção de ruídos, detecção de bordas e rotulamento de componente conexas 3D são os passos de pré-processamento executados. Depois, inicia o processo de reconhecimento, no qual todos os candidatos são projetados das superfícies 3D para o plano, informações 2D das folhas são produzidas. A transformação frequentemente projeta os pontos limite para o plano definido pelos próprios vetores computados a partir da nuvem de

pontos 3D. Cada objeto no *buffer* de reconhecimento contém também sua distribuição. Com a ajuda da computação de autovetores, valores de orientação, ponto médio, entre outros são computados e armazenados para cada objeto. Os dois maiores autovalores com seus correspondentes autovetores definem o plano de projeção para a análise do formato. Pontos limite do formato projetado são utilizados para computar descritores elípticos de Fourier, os quais são utilizados para a classificação do formato [158]. A classificação é realizada com a ajuda do classificador Support Vector Machine (SVM). Os resultados obtidos com o sistema proposto demonstram uma alta taxa de detecção sob condições simples, entretanto, sob condições complicadas (erva localizada sob constante oclusão) o sistema necessita de um melhora na taxa de detecção.

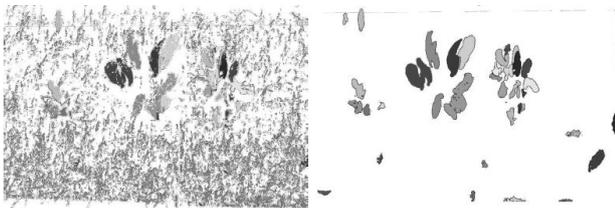


Fig. 50. Representação do método proposto por Seatovic. Figura extraída de [155].

Jidan Zhou[159] introduz um conjunto de características com base em formato denominado Histograms of Categorized Shapes (HCS), para o reconhecimento de objetos 3D. Mais especificamente, o conjunto apresentado é aplicado em um procedimento para detecção de orelha que usa uma janela deslizante, como caso de teste. Inicialmente, para localizar a orelha em uma imagem de profundidade, a imagem é escaneada do ponto superior esquerdo até o ponto inferior direito com um janela de detecção de tamanho fixo. Em cada posição, o vetor de característica HCS é extraído e empregado para treinar um classificador binário, no caso foi utilizado o Support Vector Machine (SVM). Para construir o descritor HCS, cada pixel na imagem é atribuído uma categoria de formato e uma magnitude com base em seu índice de formato e valor de curvatura. A janela de detecção da imagem a ser codificada é preenchida com blocos sobrepostos de vários tamanhos dos quais os HCSs são construídos pela agregação das respostas dos *pixels* dentro de blocos. Os histogramas em cada bloco são concatenados para formar um único histograma codificando a janela da imagem. Este histograma pode então ser utilizado como vetor de característica. O SVM é então utilizado para determinar se a janela atual contém ou não uma orelha pela avaliação do valor de classificação com base no vetor de característica extraído da janela. A janela de escaneamento com tamanho fixo permite apenas a detecção da orelha em uma única escala. Para alcançar detecção em múltiplas escalas, a imagem é gradativamente redimensionada, utilizando um fator de escala, e a janela de detecção é iterativamente aplicada em cada uma das imagens redimensionadas. Após o escanear a janela de detecção através da imagem em múltiplas escalas, múltiplas detecções geralmente ocorrem ao redor das regiões alvo e é útil fundir janelas de detecção sobrepostas em uma única detecção [160]. Seleciona-se um

método de non-maximal suppression (NMS) proposto em [161] como a solução para a fusão dessas regiões sobrepostas, nas quais cada detecção é mapeada para uma posição 3D respectiva e o espaço da escala é ponderado pelos valores de classificação. Um estimador não-paramétrico de densidade é empregado para estimar a função de densidade correspondente, onde os picos resultantes da função de densidade constituem as detecções finais, com posições, escalas e pontuações de classificação fornecidos pelas posições dos picos. Depois do NMS, o sistema de detecção retorna uma caixa delimitadora com uma pontuação de detecção associada representando a região da orelha. Como caso de teste, empregou-se o método de codificação de características 3D em um sistema de detecção de características. A base de dados utilizada foi um subconjunto da base de dados da universidade de Notre Dame (UND) [112] e os resultados foram demonstrados em termos de taxa de detecção por número de falsos positivos por imagem e comparados com outros métodos([113], [112], [162]).



Fig. 51. Representação do método proposto por Jidan Zhou. Figura extraída de [159].

Wohlkinger [163] propõe explorar o conhecimento contextual através do uso de restrições de sensores e hardwares dos domínios da robótica e da casa e mostrar como utilizar a internet como fonte para obter os dados requisitados para construir um rápido sistema de categorização de objetos, com base em visão, para a robótica. No trabalho em questão, uma visão geral das restrições disponíveis e das vantagens de se utilizar um robô para setar prioridades para a classificação de objetos é fornecida, e propõe-se um sistema o qual cobre a aquisição automática da internet, simulação do domínio, geração de descritores, processamento de dados 3D de imageamento estéreo denso e classificação para um cenário de robô em um ambiente doméstico conectado à internet. Para tanto, inicia-se com o requerimento de inventar um *framework* que pode facilmente ser estendido para uma nova classe de objetos. Para este fim, utiliza-se a internet como fonte para obter os modelos para novos objetos. Alternativamente ao trabalho apresentado em [164], propõe-se usar apenas dados 3D perfeitos e transformá-lo em um dado simulado de sensor para lidar com típicos problemas de aplicações reais tais como, apenas uma visão do objeto (2.5D) com auto-occlusão, modelos incompletos, efeitos de *aliasing* e níveis realísticos de ruído de dados estéreo. Então, utiliza-se este dado para calcular classificadores de objetos estendendo o descritor 3D Harmonics [165] com as restrições do domínio robótico e correspondendo-os contra a base de dados para encontrar a classe mais próxima do objeto. Para a parte aquisição dos modelos com base na internet utiliza-se a mesma abordagem apresentada pelo autor em [166]. Para a parte de particionamento semântico dos dados, utilizam-se os seguintes passos: filtragem, removendo *outliers* e pontos esparsos; Detecção do plano suporte, por meio do RANSAC; Clusterização, utilizando um método de clusterização com base no *flood*

filling nos pontos remanescentes acima do plano chão; Filtro de região prioritária, rejeitando todos os *clusters* fora de uma região de suporte; Filtro de candidatos suportados, ao fazer a suposição que o objeto deve estar em um plano de suporte ela também implica que os objetos candidatos estão no plano de suporte, desta maneira apenas são processados *clusters* de pontos que estão ligados aos seus planos suporte. Os objetos candidatos são submetidos ao processo de classificação. O objetivo do processo de classificação é atrelar o rótulo correto para um dado *cluster* de dados. Isto pode ser visto também como encontrar o objeto mais similar ao dado consultado e atribuir o rótulo da correspondência mais similar. Esta etapa é realizada com o uso do descritor Spherical Harmonics, representado e armazenado como um histograma 2D na base de dados, permitindo o cálculo do k-nearest neighbours utilizando a distância euclidiana entre os dados consultados e as entradas na base de dados. Os testes foram realizados em um conjunto de dados coletados localmente e os resultados de classificação foram demonstrados em termos de curvas de precisão e lembrança.

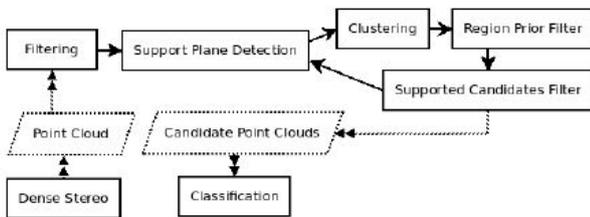


Fig. 52. Representação do método proposto por Wohlkinger. Figura extraída de [163].

Tombari [167] propõe um método para detecção de formas livres no espaço 3D, para utilização na tarefa de reconhecimento de objetos em cenas 3D com um significativo grau de oclusão e desordem. O método apresentado utiliza a detecção e descrição de características 3D para computar um conjunto de correspondências entre o modelo 3D e a cena atual. Em adição, cada ponto de característica está associado com sua posição relativa com relação ao centroide do modelo, de tal forma que cada característica correspondente da cena pode gerar um voto em um espaço 3D Hough. Para votar corretamente de acordo com a atual pose do objeto procurado, conta-se com o Local Reference Frame (LRF) associado com cada par de características correspondidas. O sistema como um todo pode ser dividido em duas etapas. A primeira etapa é uma etapa offline, na qual ocorre a detecção e descrição dos modelos bem como a inicialização do acumulador do Hough. A segunda etapa é a etapa online, onde analisa-se a cena detectando e descrevendo as características da cena para a posterior correspondência das características entre cena e modelo seguida da etapa de voto no espaço 3D Hough. O uso do esquema de voto visa acumular evidências sobre a presença do objeto sendo buscado. Se existe uma quantidade suficiente de votos de características para a presença de um objeto em uma determinada posição, então o objeto é detectado e a sua pose determinada por meio das correspondências computadas. Os experimentos foram realizados em duas bases de dados diferente (uma base de dados 3D pública construída por Milan

[138] e outra adquirida localmente pelos autores por uma técnica Spacetime estéreo e que consiste de 6 modelos e 12 cenas, para um total de 72 instâncias de reconhecimento de objetos) e os resultados avaliados em termos de curvas de precisão e lembrança.

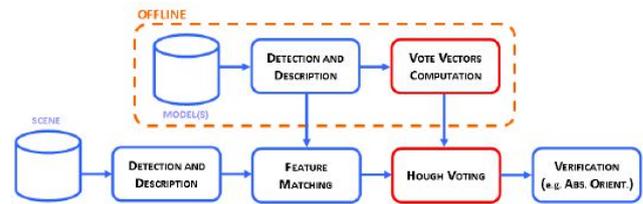


Fig. 53. Representação do método proposto por Tombari. Figura extraída de [167].

Um *pipeline* para o reconhecimento de objetos 3D com base em aprendizagem é descrito por Owechko [168]. O primeiro passo é um processo de sugestão espacial consistindo de uma segmentação da nuvem de pontos em objetos potenciais ou sugestões. Sugestões relacionadas são então fundidas e a segmentação resultante é utilizada para selecionar um classificador 3D com base nas propriedades espaciais das sugestões. Então, regiões de interesse centradas nas segmentações são utilizadas para gerar um conjunto de características 3D que capturam propriedades geométricas e topológicas dos grupos de pontos dentro da região de interesse de uma maneira invariante ao eixo-Z. Classificadores estatísticos com base em árvores de decisão são treinados com essas características. O resultado é um conjunto de regiões de objeto segmentadas e rotuladas semanticamente. No intuito de implementar o processamento espacial de sugestões, a cena 3D é representada como um grid de tiras 3D verticais, onde cada uma das quais está verticalmente dividida em células e é representado por uma distribuição de densidade populacional através das células. As estatísticas globais das tiras são calculadas usando a distribuição de densidade vertical. A segmentação com base na representação anteriormente descrita consiste em agrupar tiras que possuem atributos similares. Uma abordagem com base em grafos é utilizada para segmentar estas partições de nodos em componentes de modo que cada região corresponde idealmente a um objeto 3D no mundo. A performance de reconhecimento do *pipeline* proposto para 17 cenas urbanas 3D é apresentada em termos de probabilidade de identificação.

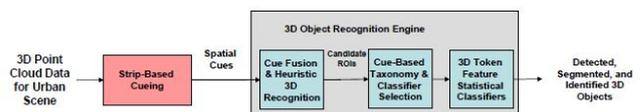


Fig. 54. Representação do método proposto por Owechko. Figura extraída de [168].

Um método para o reconhecimento de objetos 3D com base na comparação de suas invariantes de projeções de curvas 2D é apresentado por Unel[169]. O método proposto inicia com uma imagem de profundidade ou uma representação tessellada de um objeto. Em ambos os casos, computa-se a orientação do objeto analisado. Quando a imagem de entrada

é a de profundidade, a abordagem ajusta uma superfície algébrica ao objeto. Os autovetores da matriz de momento de segunda ordem dos dados da superfície são então computados. Estes autovetores implicam em três direções ortogonais no espaço ao longo do qual os dados de superfície se espalham. Cruzamentos da superfície algébrica com planos ortogonais ao seu eixos principais produz curvas de projeções no plano de coordenadas. Quando a entrada é a imagem tessellada, computa-se a orientação do objeto analisado empregando um casco quase convexo do limite do objeto. O casco quase convexo induz uma orientação do objeto de entrada em termos de eixo de inercia, como definido pelo momento de inercia do tensor. Cruzamentos são obtidos através da intersecção dos planos perpendiculares ao eixo de inercia e o modelo tessellado. Dado que cruzamentos são curvas planares, tratam-se eles como curvas de projeção, i.e., considera-se reorientar os objetos de modo que os eixos de inercia coincidem com os eixos de coordenada. Mostra-se que essas curvas de projeção são *affine equivalent* e propõe-se dois métodos para construção de invariantes algébricas e geométricas das curvas de projeção. Para o reconhecimento dos objetos 3D, emprega-se uma medida média de similaridade dos vetores invariantes de um certo número de curvas de projeção. Adicionalmente, emprega-se uma medida de distância entre o conjunto de invariantes. Os experimentos foram realizados com base em objetos definidos por nuvem de pontos e modelos tessellados disponíveis nas bases de dados de Princeton [39] e McGill [170]. Os resultados desse experimentos foram demonstrados em termos de matrizes de confusão, similaridade média e tempo de computação.

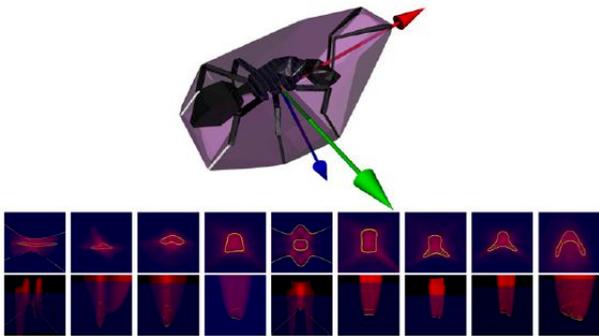


Fig. 55. Representação do método proposto por Unel. Figura extraída de [169].

Westell [171] apresenta um sistema de reconhecimento e localização de objetos dentro de ambientes fechados e desconhecidos. No sistema, o usuário pode descrever o objeto alvo por meio de descrições de formato, tamanho ou cor e o sistema utiliza estas descrições para selecionar automaticamente o objeto alvo de uma base de dados de objetos. Uma vez que o objeto alvo foi selecionado, a inspeção de múltiplas visões pode ser conduzida afim de identificar e localizar o objeto dentro do ambiente. Ao entrar em um ambiente e capturar várias imagens da cena de diferentes pontos de vista, um algoritmo de reconhecimento de objetos é utilizado para localizar possíveis correspondências e um dispositivo de imagem estereo é usado para resolver as coordenadas 3D dessas correspondências,

de modo que o robô possa subsequentemente inspecionar cada região de três pontos de vista distintos para identificar correspondências positivas dentro do ambiente. O algoritmo de reconhecimento possui dois estágios extração dos dados de treinamento e identificação do objeto. No estágio de extração de dados, a composição de cor das imagens da base de dados são utilizadas para construir uma representação única de cada objeto. Para limitar os efeitos da luminosidade utiliza-se apenas a matiz e a saturação do espaço HSV, assim como aplica-se o algoritmo *white patch retinex* [172] para reduzir os efeitos de mudança da iluminação. Em cada uma das imagens armazenadas na base de dados extraem-se remendos quadrados contendo apenas o objeto, sem o plano de fundo, e para cada remendo extraído monta-se um histograma 2D normalizado referente a matiz-saturação de cada remendo. Realizando a extração para cada uma das 36 imagens de um objeto gera-se uma quantidade grande de vetores de *bins* do histograma. Para obter uma representação compacta, esses vetores são *clusterizados* para 100 histogramas representativos utilizando o k-means. A matriz resultante fornece uma representação única de cada objeto e pode ser utilizada para localizar instâncias na base de dados. Para identificar um objeto dentro de uma imagem de cena, o estágio de identificação do objeto produz um mapa de confiança com base nas informações de correspondência da base de dados para a cena. Assim como na etapa offline de treinamento, os remendos da imagem são extraídos de uma variedade de escalas e posições dentro da imagem. Para cada remendo de imagem extraído, monta-se o histograma 2D referente a matiz e saturação e este é comparado com cada um dos 100 histogramas representando o objeto alvo. Após comparar os histogramas entre a cena e o objeto alvo gera-se uma pontuação de intersecção, a qual se for maior que 75% as localizações do remendo de imagem correspondente é incrementado em um mapa de confiança. Após esta análise, as regiões são identificadas na máscara de confiança onde existem possíveis correspondências. Estas regiões são analisadas com base em duas métricas, confiança média e taxa de correspondência, e estas duas pontuações fornecem uma probabilidade significativa que o objeto alvo esta presente ou não na região. Para testar a acurácia do algoritmo de reconhecimento, uma verificação cruzada dentro da base de dados de imagens foi realizada e os resultados foram demonstrados em termos de curvas Receiver Operating Characteristics (ROC). Para testar os resultados em ambientes reais, vários experimentos foram conduzidos nos quais o sistema inteiro foi examinado demonstrando os resultados com base nas pontuações de confiança média e taxa de correspondência.

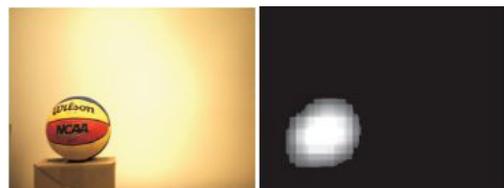


Fig. 56. Representação do método proposto por Westell. Figura extraída de [171].

Ayoub [173] propõe um método para o reconhecimento de objetos 3D capturados por um sensor ativo de visão estéreo. A aplicação médica, na qual o sistema é utilizado requer a diferenciação entre dois tipos de pólipos durante o imageamento 3D de uma endoscopia. O sistema de reconhecimento e classificação proposto é dividido em três partes: aquisição dos dados, extração de características e classificação. A parte de aquisição de dados é realizado por meio de um sistema de visão estéreo, como sensor de captura, e da aplicação de técnicas de processamento com a finalidade de melhorar a qualidade e preparar para o estágio de reconstrução. No final desta parte obtém-se uma nuvem de pontos 3D. Na parte de extração de características, dado a nuvem de pontos adquirida dos objetos de estudo na etapa anterior, extraí-se características as quais serão utilizadas em um classificador Support Vector Machine. O classificador SVM foi avaliado com dois diferentes conjuntos de características. O primeiro é o tradicional conjunto características constituído dos pontos da nuvem 3D da superfície sendo estudada. O segundo é um conjunto de características adequadas com base em histograma de profundidade. O sistema de reconhecimento como um todo envolve algumas etapas a serem realizadas previamente durante uma análise offline. Estas atividades incluem calibração do sistema estéreo, processo de seleção de características para redução da dimensionalidade do espaço de características, treinamento do SVM, seleção do modelo e validação cruzada para encontrar os melhores parâmetros do classificador. Os experimentos realizados tem como base a comparação dos dois métodos de extração de características apresentados em termos de acurácia de classificação.

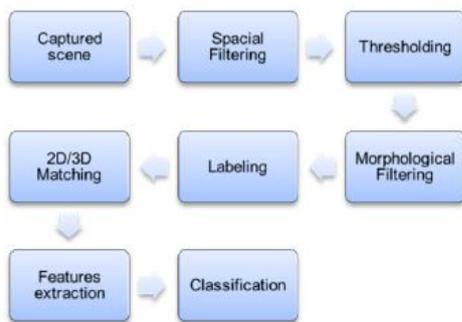


Fig. 57. Representação do método proposto por Ayoub. Figura extraída de [173].

Um método para o reconhecimento de objetos em alta velocidade, utilizando um método de subespaço linear em um espaço de características 3D com base em características Color Cubic Higher-order Local Auto-Correlation (Color-CHLAC)[174], é proposto por Kanezaki [175]. As características Color-CHLAC são calculadas utilizando informações de formato e cor dos dados de voxel 3D. O sistema de reconhecimento proposto com base no método de subespaço linear [176], pode ser descrito pelos seguintes passos: Como passo de pré-processamento, calcula-se os vetores de características Color-CHLAC para as partes subdivididas de cada modelo na base de dados. Então, utilizam-se esses vetores de características para computar a bases para um subespaço

definido por cada objeto. Na etapa de reconhecimento, um vetor de características é extraído de uma parte da cena para consulta. Então, esta parte consultada é correspondida contra os objetos na base de dados através da projeção deste vetor de características em cada subespaço dos objetos e do cálculo de uma pontuação de similaridade, a qual é utilizada para o ranqueamento dos objetos da base de dados e eleição do objeto candidato. Para os experimentos testou-se a performance da abordagem proposta no reconhecimento de objetos 3D em cenas reais, comparando-se os resultados obtidos com abordagens convencionais utilizando Spin-Image (SI)[16] e Textured Spin-Image (TSI) [177] em termos da média de taxa correta e do tempo de computação.

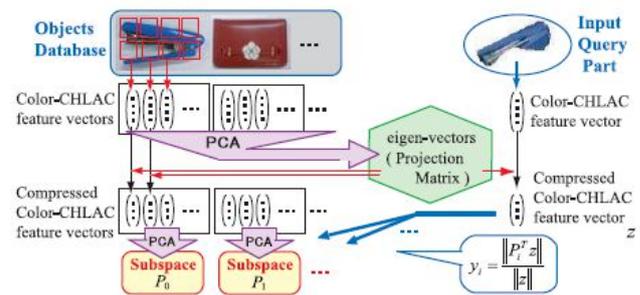


Fig. 58. Representação do método proposto por Kanezaki. Figura extraída de [175].

Usui [178] propõe um método de correspondência com base em confiança para o reconhecimento de objetos 3D. O *framework* proposto tem como base sistemas de reconhecimento de objetos 3D com base em ponto, no qual características são extraídas de uma imagem de entrada e então correspondidas com um descritor do modelo. Para a parte de extração de características utiliza-se o Affine SIFT [179]. Depois, aplica-se o Best Bin First (BFF) proposto por Lowe's [47], o qual é um método de busca que compara a distância de um ponto até seu vizinho mais próximo com a sua distância ao segundo vizinho mais próximo. O BFF é utilizado para procurar o descritor mais próximo. O processo de decisão é realizado para todos os pontos de característica, após o qual o número total de pontos correspondentes é comparado com o valor de limiar. Se o número de pontos correspondentes for maior que um limiar, a imagem de entrada é classificada como correspondida. A parte de decisão apresentada no algoritmo de Lowe's é substituída por uma look up table (LUT) com base em valores de confiança, onde a distância euclidiana entre a característica consultada e o seu vizinho mais próximo, e o valor de confiança estão relacionados. A LUT é treinada utilizando um conjunto de dados de treinamento rotulados. Para avaliar o método proposto alguns experimentos foram realizados utilizando imagens de objetos de comuns em uma casa. Os resultados dos experimentos foram demonstrados em termos de curvas Receiver Operating Characteristics (ROC).

Rodner [180] apresenta um abordagem para o reconhecimento de objetos genéricos com informações de profundidade, obtidas por uma câmera Time-of-Flight, e imagens coloridas, obtidas com um sensor visual. Para combinar a informação dos dois sensores, utiliza-se um Gaussian Process (GP) de

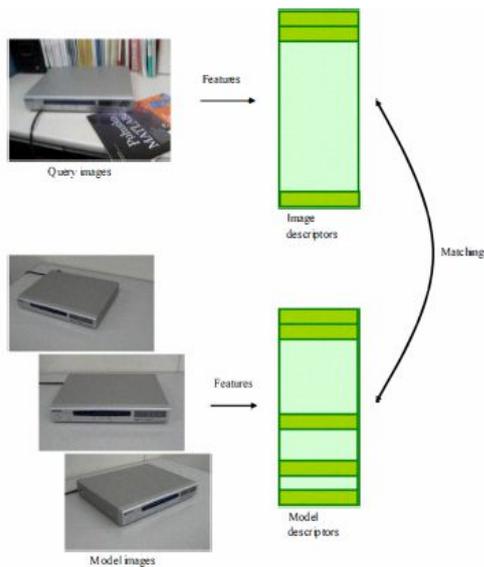


Fig. 59. Representação do método proposto por Usui. Figura extraída de [178].

classificação [181], o qual permite uma estimativa de hiper-parâmetro e integração da informação de múltiplos sensores por meio de uma combinação de kernel em um *framework* Bayesiano. A estimativa do hiper-parâmetro é realizada pela extensão de técnicas de múltiplas tarefas para regressão GP para aproximação Laplaciana. A computação das funções *kernel* com base em imagem, utilizando imagens de profundidade, são realizadas através da aplicação do *framework* Spatial Pyramid Matching Kernels [182] (SPMK) para os diferentes tipos de características locais de profundidade (profundidade do pixel, normais da superfície e curvatura) utilizados. O que resulta em diferentes *kernel* para cada tipo de característica local, os quais podem ser combinados com o GP de seleção de modelo. Para as características de cor utilizou-se as características do Opponent-SIFT [183]. O processo de classificação em múltiplas classes segue a técnica um contra todos [184]. Ou seja, para cada classe um classificador binário é treinado, utilizando todas as imagens de classe i como exemplos positivos e as imagens remanescentes como exemplos negativos. A classificação é realizada pelo retorno da classe com maior probabilidade estimada pelo classificador binário correspondente. A abordagem um contra todos, também possibilita a realização da seleção do modelo pela otimização do conjunto de hiper-parâmetros para todos os problemas binários envolvidos. A função objetivo é a simples soma de todos os \log de probabilidade binária negativa para a regressão GP. Para os experimentos utilizou-se a base de dados em [185], na qual os objetos pertencem a 5 diferentes categorias e cada categoria contém 7 objetos com 32 pares de imagens. Inicialmente, comparou-se a performance para todas as características de profundidade apresentadas, utilizando para a classificação a regressão GP. Depois, comparou-se a performance da regressão GP com a classificação GP com aproximação Laplaciana. Por fim comparou-se a abordagem com base no GP com o trabalhos anteriores [185], [186]. Os

resultados das comparações foram demonstrados em termos de taxa de reconhecimento.



Fig. 60. Representação do método proposto por Rodner. Figura extraída de [180].

Um algoritmo para o reconhecimento de objetos 3D na presença de desordem e oclusões em dados esparsos, não segmentados e com ruído é apresentado por Papazov [187]. O método é composto por duas fases, pré-processamento do modelo e reconhecimento, e utiliza um descritor geométrico, uma técnica de *hashing* e uma estratégia de amostragem similar ao RANSAC. Na fase de pré-processamento do modelo, executada offline, cada modelo de objeto passa por um processo de amostragem dos pares de pontos orientados que respeitam um valor de tolerância. Para cada par de ponto, um descritor é computado, da mesma forma a realizada em [188], e armazenado em uma tabela de espalhamento. A fase de reconhecimento utiliza a tabela de espalhamento, previamente computada, e pode ser descrita nos seguintes passos: Computação de uma octree para a cena, com a finalidade de produzir uma cena modificada, e cálculo do número de iterações necessárias para alcançar uma probabilidade de reconhecimento bem-sucedido maior que um valor previamente definido. Em cada iteração, realiza-se uma amostragem dos pontos da cena, estima-se a normais dos pontos para obter pares de pontos orientados na cena, computa-se os descritores da mesma forma realizada para os modelos, e utiliza-se os descritores calculados como chave para recuperar, da tabela de espalhamento dos modelos, os pares de pontos orientados dos modelos similares aos pontos de pares da cena, computando a transformação, para alinhar o modelo e a cena, e armazenando o par modelo-transformação se o par for aceito por uma função de aceitação. Como ultimo passo filtra-se da lista de solução hipóteses conflitantes. Para os testes utilizou-se a mesma base de dados empregada em [81] e compararam-se os resultados com outros dois métodos, Spin images [16] e tensor matching [81], em termos de taxa de reconhecimento em relação a porcentagem de oclusão. Também verificou-se o tempo de execução e o comportamento em relação a presença de ruído. O método proposto por esse trabalho serviu como parte central de outro trabalho desenvolvido pelo autor, no qual utiliza-se o método proposto atrelado a um sistema de controle de braço robótico para pegar objetos em cenas com desordem [189].

Akagunduz [190] propõe um método de reconhecimento de objetos 3D invariante a transformações. As superfícies 3D são representadas por estruturas de superfície 3D denominadas características multi-escala. Estas estruturas da superfície são extraídas invariantes a suas métricas de tamanho e amostragem das imagens de profundidade. Características multi-escala



Fig. 61. Representação do método proposto por Papazov. Figura extraída de [187].

são extraídas com suas escalas utilizando o espaço-escala da curvatura das superfícies 3D. Diferente de abordagens anteriores de espaço-escala, as características são extraídas como volumes conectados dentro da curvatura escala-espaço classificada. Triplas dessas características multi-escala são consideradas para representar a topologia das superfícies invariantes a transformação e elas são utilizadas em um *framework* de hashing geométrico [188] para o reconhecimento de objetos. Para os testes com o método proposto utilizou-se a base de dados Stuttgart Range Image [68]. Para observar os benefícios da utilização de características multi-escala, em oposição a extração de características em uma única escala, as características 3D foram extraídas com procura de múltiplas escalas (MSFE) e sem procura em múltiplas escalas (SSFE) e vetores de características de topologia foram utilizados no hashing. Para ver os benefícios de utilizar descritores de pontos de característica, spin images ao redor dos centros de característica também foram extraídos e anexados ao vetor de característica como realizado em [188] para ambos SSFE e MSFE. Por fim, características SIFT são utilizadas como realizados em [47]. Os resultados obtidos foram demonstrados em termos de matriz de confusão e taxa de reconhecimento para várias configuração de características e descritores.

Tanguamsub [191] propõe uma técnica de reconhecimento de objetos 3D em ambientes reais. O método proposto possui dois estágios, treinamento e detecção. Na parte de treinamento, uma caixa *template* é utilizada para investigar a área do objeto em um conjunto de imagens de treinamento. Nas imagens de treinamento, pontos chave do objeto são detectados e suas características são extraídas. Dois tipos de informações são extraídas, informações espaciais e informações de aparência. Informações de aparência incluem características locais tais como SIFT e SURF. Informações espaciais são utilizadas para representar a relação entre cada ponto chave do objeto e o centro do objeto. Esta relação é representada por um vetor de localização o qual é formado por 4 elementos: escala do ponto chave, orientação do ponto chave, rótulo da pose do objeto e vetor de distância do centro do objeto até um ponto chave. Para representar a categoria do objeto é utilizado um modelo de classes. Este modelo é composto de um conjunto de modelos de *templates*, onde cada um deles é utilizado para descrever uma instância do objeto. O modelo de *template* inclui as informações dos pontos localizados dentro da caixa de *template*, informações espaciais e de aparência de todos os pontos chave. Na parte de reconhecimento e detecção utiliza-se uma técnica de votação para reconhecer um objeto em uma imagem. A posição do centro do objeto e seus limites são

estimados a partir do resultado da votação. Visando reduzir os erros provenientes de falsas detecções, atribuiu-se um limiar de votação e utiliza-se a similaridade da aparência para rejeitar as hipóteses de candidatos com baixa confiança. O método proposto foi comparado contra outros métodos (Liebelt [192], su & sun [193], sun & su [194], INRIADouze [195], EN-SMP [196], Cambridge [197] e TKK [198]) utilizando dois conjuntos de treinamento e teste, os quais tem como base as imagens da base de dados PASCAL Challenge Visual Object Class Dataset 2006 [199]. A comparação com outros métodos se deu em termos dos resultados das curvas de precisão e lembrança.

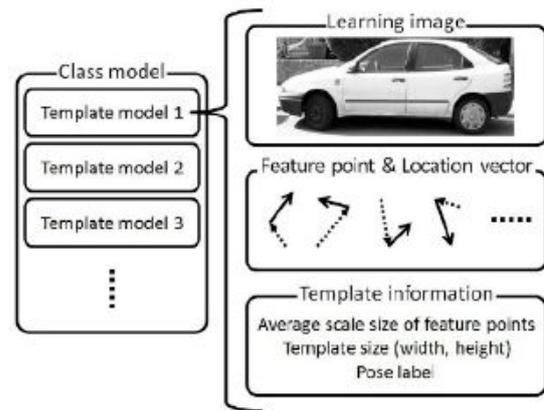


Fig. 62. Representação do método proposto por Tanguamsub. Figura extraída de [191].

Eunyoung Kim [200] apresenta uma abordagem para o reconhecimento de objetos que impulsiona a dissimilaridade entre objetos consultados e objetos com formato similar no plano de fundo na cena através da maximização do uso do contexto de visibilidade. Uma característica de pares de ponto foi projetada, contendo uma descrição discriminativa inferida do contexto de visibilidade. E, foi proposto um método de estimativa de pose que localiza os objetos utilizando a correspondência desses pares de pontos. Por fim, duas medidas de validade são sugeridas para descartar falsas detecções. O passo inicial da abordagem é estimativa da normal da superfície e a amostragem de pontos com base em múltiplas resoluções de malhas de pontos, obtendo assim uma malha simplificada. Depois, com base na imagem de profundidade original, computa-se 3 tipos de visibilidade, *Visible space* S_V , *Surface* S_S e *Invisible space* S_o , com base no valor de profundidade de cada pixel na imagem e sua correspondente posição 3D na coordenada de mundo. Esses contextos de visibilidade são utilizados para aumentar o poder discriminativo na descrição da característica e descartar poses inválidas. Então, calcula-se a característica de pares de pontos com base em dois pontos 3D orientados, de forma que cada valor de cada coordenada é mapeada em um valor inteiro e estes valores são utilizados como chaves para encontrar um conjunto de pares de pontos correspondentes extraídos dos modelos. Para cada correspondência, a pose em potencial do modelo correspondido na cena é computada e votada para o reconhecimento final. Com a finalidade de reduzir a complexidade

computacional do cálculo de tal característica, propõe-se uma nova características de pares de pontos que contém descrições inferidas dos contextos de visibilidade anteriormente descritos. O contexto de visibilidade impulsiona o poder discriminativo de cada característica de pares de pontos, através da imposição implícita de características globais de um modelo consultado, e reduz o número de correspondências espúrias. Então utiliza-se as características dos pares de pontos calculadas, para reconhecer e localizar os objetos consultados através do acúmulo de todas as possíveis poses do objeto inferido da correspondência dos pares de pontos com base em um processo de votação. O processo de votação é utilizado para validar a pose com base no contexto de visibilidade. Após a votação duas métricas, alinhamento da superfície e separabilidade da superfície, são utilizadas para verificar as poses selecionadas em termos de possíveis falsos-positivos que passaram pelo processo anterior. As poses que obtiverem o maior suporte são identificadas como corretas. Os experimentos realizados com a abordagem proposta utilizaram 10 objetos diferentes capturados com um sensor de profundidade. Para a verificação dos resultados, *Ground truth* foram marcados manualmente nas cenas e para a análise comparativa utilizou-se o método apresentado em [136] e o Spin images [16]. Os resultados dos experimentos foram demonstrados em termos de curvas de precisão e lembrança.

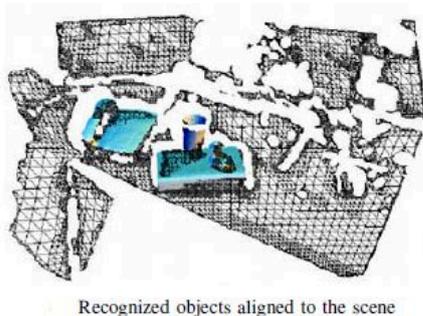
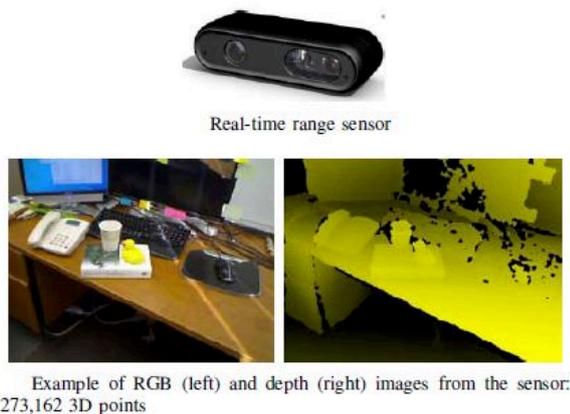


Fig. 63. Representação do método proposto por Eunyoung Kim. Figura extraída de [200].

Um descritor de formato, com base em funções de formato, para nuvens de pontos parciais é apresentado por Wohlking [166]. Este descritor, denominado ESF (Ensemble of Shape Functions), possui a capacidade de treinamento em dados sintéticos e classificação de objetos provenientes de um sensor

de profundidade em uma única visão parcial, de uma forma rápida e robusta. A tarefa de classificação é declarada como uma tarefa de recuperação 3D, localizando os vizinhos mais próximos de visões geradas sinteticamente de modelos CAD para a nuvem de pontos detectada com um sensor de profundidade de estilo Kinect. O descritor ESF é um conjunto de dez histogramas, com tamanho de bin 64, de funções de formato descrevendo propriedades características do *cluster* de nuvem de pontos. Estes dez histogramas são formados pelas seguintes funções de formato: D2, utilizada como apresentado pelo autor em [201]; A3, uma função de formato que codifica o ângulo delimitado por duas linhas criadas por amostragem aleatória de três pontos da nuvem de pontos e que, assim como a D2, gera três histogramas distintos; D3, função de formato a qual tem como base a D2 com a adição de mais uma dimensão, gerando assim outros 3 histogramas; Taxa de distância das linhas é o último histograma gerado, compondo assim os 10 histogramas do ESF. A correspondência do ESF dos modelos com os ESF da nuvem de pontos são realizados por meio da L1-distance. A base de dados utilizada para os experimentos, foi criada de forma semi-automática através do download de modelos 3D principalmente da *Google Warehouse* [202] e do armazenamento dos modelos de forma hierárquica similar ao WordNet[203]. Esta base de dados criada, que em outro trabalho do autor serviu como alicerce para o reconhecimento de classes de objetos em larga escala partindo de modelos CAD [204], é formada por modelos CAD e o treinamento com base nestes modelos é realizado através da renderização e amostragem do z-buffer a partir de visões ao redor de cada modelo e do armazenamento destas visões parciais como nuvem de pontos. Os resultados dos experimentos foram demonstrados em termos de resultados de classificação e comparados com outros descritores (Shape Distribution on Voxel Surfaces[201], Viewpoint Feature Histogram [145] , Clustered Viewpoint Feature Histogram [205] e SHOT [150].

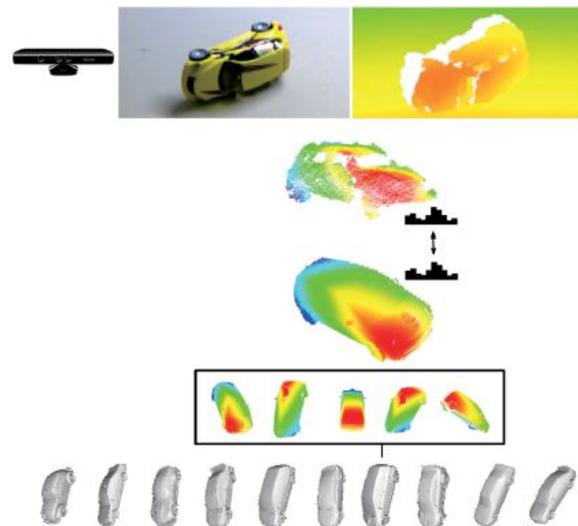


Fig. 64. Representação do método proposto por Wohlking. Figura extraída de [166].

Zarpalas[206] apresenta um descritor para o reconhecimento de objetos cenas 2.5D, produzidas escaneamentos de pro-

fundidade, com oclusão e desordem. O descritor compacto de formato regional proposto, denominado projection images, foi elaborado para ser robusto contra ruído, oclusão parcial e desordem. O projection images é formado por projeções dos pontos sobre o plano centrado no ponto base, o qual é perpendicular ao eixo de visualização. Para o processo de reconhecimento, o projection images dos objetos conhecidos são extraídos e armazenados em uma base de dados. Dado um escaneamento de uma cena, Projection images são extraídos e comparados com os que estão armazenados na base de dados, encontrando uma correspondência entre os pontos. Depois, realiza-se um processo para verificar quando um objeto é identificado na cena: Primeiro, uma abordagem de clusterização é aplicada na nuvem de pontos da cena, explorando o fato que os pontos que representam um objeto estão próximos uns dos outros em termos de distância euclidiana. Neste escopo, começando de um ponto randômico, seus vizinhos que estão dentro de uma distância μ são computados. Então, para cada um destes vizinhos, seus correspondentes vizinhos são computados e assim por diante. Deste modo o conjunto de pontos expande-se até não existirem mais pontos dentro da distância μ de um ponto do conjunto. Desta forma, clusters são criados que realizam uma conectividade μ . O cluster com maior taxa de correspondência, isto é, a taxa do número de pontos correspondentes dentro de um cluster em relação ao população do cluster, é considerado a melhor correspondência. Com o intuito de comparar a performance do método proposto contra outros métodos utilizou-se a base de dados disponível em [207]. Os outros métodos utilizados no estudo comparativo foram o spin images [16] e o distance maps [208] em termos de taxa de reconhecimento em relação a porcentagem de oclusão.

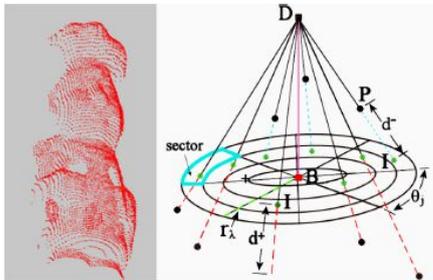


Fig. 65. Representação do método proposto por Zarpalas. Figura extraída de [206].

O trabalho apresentado por Wohlkinger [201] apresenta uma adaptação do descritor 3D D2 shape distribution descriptor [209], originalmente definido unicamente para correspondência de modelos 3D. A adaptação apresentada do descritor de formatos 3D para dados 2.5D permite o cálculo de características em tempo real diretamente dos pontos 3D. Além disso, mostra-se como tal descritor pode ser utilizado em um *framework*, o qual utiliza uma abordagem semi-automática para adquirir da internet os dados requeridos para treinamento na forma de modelos 3D, o que reduz significativamente o tempo para aprender novas categorias. A entrada para o sistema de aquisição de modelos é o nome da classe do objeto. Com essa palavra chave, inquirir-se a base de dados léxica WordNet[203] para retirar a ambiguidade da palavra-

chave através da apresentação dos diferentes significados ao usuário para o mesmo selecionar o mais apropriado. Uma vez que o significado da palavra é conhecido, os sinônimos e hipônimos fornecidos pelo WordNet e as traduções para as linguagens mais comuns são utilizadas para a procura de modelos 2D na 3D *Google Warehouse* [202] para obter-se o maior número de modelos. Para utilizar os modelos adquiridos na internet para a correspondência de imagem de profundidade, imagens sintéticas de profundidade são geradas através da renderização dos modelos 3D e da amostragem do *z-buffer* de 80 visões igualmente espaçadas ao redor do modelo. Para uma destas visões do modelo, o D2 shape distribution descriptor é calculado e armazenado na base de dados. O modelo mais similar e as visões podem ser encontradas pela comparação dos descritores calculados, a partir da imagem de profundidade fornecida pelo sensor, com todos os descritores armazenados na base de dados. A adaptação apresentada classifica cada linha, criada por dois pontos amostrados randomicamente, como fora ou dentro da superfície de nuvem de pontos. Esta etapa é realizada pelo tracejamento da linha com a ajuda do algoritmo 3D Bresenham em um *grid* de *voxels*. Este *grid* de *voxels* grosseiro serve como uma aproximação da superfície real, mas pode ser criada de uma forma computacionalmente eficiente. Após o tracejamento, as classes de linha são colocadas em 3 classes de histogramas diferentes, representado ON, OFF e distâncias misturadas. O descritor final é juntado por um histograma grosseiro com 32 *bins* e um histograma refinado com 128 *bins* do D2 shape distribution. Isto é estendido para 3 histogramas de 64 *bins* representando as distâncias classificadas na superfície do *voxel* e finalmente por um histograma de 64 *bin* da proporção das linhas das distância misturadas, o qual fornece um poder descritivo adicional. A correspondência desses histogramas é realizada por *L1-distance*, o qual provou ser capaz de lidar com o ruído do sensor sem sacrificar a performance de classificação. Para a parte de correspondência utilizou-se uma ideia de *hashing* para evitar problemas com uma grande base de dados. A base de dados utilizada para os testes consiste em 163 categorias com 2460 modelos 3D e 196800 visões, as quais foram organizadas em uma tabela *hash* e os resultados foram demonstrados em termos classificação por classe de objeto e taxa de classificação.

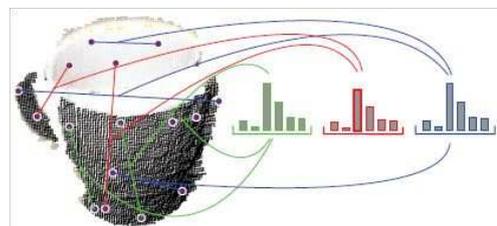


Fig. 66. Representação do método proposto por Wohlkinger. Figura extraída de [201].

Jindan Zhou [210] apresenta um sistema completo de reconhecimento de objetos 3D combinando características locais e holísticas de uma forma computacionalmente eficiente. Este sistema foi avaliado em uma tarefa de classificação

de orelhas. O sistema é composto de quatro componentes primários, nomeados, segmentação 3D da orelha, extração das características locais e correspondência, extração de características holísticas e correspondência e um *framework* de fusão combinando as características locais e holísticas a nível pontuação de correspondência. Para o componente de segmentação empregou-se o método apresentado em [159] para segmentação de orelhas 3D. Para o componente de extração de características locais e correspondência, foi introduzido o descritor de características Histogram of Indexed Shapes (HIS) e estendido para um descritor de formato 3D centrado no objeto denominado Surface Patch Histogram of Indexed Shapes (SPHIS), para representação de segmentos da superfície e correspondência. Para o componente de extração e correspondência de características holísticas, foi proposto a "voxelização" [211] da superfície da orelha para gerar uma representação, da qual uma eficiente comparação *voxel-wise*, do pares de modelos da galeria e de teste, pode ser realizada. As pontuações de correspondência obtidas das correspondências entre características holísticas e locais, dos modelos da galeria e de teste, são fundidos para gerar a pontuação de correspondência final, por meio de uma técnica de soma ponderada [212]. Os experimentos realizados utilizando a base de dados da Universidade de Notre Dame (UND) coleção G [112], foram comparados com outros métodos ([112], [113], [213] em termos de taxa de reconhecimento e os demais experimentos realizados tiveram seus resultados demonstrados em termos das curvas Receiver Operating Characteristic (ROC) e Cumulative Match Characteristic (CMC) e do *equal error rate*.

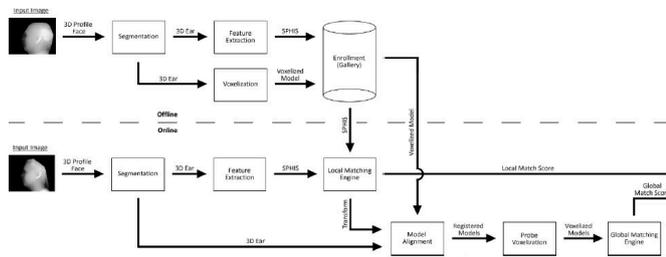


Fig. 67. Representação do método proposto por Jindan Zhou. Figura extraída de [210].

Decker [214] apresenta duas abordagens de classificação automática de objetos 3D em imagens 2D. A primeira tem como base modelagem estatística de características de wavelets e utiliza a estimativa por semelhança máxima para determinar a classe do objeto na cena. De forma geral, essa abordagem pode ser resumida em treinamento e classificação. Na parte de treinamento, a imagem do objeto é adquirida de diferentes pontos de vista, pré-processada em um dos espaços de cor investigado, extraído características, definida a área de objeto e estimada a função de densidade de probabilidade. Na parte de classificação, um conjunto de vetores de características é determinado de uma imagem de teste e avaliado contra as funções de densidade de todas as classes de objetos consideradas. A segunda abordagem tem como base descritores de pontos locais robustos. Inicialmente, para o treinamento,

características SURF são extraídas partindo de imagens de treinamento. Depois, para a parte de reconhecimento, as características da imagem são correspondidas geometricamente através da correspondência entre os descritores das imagens de treinamento e de teste, de forma a encontrar o objeto com maior correspondência à imagem pesquisada. Os experimentos foram realizados para diferentes conjuntos de treinamento com diferentes planos de fundo da base de dados 3D-REAL-ENV [215], e os resultados foram demonstrados em termos de taxa de classificação comparando as duas abordagens apresentadas.

Ryo Hanai [216] apresenta uma base de dados de peças eletrônicas e inicialmente compara algumas características utilizadas, aplicando-as a base de dados. Com base em um *survey* apresentado por Akgul [217] escolheu 2 métodos de extração de características, Density-based Framework (DBF) [217] e o CRSP [218], e adicionou mais um método, o Surflet pair relation histograms (SPRH) [219], visando analisar qual método é o mais apropriado para a classificação da base de dados em questão. As medidas para avaliar os métodos foram o Discounted Cumulative Gain (DCG) and Nearest Neighbor (NN). No NN, utiliza-se a porcentagem das primeiras correspondências mais próximas que pertencem a classe investigada. Uma alta pontuação do NN indica o potencial do algoritmo em uma aplicação de classificação. O DCG avalia uma sequência inteira de objetos ordenados através de uma medida de similaridade e fornece um maior peso para os melhores classificados na relação de corretos e incorretos. Os três métodos foram aplicados a toda a base de dados e utilizou-se o DCG e o NN para avaliar os resultados. Por fim, avaliou-se o uso de combinação de características para melhorar os resultados, obtendo uma grande melhoria demonstrada pela pontuação derivada da avaliação utilizando o DCG e o NN.

Uma abordagem utilizando um filtro de partículas de duas camadas é proposto por Sukhan Lee [220] para o reconhecimento e estimativa da pose de objetos 3D. Na camada superior, um conjunto de poses de objetos candidatos é identificada e preservada no espaço de busca como um conjunto de super-partículas, onde para cada uma é atribuída uma probabilidade de pose verdadeira, a qual é desenvolvida ao longo do tempo com o acúmulo de evidências futuras. Para definir as candidatas a pose do objeto, primeiro, adquire-se, inicialmente, evidências fracas interpretando-as em termos de possíveis poses do objeto no espaço. Estas interpretações servem como região de interesse para uma detalhada investigação, pela qual as probabilidades das poses são computadas, para interpretações individuais, com base na probabilidade e improbabilidade de várias características disponíveis nas correspondentes regiões de interesse. Durante o processo de computação das probabilidades, são selecionados os candidatos a pose do objeto para serem utilizados como super-partículas na camada superior. Na camada inferior, as incertezas de pose associadas com os candidatos individuais são representados como partículas que estão sujeitas à propagação no tempo. Mais especificamente, a camada superior está lidando com problemas de ambiguidade e a camada inferior está lidando com problemas de incerteza. Comparado com o convencional filtro de partículas, onde a informação de observação apenas atualiza a probabilidade da partícula

propagada, na abordagem proposta atualiza-se a probabilidade e a pose da super-partícula simultaneamente em um estágio de fusão. Neste estágio, de forma a evitar fusões randômicas, mede-se a relação de suporte entre a super-partícula propagada e uma nova super-partícula observada em termos de distância de mahalanobis. A fusão irá gerar novas super-partículas com poses e probabilidades atualizadas, de forma que após a fusão o número total de super-partículas é aumentado. Assim sendo, por uma questão de balanço computacional, uma reamostragem é utilizada para gerar um certo número de super-partículas. De forma resumida, o algoritmo de filtro de partículas com duas camadas pode ser descrito pelas seguintes etapas: Inicialização, propagação, geração de novas super-partículas, fusão de super-partículas e reamostragem das super-partículas. O sistema proposto foi experimentalmente avaliado para geração de múltiplas interpretações, propagação e fusão de super-partículas e filtragem de estimativas de pose. Os experimentos utilizaram um Pioneer 3-AT robot platform e uma câmera estéreo, bumblebee e os resultados foram demonstrados em termos da análise de erro de convergência.

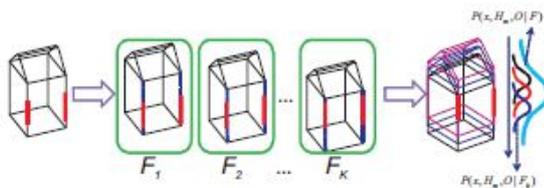


Fig. 68. Representação do método proposto por Sukhan Lee. Figura extraída de [220].

Um *framework* para classificação de objetos de formato livre em nuvens de pontos é apresentado por Eunyong Kim [221]. O *framework* proposto, inicialmente, segmenta o objeto candidato da cena e então identifica a classe para cada candidato. O *framework* pode ser resumidamente descrito em três partes principais: A estrutura hierárquica da base de dados, o classificador de objetos e o processo de aprendizagem online. A estrutura hierárquica ou hierarchical structured database (HSD) foi construída no intuito de fornecer uma rápida inferência no rotulamento e na aprendizagem online. O modelo utilizado para construir tal estrutura foi o modelo TAX[222], o qual constrói um modelo hierárquico das classes de objetos a partir de imagens de profundidade não rotuladas, através de um mapeamento de cada imagem para um caminho em uma árvore composta por L nodos. À esta estrutura foi proposto uma representação de objetos, denominada *visual word*, a qual utiliza o contexto espacial entre uma superfície plana e o objeto; O classificador de objetos utiliza informações dos padrões dos objetos, a distribuição, representado pela frequência de todas as *visual words* no objeto, a estrutura montada pela HSD, o k-mean, para o processo de clusterização das visual words e Bhattacharyya coefficient para inferir os rótulos dos objetos; No processo de aprendizagem online, um processo de aprendizagem é utilizado para inferir a estrutura da HSD, entretanto, em muitas aplicações, a HSD deveria crescer com o passar do tempo. Dado que é inviável rodar o algoritmo em lote repetidamente toda vez que uma nova

imagem é adicionada propõe-se um algoritmo de aprendizagem online que incrementalmente atualiza a HSD existente. Dado uma HSD inicial treinada pelo algoritmo em lote, a base de dados aprende um novo objeto através da atualização local graças a sua estrutura de árvore. Diferente do processo de aprendizagem em lote no qual as atribuições de tópico, nível e caminho são iterativamente amostrados em turnos, para a atualização local, o processo de aprendizagem online inicia escolhendo um caminho ao qual a imagem deveria pertencer. Então, atribui as variáveis de nível e tópico para cada ponto de interesse na imagem, por uma amostragem de Gibbs. Depois, deve-se reamostrar as atribuições de tópico e nível das imagens existentes associadas ao caminho em questão, de forma a atualizar as distribuições relacionadas com o caminho, condicionando não apenas através dos dados de treinamento, mas também ao novo objeto. Para avaliar a performance do *framework* proposto, foram utilizados modelos 3D disponíveis montando uma base de dados de imagens de profundidade e esta base foi utilizada para treinamento e testes. Os resultados foram demonstrados em termos de resultados de classificação, performance da aprendizagem incremental e tempo de computação.

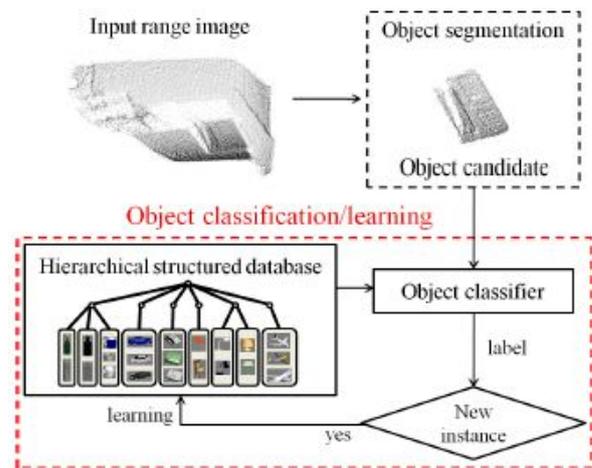


Fig. 69. Representação do método proposto por Eunyong Kim. Figura extraída de [221].

Akbar [223] discute o uso do Clonal Selection Algorithm (CLONALG)[224] and Particle Swarm Optimization (PSO) [225] para o reconhecimento de objetos 3D. Ao invés de utilizar um modelo pré-definido para extrair as informações geométricas, o objeto 3D é modelado com base em sua aparência 2D. Primeiro, a imagem do objeto é segmentada utilizando o método OTSU thresholding. Depois, um conjunto de características, Hu moments invariants, é extraído. Então, o CLONALG e o PSO são utilizados, separadamente, para inicializar os pesos da rede neural e a rede neural tem seu treinamento continuado pelo algoritmo Levenberg-Marquardt. Por fim, utiliza-se a rede treinada para classificar os objetos. Para os experimentos utilizou-se um subconjunto da base de dados Columbia Database [23], composto por 5 objetos e comparou-se o resultado das duas abordagens de inicialização da rede neural com outras abordagens de

inicialização/treinamento (Levenberg-Marquardt (LM), Scaled Conjugate Gradient (SCG), Gradient Descent (GD), Gradient Descent with Momentum (GDM), Gradient Descent with Adaptive Learning Rate (GDA), and Gradient descent with momentum and adaptive learning rate (GDMA)) em termos de acurácia de classificação.

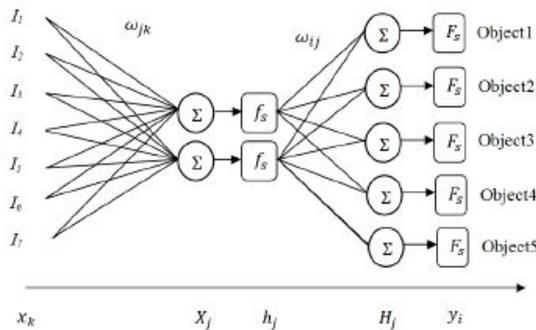


Fig. 70. Representação do método proposto por Akbar. Figura extraída de [223].

Woongji Jeong [226] apresenta a seleção de características com modelagem estatística em ambientes reais para o reconhecimento de objetos 3D e estimativa de pose. Para o robusto reconhecimento de objetos e estimativa de pose em vários ambientes, tentou-se utilizar várias características (SIFT[42], line[227] e cor). No entanto, a confiabilidade de cada característica muda com as mudanças do ambiente, como por exemplo mudança na iluminação, oclusão e variação na distância. Para reduzir o efeito dessas mudanças ambientais, estimou-se a mudança das características em diferentes ambientes para realizar uma seleção de características razoável utilizando os seguintes métodos. Previu-se a quantidade esperada da característica pela combinação da probabilidade de detecção (modelo estatístico - modelo probabilístico que extrai características de variáveis ambientais medidas a partir de regiões de interesse) de cada característica e quantidade de característica ociosa (modelo do objeto) de um objeto que se pode ver do ponto de visão atual. Além disso, calculou-se a confiabilidade de cada característica através da combinação de função de utilidade R para decidir se as características esperadas são válidas no processo de reconhecimento do objeto e estimativa de pose. A função R fornece o peso com base em valores experimentais antes de decidir a probabilidade final de cada característica. Com base na probabilidade final, decide-se a característica ótima. Selecionando a característica ótima em mudanças ambientais permite um aumento na acurácia da fusão e da filtragem [228], [115] e benefícios nos cálculos, dado que não se utiliza características com baixa confiabilidade. No intuito de provar o método de seleção de características proposto, testou-se o método em vários ambientes em diferentes condições de distância e iluminação, mostrando os resultados para as três características estudadas em termos de probabilidade de ocorrência de cada característica nos diferentes ambientes.

Aldoma [229] propõe um método de verificação de hipóteses para o reconhecimento de objetos 3D. O mesmo autor utilizou este método para propor um *framework*

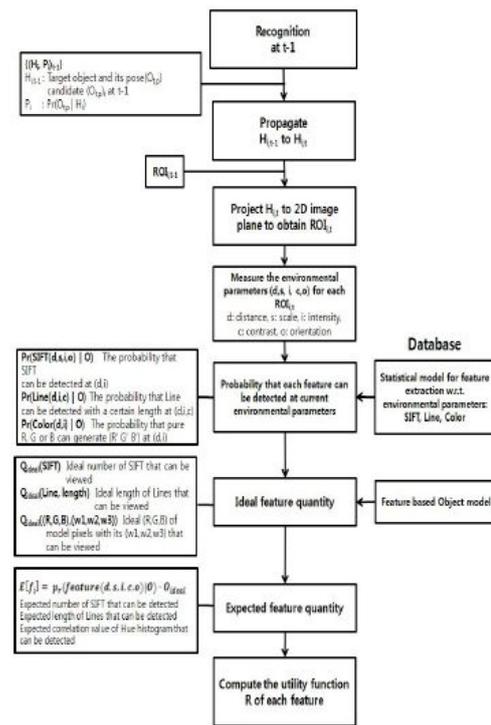


Fig. 71. Representação do método proposto por Woongji Jeong. Figura extraída de [226].

de verificação de hipóteses global em cenas desordenadas [230]. Inicialmente, pontos chave são extraídos em posições amostradas uniformemente na superfície da cena e do modelo. Então, o descritor local, SHOT [150], é computado para cada ponto chave sob um tamanho de suporte definido por um raio. Depois, os descritores calculados são correspondidos para atingir uma correspondência ponto-a-ponto. Para lidar com casos com múltiplas instâncias do mesmo modelo, cada descritor da cena é correspondido, via indexação rápida, com todos os descritores dos modelos. Sucessivamente, um algoritmo de correspondência de grupo é utilizado para obter o modelo de hipótese a ser inserido no algoritmo de verificação de hipóteses proposto. A abordagem de correspondência de grupo iterativamente agrupa subconjuntos de correspondências baseadas na verificação da consistência geométrica do pares de correspondência. Dado que cada subconjunto de correspondências fornecidos pela correspondência de grupo define um modelo hipótese, um valor limiar influencia na cardinalidade final do conjunto de hipóteses e, conseqüentemente, na eficiência computacional do estágio de verificação de hipóteses. Como um refinamento adicional, utiliza-se o RANSAC em cada subconjunto obtido na etapa anterior, sendo o modelo uma transformação 6DOF fornecido pelo método de Absolute Orientation (AO) [231]. Por fim, utiliza-se o Iterative Closest Point (ICP) em cada subconjunto para refinar a pose 6DOF fornecida pelo AO. Neste ponto, o conjunto de hipótese está pronto para ser alimentado no estágio de verificação de hipóteses proposto. O estágio de verificação de hipóteses utiliza uma função de custo definida sob um espaço de solução de forma a selecionar um número arbitrário de hipóteses per-

tencentos ao conjunto de hipóteses candidatas, maximizando o número de soluções corretas e minimizando o número de reconhecimentos errados. Ou seja, define-se uma função de custo e sobre esta utiliza-se um processo de otimização do custo com base em sugestões geométricas evidenciadas pelo autor. Dois experimentos foram conduzidos, um realizado na base de dados denominada *Kinect*, onde correspondem-se um conjunto de modelos CAD contra uma cena adquirida pelo *Kinect* e comparam-se os resultados obtidos pelo método proposto com os resultados do método apresentado em [232]. O segundo experimento utiliza a base de dados para reconhecimento de objetos 3D adquiridos por um escâner a laser [81]. Em ambos os experimentos os resultados foram comparados em termos de taxa de reconhecimento por quantidade de oclusão.

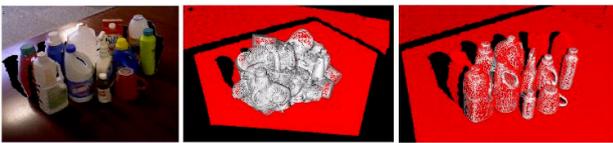


Fig. 72. Representação do método proposto por Aldoma. Figura extraída de [229].

Petricek [233] apresenta um método com base em características para o reconhecimento de objetos 3D em cenas desordenadas. O método proposto é aplicado sobre malhas poligonais, construídas partindo dos dados adquiridos pelos sensores de profundidade, para estabelecer o Local Reference Frame (LRF) e criar descritores de características. Para obter o vetor base e criar um LRF repetível, o qual seja robusto ao ruído aplica-se um processo similar a [234], o qual estima a orientação local da imagem. Neste caso, procura-se a direção a qual concorda com a maioria das normais da superfície. Já que foi assumido que as superfícies normais são constantes sobre polígonos individuais, utiliza-se suas áreas das superfícies como pesos e maximiza-se sobre um sujeito b , com base em uma fórmula descrita pelo autor. Os dois valores correspondendo aos dois maiores valores, em ordem decrescente, fornece os dois vetores base. Posteriormente, aplica-se um procedimento com base em [235] para obter um LRF não ambíguo. Todo esse procedimento serve para calcular um LRF único e não ambíguo. Para abstrair de pequenos deslocamentos espaciais ou angulares do LRF escolheu-se um descritor com base em histograma, similar ao MeshHOG [236]. O histograma essencialmente acumula projeções das normais das faces. Cada normal da face da área ponderada e seu centroide são projetados em três planos ortogonais com eixos alinhados com os vetores base. A coordenada angular da normal projetada determina pesos dos *bins* de orientação e as coordenadas da projeção do centroide determinam os pesos dos *bins* espaciais. Cada voto, ponderado pela coordenada radial da projeção da normal, é interpolado linearmente entre os dois *bins* de orientação mais próximos e *bilinearmente* entre 4 *bins* espaciais. Três planos de projeção, 4 *bins* espaciais e 12 *bins* de orientação fornecem um descritor de tamanho 144, no geral. O método de reconhecimento com base no descritor previamente apresentado consiste de duas fases, aprendizagem offline e reconhecimento online. Na fase de aprendizagem

as características são computadas densamente para o modelo. Na fase de reconhecimento características na mesma escala são computadas para a cena, espaçadamente em localizações randômicas. Estas características da cena são então correspondidas com as do modelo para formar um conjunto de tentativas de correspondência, cada uma fornecendo uma estimativa preliminar da pose do objeto. A hipótese final do objeto é gerada por um procedimento com base em consenso. Nos experimentos utilizou-se a base de dados de [138], [81] e os resultados obtidos com o método proposto foram comparados com outros métodos ([16], [136], [138], [81], [237]) em termos de taxa de reconhecimento.

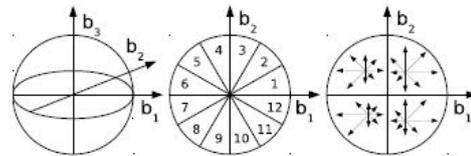


Fig. 73. Representação do método proposto por Petricek. Figura extraída de [233].

Yabushita [238] propõe um *framework* para o reconhecimento de objetos 3D que necessita de poucas imagens de referência. O *framework* proposto, primeiro, estima um modelo 3D de um objeto a partir de uma sequência de vídeos e gera uma única imagem alvo através de uma projeção esférica do modelo 3D e da textura. O *framework* reconhece o objeto pela correspondência da imagem alvo contra a imagem de referência armazenada na base de dados. A imagem armazenada na base de dados é uma foto do objeto em uma posição arbitrária. O processo de correspondência, inicialmente, detecta pontos chave na imagem alvo e nas imagens da base de dados e os descreve utilizando o vetor de características do descritor SURF. Depois, calcula-se a distância entre os vetores das imagens alvo e da base. Para cada ponto chave na imagem alvo, a técnica procura os k -nearest neighborhood pontos chave nas imagens da base de dados e calcula a pontuação de votos a partir das distâncias para as k imagens. A imagem recebendo o maior número de votos torna-se o resultado de reconhecimento. Dois experimentos foram conduzidos para verificar a performance da técnica proposta. Para os experimentos utilizou-se a base de dados KIT ObjectModels Web Database [239] e os dois experimentos procuraram avaliar o uso da imagem alvo a partir da projeção esférica e a técnica de correspondência proposta em relação a outras duas técnicas de correspondência (correspondência utilizando apenas os pontos chaves observados, sem os pontos da vizinhança e o Changed NBNN Technique) em termos de acurácia obtida.

Um método para extração de características geométricas com base em modelagem do olhar, para o reconhecimento de objetos 3D com base em modelo, é proposto por Maeda [240]. No processo de modelagem, superfícies locais do modelo são independentemente estimadas para partes dos dados de profundidade restringidos por vários domínios do olhar. Então, dado que as características são extraídas independentemente de cada domínio do olhar, características inconsistentes ou inco-

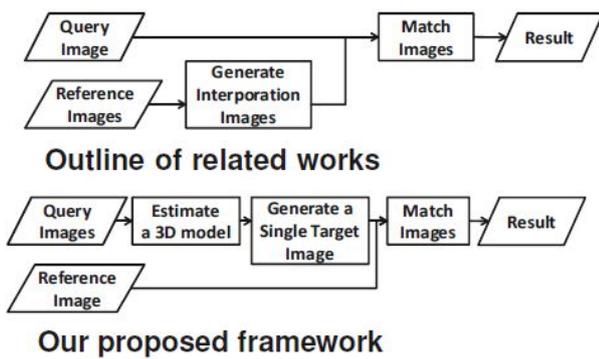


Fig. 74. Representação do método proposto por Yabushita. Figura extraída de [238].

erentes podem ser obtidas. Portanto um método estocástico é introduzido que permite integrar tais características pela avaliação da confiabilidade de cada modelo de olhar. Visando evitar a geração de vários descritores para uma imagem, tenta-se gerar um descritor de formato para um número limitado de pontos de característica. O descritor proposto, curvature distribution image (CDI), é gerado com base nas taxas entre curvaturas da superfície para codificar eficientemente a distribuição da curvatura perto de um ponto de referência. Afim de discutir a performance do descrição de formato 3D pelo tipo de curvatura utilizada, 5 tipos de curvaturas foram utilizadas (Gaussian curvature, mean curvature, principal curvatures (maximum, minimum) e shape index [113], [241]). Consequentemente, uma vez que um objeto é representado por um conjunto de CDIs descritivos, o reconhecimento é realizado pela descoberta do objeto na base com os CDIs similares ao objeto inquerido. Com o intuito de avaliar a habilidade de distinguir um formato 3D com base em CDIs, dois experimentos com base em 15 objetos foram realizados. O primeiro experimento, avaliou os 5 tipos de CDIs, com base nos tipos de curvaturas, e comparou-os com o spin image em termos de taxa de reconhecimento. O segundo experimento, testou o descritor na presença de oclusão própria dos objetos na cena, comparando os CDIs gerados por parte dos dados de 3 objetos com os CDIs dos objetos armazenados anteriormente.

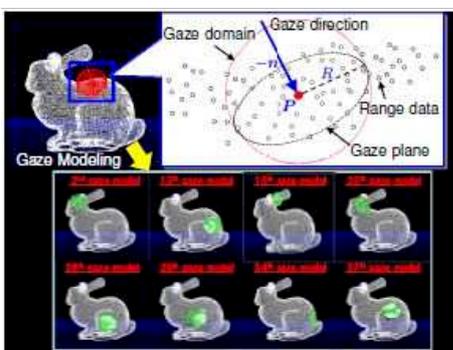


Fig. 75. Representação do método proposto por Maeda. Figura extraída de [240].

Sheta [242] apresenta um *pipeline* de processamento de

imagem para reconhecer objetos 3D com base nas imagens 2D deles. O *pipeline* completo apresenta os seguintes estágios de processamento: aquisição de imagens, pré-processamento, extração de características e classificação utilizando um modelo matemático Fuzzy das características extraídas. Na parte de aquisição de imagens três câmeras foram utilizadas para adquirir o objeto de interesse, posicionado em uma mesa giratória, de diferentes posições. Para cada objeto são adquiridas 72 imagens. A parte de pré-processamento consiste da segmentação automática da imagem capturada utilizando um método de *thresholding* iterativo, seguido por processos de encolhimento e crescimento para fornecer uma separação clara entre objeto e plano de fundo [243]. No estágio de extração de características, um conjunto de experimentos foi implementado para extrair características de *moments* dos objetos segmentados. As características incluem sete Hu's moments[244], onze Zernike's moments [245] e seis affine moments [246], [247]. A parte de classificação utilizando um modelo matemático Fuzzy tem como objetivo considerar as características extraídas das imagens 2D como entradas e fornecer como saída o número do objeto. O modelo Fuzzy Takagi-Sugeno foi utilizado, por ser considerado adequado para classificar um conjunto de dados dinâmico [248]. Para avaliar a performance do modelo fuzzy desenvolvido para classificação de objetos 3D utilizou-se o critério Variance-Accounted-For (VAF). Um total de 792 imagens foram utilizadas para implementar o modelo proposto. Metade das imagens foram utilizadas para o treinamento e o resto para teste. O critério VAF foi utilizado para medir quão próxima está a relação entre o objeto atual e o número estimado pelo sistema de classificação fuzzy. Quanto maior o VAF melhor é a acurácia de classificação.

Um sistema para reconhecimento e rastreamento de objetos que utiliza a informação de profundidade de um sensor de baixo custo é apresentado por Bongale [249]. O reconhecimento do objeto envolve dois estágios distintos, treinamento e teste. Na estágio de treinamento, captura-se uma foto 3D do objeto, sendo treinado, a qual é mantida num plano de fundo plano, tal como uma mesa ou o chão. Esta foto retirada com os sensores de profundidade é armazenada no formato de nuvem de pontos. O objeto de interesse nesta foto, é extraído utilizando o Euclidean Cluster Extraction. Esta clusterização resulta em um número de *clusters* indo de 2 até valores altos como 20 *clusters*, dependendo da uniformidade do plano de fundo, um destes *clusters* é o objeto. O cluster que representa o objeto de interesse é escolhido e o resto é descartado. Os Viewpoint Feature Histograms (VFH) [145] deste *cluster* são obtidos. Este processo é repetido para diferentes poses do mesmo objeto e para diferentes objetos. Deste modo, todas as poses e seus respectivos descritores VFH são obtidos e armazenados em uma base de dados. Todos esses histogramas para diferentes poses são convertidas no formato FLANN [146] e adicionados a base de dados. Uma árvore KD da base de dados sendo treinada é obtida para propósitos de reconhecimento. No estágio de teste, os objetos são mantidos na frente da câmera 3D para o reconhecimento. A imagem de profundidade obtida é *clusterizada* utilizando o Euclidean Cluster Extraction. Para cada *cluster* na imagem de profun-

didade o VFH é estimado. Os descritores VHF de cada um destes *clusters* são correspondidos com os descritores VFH dos objetos presentes na base de dados utilizando a árvore KD treinada. O vizinho mais próximo da base de dados fornece o objeto e a pose se requerida. O método proposto foi testado com 5 objetos e os resultados do reconhecimento dos objetos foram demonstrados em termos de taxa de sucesso.



Fig. 76. Representação do método proposto por Bongale. Figura extraída de [249].

Bariya [250] apresenta um *framework* para explorar a variabilidade da escala geométrica 3D em imagens de profundidade que fornecem ricas informações para caracterizar a geometria geral. Para tanto, deriva-se uma representação escala-espço, através da convolução do campo normal de uma imagem de profundidade com *kernels* Gaussianos, a qual os autores se referem como escala-espço geométrico. Esta representação, fielmente codifica a variabilidade da escala da geometria da superfície e deriva detectores para extrair características proeminentes e identificar a escala natural delas. Um *framework* similar foi apresentado pelo autor para o reconhecimento de objetos 3D em imagens desordenadas [237]. O resultado é um conjunto hierárquico de características de diferentes escalas, as quais os autores se referem como características geométricas dependentes da escala. Então, gera-se um descritor de formato local que representa as estruturas da superfície que dão origem àquelas características por esculpir e codificar a superfície local que cai dentro da região de suporte das características. Isso leva a um descritor de formato local dependente da escala ou invariante a escala que transmite informação discriminativa significativa da geometria do objeto. Esse descritor fornece fortes deixas para a correspondência e alinhamento de dados geométricos 3D. Isto é demonstrado através da execução do registro de imagens de profundidade. Também mostra-se como a hierarquia induzida pela variação da escala de estruturas geométricas locais pode ser empregada no auxílio ao reconhecimento de objetos 3D de forma precisa. Para o reconhecimento de objetos os seguintes passos principais são executados: inicialmente constrói-se uma biblioteca de modelos 3D dos objetos que se existe o interesse de reconhecimento nas cenas alvo. Depois, utiliza-se uma *interpretation tree*[251], [252] para realizar todas as possíveis correspondências entre as características da cena e do modelo. Entre todas as hipóteses representadas pelos

nodos folha da *interpretation tree*, são escolhidas apenas as hipóteses mais fortes para verificação, o que implica em utilizar a transformação geométrica definida pela hipótese para transformar o modelo 3D da biblioteca de objetos para a cena e avaliar sua acurácia dada pela área de sobreposição entre o modelo transformado e a cena. Então, escolhe-se a hipótese que produz a área máxima de sobreposição como sendo a melhor hipótese, a qual é refinada utilizando o Iterative Closest Point (ICP). A efetividade da abordagem proposta é demonstrada em experimentos de reconhecimento em imagens de duas bases de dados University of Western Australia [81] e Queen's University [253] em termos de taxa de reconhecimento por porcentagem de oclusão e desordem.



Fig. 77. Representação do método proposto por Bariya. Figura extraída de [250].

Socher [254] apresenta um modelo com base na combinação de redes neurais convolucionais e recursivas para aprendizagem de características e classificação em imagens de RGB-D. O modelo inicia com imagens RGB e de profundidade e, primeiramente, de forma separada, extrai características das imagens. Cada modalidade de imagem é inserida numa camada simples da rede neural convolucional (CNN)[255], a qual fornece características de baixo nível invariantes a translação, como por exemplo bordas, e permite que partes de um objeto sejam deformáveis até uma certa extensão. As saídas desta camada são então inseridas em redes neurais recursiva (RNN)[256], as quais podem aprender características composicionais e as interações entre as partes. As RNNs hierarquicamente projetam as entradas em espaços de menor dimensão através de múltiplas camadas com pesos e não-linearidades atreladas. Por fim, a concatenação de todos os vetores resultantes formam o vetor de características final utilizado em uma classificador softmax. Todos os experimentos realizados foram conduzidos utilizando a base de dados RGB-D dataset [257], na qual existem 51 classes de objetos e 300 instâncias destas classes. Cada instância de objeto é imageada de 3 diferentes ângulos resultando em aproximadamente 600 imagens por instância. Os resultados obtidos com o modelo proposto foram demonstrados em termos de matriz de confusão e comparados com outros métodos (Linear SVM, Kernel SVM, Random Forest, SVM, CKM e SP+HMP) em termos de taxa de reconhecimento.

Hao Ding [258] propõe um novo generalizado *affine moment invariants* denominado illumination invariant MSA moments. Este método combina os tradicionais affine moment invariants (AMI) [259], multiscale autoconvolution (MSA)[260] e MSA *moment* e as ideias básicas utilizadas para construir invariantes de iluminação. A performance do MSA *moments* degenera quando a iluminação varia. Isto indica que seria válido con-

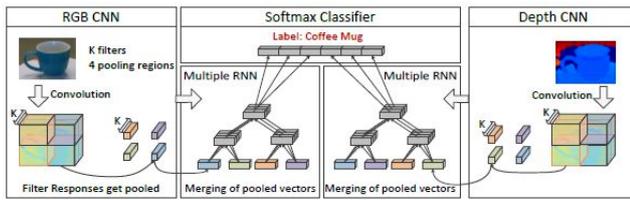


Fig. 78. Representação do método proposto por Socher Socher. Figura extraída de [254].

struir MSA moment invariante a iluminação. Ao contrário da construção do MSA moments, o illumination invariant MSA moments utiliza seis invariantes do tradicional Affine Moment, substituindo todos os *moments* centrais com *moments* centrais do MSA. Com o propósito de obter invariância a iluminação, elimina-se o fator λ da formulação proposta através de dois passos. O primeiro passo é a aplicação de uma normalização, fazendo com que cada componente seja inversamente proporcional ao fator λ . No segundo passo, cada componente é dividido pelo sexto polinômio do tradicional Affine Moment para assegurar a invariância a iluminação. Nos experimentos para ilustrar a performance do illumination invariant MSA moment contra o tradicional AMI e o MSA, na recuperação de formatos e no reconhecimento de objetos, utilizou-se o classificador por nearest neighbor em três diferentes conjuntos de dados. O conjunto das imagens binárias foi utilizado para a recuperação do formato, as imagens em tons de cinza foram utilizadas para avaliar a robustez das características e imagens em tons de cinza com vários pontos de visão foram utilizadas para o reconhecimento de objetos 3D, para este último experimento, utilizou-se um subconjunto das imagens da base de dados Coil-100 image [23]. Os resultados dos experimentos foram demonstrados em termos de taxa de reconhecimento.

Um descritor de características locais para o reconhecimento de objetos 3D é proposto por Youngkyoon Jang [261]. O descritor é uma combinação de Local Angle patterns (LAP), orientação de gradientes e informações de cor. O sistema de reconhecimento como um todo pode ser descrito nos seguintes passos: primeiro, pontos de características são extraídos da imagem. Para objetos com pouca textura o algoritmo extrai pontos de características através de uma amostragem randômica de posições nas bordas do objeto. Depois, com base nos pontos de característica extraídos, constrói-se descritores de características através da combinação de padrões LAP, orientação de gradiente e histogramas de cor dos fragmentos ao longo dos espaços R, G e B. Então os descritores são rotulados para aprender um livro de códigos através do Random Forest. Através do livro de códigos apreendido o algoritmo proposto monta um Bag-of-Words (BoWs). Para o teste, um classificador classifica o consultado BoWs para assegurar que a maior probabilidade é selecionada como um objeto reconhecido. Para avaliar a performance da taxa de reconhecimento do algoritmo proposto utilizou-se 10 objetos capturados em um ambiente controlado, para treinamento, e em um ambiente com rotações, variações de iluminação e com vários objetos na cena, para teste.

Yong-Jin Liu [262] propõe um método de reconhecimento

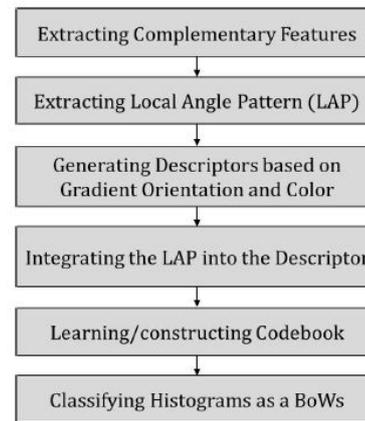


Fig. 79. Representação do método proposto por Youngkyoon Jang. Figura extraída de [261].

de objetos 3D com base no desenho de linhas. Inicialmente, apresenta-se uma representação visual circular de características utilizando componentes excitatórios e inibitórios, para extrair informações distintas dos desenhos das linhas. Esse procedimento tem como base o processo de formação da imagem de um objeto 3D na retina. Para simular esse processo, o modelo de objeto 3D é colocado de forma a coincidir o centro de gravidade com o centro de uma esfera a qual delimita o objeto 3D. Uma amostragem densa de pontos de visão é aplicada na superfície da esfera. Para cada ponto de visão, um modelo de iluminação é aplicado para gerar uma imagem de sombreamento do objeto e esta imagem de sombra é convertida em um desenho de linhas utilizando o algoritmo CLD [263]. Devido a grande quantidade de pontos, aplica-se um método de redução de dimensionalidade para obter um conjunto menor, mas representativo de linhas de desenho para cada objeto 3D. Com base na representação por linhas de desenho aplica-se o método de *Halton's quasi-random point sequence* para amostrar uniformemente um número de pontos dentro de uma região. Em cada ponto amostrado um histograma circular é estabelecido, no qual cada bin circular tem a mesma diferença em raio e o círculo máximo tem o raio igual a um quinto do tamanho da diagonal da área delimitada. O histograma de cada ponto de amostragem possui um número de bins e se o número de pontos pretos que caíram no bin for maior que um limiar então esse bin é ativado e utilizado no processo de reconhecimento. Com base nesta representação por histograma de características utiliza-se uma abordagem com base em livro de códigos para organizar e posteriormente fazer o processo de correspondência entre livros de códigos com base em uma métrica de similaridade apresentada. Para os experimentos comparou-se a o método proposto contra 5 outros métodos em termos de taxa de precisão e lembrança, utilizando a base de dados McGill 3D Shape Benchmark [170].

Wang ping[264] propõe o uso da entropia Tsallis para gerar uma representação concisa do spin image (SI). Esta representação, Tsallis entropy vector of spin image (TESI), pode converter um SI em um escalar pela utilização da entropia Tsallis do SI, de modo que um objeto 3D pode ser representado por um vetor, o qual é composto de valores de entropia

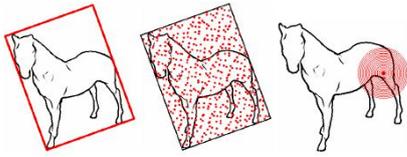


Fig. 80. Representação do método proposto por Yong-Jin Liu. Figura extraída de [262].

Tsallis. O SI é essencialmente um histograma 2D que representa a probabilidade distribuída quando coordenadas obtidas do spin map caem em um *bin* retangular. Esta probabilidade é conveniente para se calcular a entropia Tsallis. A entropia Tsallis, uma generalização da entropia Shannon, representa o grau de confusão do SI. Deste modo, um objeto 3D pode ser descrito por um vetor composto destes valores de entropia. Uma vez que um vértice de um Objeto 3D possui um SI, o qual foi denotado pela entropia Tsallis, a dimensionalidade do vetor é igual ao número de vértices. No processo de reconhecimento do objeto, um passo essencial é o cálculo da similaridade entre os dados do modelo e da cena. O trabalho apresentado utiliza o coeficiente de correlação utilizado em [265], [266]. Todas as características de invariância que o SI possui sem a entropia Tsallis continuam sendo satisfeitas ao se utilizar a entropia. O algoritmo para reconhecimento de objetos é descrito pelos seguintes passos: Para um modelo, todas as coordenadas dos vértices são normalizados; Constrói-se um SI de cada vértice e computa-se a entropia Tsallis; Com todas as entropias Tsallis dos vértices, constrói-se o vetor de representação como sendo o descritor do modelo; O descritor de todos os modelos na base de dados é calculado e feito um índice; Para uma imagem de entrada, todos os vértices são normalizados e calcula-se o vetor de entropia Tsallis dos objetos na cena; Computa-se a correlação entre os vetores de entropia Tsallis do modelo e da cena. Um limiar é setado afim de julgar se os dois objetos correspondem, ou seja, se a correlação calculada anteriormente é menor que o limiar setado então os dois objetos, muito provavelmente, são o mesmo objeto. As imagens utilizadas para os experimentos são da base de dados da universidade de florida. Nestes experimentos, o TESI foi avaliado no reconhecimento de objetos, segundo um valor de correlação do TESI do modelo e da cena, sob condições normais e na presença de ruído ou oclusão.

Uma abordagem híbrida com base em rede neural e na combinação do Hu & Zernike moments com descritores geodésicos é proposto por Bencharef [267] para o reconhecimento de objetos. Inicialmente, o objeto é segmentado do seu plano de fundo através dos operadores Sobel e um detector de borda, os quais são utilizados para calcular um valor de *threshold*. Após a extração do objeto da imagem, calcula-se o centro de massa do mesmo e com base nesse centro computa-se os descritores geodésicos. O descritor geodésico é definido como o vetor composto pelas diferentes distâncias geodésicas entre o centro e os pontos das bordas. Junto ao descritor geodésico computa-se também o Hu & Zernike moment para cada nível de cor da imagem colorida. O Zernike moment é um série de cálculos que transforma a imagem em vetores

com componentes reais representando os momentos. O Hu moments são usualmente utilizados na física para descrever a distribuição de massa em um corpo, através da combinação do nível de cinza de um ponto na imagem com a massa do corpo em cada ponto. Ambos os momentos são aplicados para extrair o máximo de informações do objeto em relação a cor e o descritor geodésico descreve informações do formato do objeto. Os descritores são utilizados como vetor de entrada para uma *back propagation artificial neural network*. Nos experimentos realizados utilizou-se a base de dados Coil 100 [23] e para o reconhecimento de objetos, testaram-se diferentes tamanhos de padrões de entrada para o processo de aprendizagem, ou seja, diferentes combinações dos descritores utilizados, sendo os resultados analisados e comparados com base na taxa de reconhecimento.

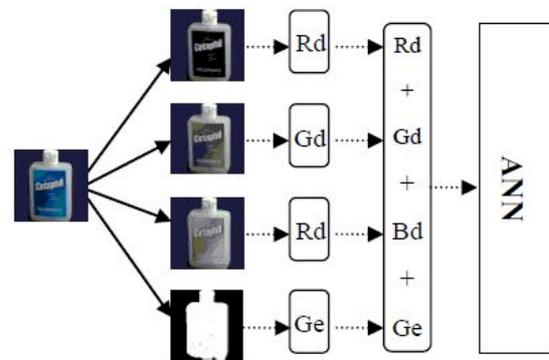


Fig. 81. Representação do método proposto por Bencharef. Figura extraída de [267].

Liang-Chia Chen [268] apresenta um esquema para detecção automática de objetos. Na abordagem proposta, o esquema de detecção de objetos pode identificar objetos alvo automaticamente em imagens de profundidade utilizando um processo inicial de segmentação do objeto, para sub-dividir todos os possíveis objetos na cena, e então aplicar um processo de classificação de objetos com base em restrições geométricas e histograma de ângulo de visão para a classificação do objeto. O método de segmentação dos objetos na cena está detalhado em [269]. O *framework* de classificação utiliza uma combinação de várias características do objeto. O objetivo da classificação do objeto é alcançado encontrando a correlação das dimensões relevantes, densidade dos pontos e histograma de ângulos de visão entre a nuvem de pontos segmentada e os modelos de objetos armazenados em uma base de dados. Três parâmetros de dimensão são calculados dos dados segmentados (altura, largura e comprimento). Para este cálculo emprega-se o principal component analysis (PCA) para encontrar os autovalores e autovetores da matriz de covariância para estimar a orientação do objeto. Com a determinação dos 3 autovetores que definem a direção de orientação da nuvem de pontos, pode-se medir os parâmetros de dimensão ao longo destes 3 autovetores. Após a aquisição dos três parâmetros, estima-se a densidade da superfície do objeto dentro do cubo criado pelas três dimensões encontradas. Para comparar os quatro parâmetros, o método proposto utiliza um conjunto fuzzy para incluir todos os objetos modelo na base

de dados. Cada objeto modelo na base de dados contém os quatro parâmetros com quatro funções triangulares de adesão. Para cada função de adesão, um valor ativado pertencente a $[0,1]$ é fornecido para mostrar o fator de correlação do valor de entrada com o valor correspondente do objeto modelo. O processo de defuzzificação utiliza o método de centro da área para determinar o valor de decisão para cada objeto com coeficientes de peso como valores de entrada. Estes coeficientes são setados de acordo com as características geométricas do objeto, em que se destaca a importância da largura, altura, comprimento ou densidade. Outra característica utilizada para a correspondência entre os objetos é o histograma de ângulo de visão, o qual é calculado com base no ângulo entre a normal da superfície de um ponto na nuvem de pontos e seu vizinho. Com base neste histograma calcula-se um valor de similaridade entre histogramas, para os objetos, o qual é o valor da correlação cruzada de dois histogramas definido pela fórmula de Pearson's para o cálculo de correlação. Quando comparando uma nuvem de pontos segmentada com um modelo na base de dados, o histograma de nuvem de pontos é comparado com todos os histogramas gerados para o modelo afim de encontrar o maior valor de coeficiente de correlação. O valor final de coeficiente de correspondência utiliza o valor de correspondência dos histogramas associado ao valor final de correspondência para os quatro parâmetros. O método proposto foi testado em um conjunto de dados produzidos por um sistema de escaneamento a laser e os resultados foram demonstrados em termos de classificação dos objetos testados.

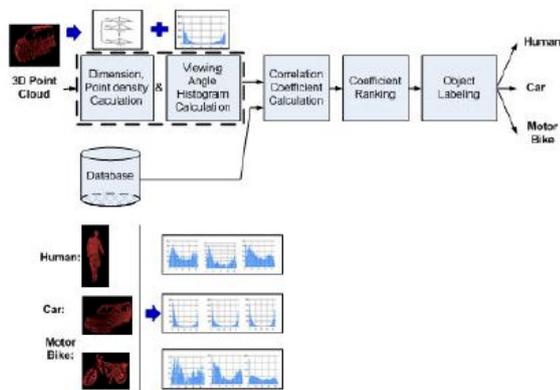


Fig. 82. Representação do método proposto por Liang-Chia Chen. Figura extraída de [268].

Megherbi [270] apresenta uma comparação de uma variedade de abordagens de classificação para identificação automática de objetos de ameaça em imagens de tomografia computadorizada. Através da combinação de técnicas de segmentação de imagens médicas 3D com métodos de classificação com base no formato 3D e recuperação, comparam-se 5 abordagens de estágios finais de classificação e apresenta-se conquistas significativas de desempenho na detecção automatizada de itens exemplares especificados. A abordagem de detecção de objetos, pega os objetos pré-segmentados da imagem de tomografia computadorizada utilizando a abordagem de segmentação apresentada em [271].

Os objetos são pré-segmentados, normalizados e, para cada instância de objeto 3D, uma representação por vetor de características é construída. Os objetos 3D segmentados são descritos pelos descritores de característica 3D Zernike [272] e o Histogram of Shape Index (HSI) [241]. Para o treinamento inicial, offline, do classificador, esta metodologia é utilizada para construir um conjunto de exemplos positivos e negativos de vetores de características. Na fase de classificação/reconhecimento uma tomografia computadorizada não vista é segmentada em múltiplos objetos conectados, dos quais uma representação por vetor de características é similarmente construída para cada objeto para permitir a classificação via um classificador. Cinco abordagens de classificação são comparadas para a classificação dos itens: Support Vector Machines [273], Artificial Neural Network [274], Decision Trees [275], Boosted Decision Trees [276] e Random Forests [277]. Cada abordagem é treinada, utilizando uma abordagem de validação cruzada, para fornecer uma detecção binária, presente ou não, para uma classe de objeto dada. A avaliação tem como base um conjunto de imagens tomográficas de bagagens contendo duas classes de objetos de ameaça, armas de fogo e garrafas [278][279]. Os resultados para cada classificador foram demonstrados em termos de acurácia, precisão, taxa de verdadeiros negativos e lembrança, utilizando o vetor de características apenas com Zernike, apenas com HSI e com os dois combinados.

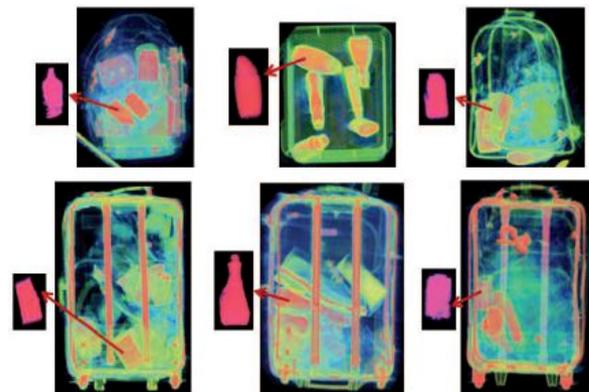


Fig. 83. Representação do método proposto por Megherbi. Figura extraída de [270].

Shaiek [280] propõe um método de reconhecimento de objetos 3D que emprega um conjunto de pontos chave 3D extraídos de uma representação de nuvem de pontos de visões 3D. O método proposto inicia com a reamostragem da nuvem de pontos inicial e a montagem de uma malha com os pontos reamostrados. Como próximo passo tem-se a detecção de pontos chave utilizando um detector proposto da combinação dos critérios de classificação do par mean-Gaussian curvatures (HK) e do par shape index-curvedness (SC) denominado $SC_H K_{connex}$. Após a detecção dos pontos chave, computa-se um descritor 3D ao redor de cada ponto de interesse. O descritor 3D apresentado, denominado IndSHOT, tem como inspiração os descritores CSHOT e SHOT. Para a construção do IndSHOT, inicialmente, acumula-se a contagem de pontos em bins de acordo com uma função cosseno do ângulo entre a

normal de cada ponto dentro da parte correspondente do grid e a normal no ponto da característica. Para cada histograma local, um agrupamento grosseiro é criado para as direções próximas a direção da normal de referência e um agrupamento melhor é criado para direções ortogonais. Depois, valores de *shape index* dos pontos de característica e de seus vizinhos são agrupados em bins. Por fim, os valores de *shape index* e os valores de cosseno são fundidos formando o IndSHOT. O descritores finais compostos por ID do modelo, *index shape* justaposto com o histograma dos cossenos, tipo da superfície, as coordenadas 3D dos pontos chave e média e desvio padrão do *shape index*, são armazenados e utilizados no processo de correspondência. O detector e descritor propostos foram validados utilizando uma abordagem de correspondência entre visões. Dado um objeto de teste, computa-se uma medida de similaridade entre os descritores extraídos das visões de teste e das visões dos modelos na base de dados. Um sistema de voto incremental para cada classe fornece a classe correspondente. Para os experimentos com o método proposto foram utilizadas uma base de dados proprietária, capturada com Kinect e composta de 20 objetos, e uma base de dados pública, RGB-D Object Dataset [257]. Para avaliar os resultados foram comparados os resultados obtidos em termos de taxa de reconhecimento e matriz de confusão.

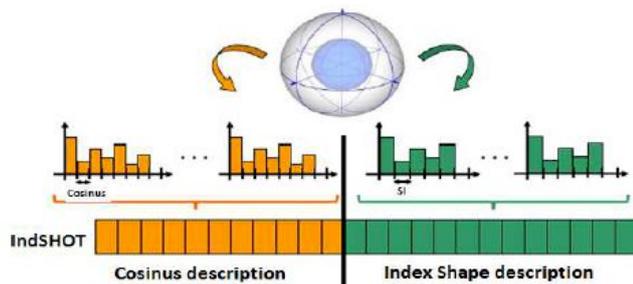


Fig. 84. Representação do método proposto por Shaiek. Figura extraída de [280].

Rodola [281] apresenta um *framework* para o reconhecimento de objetos 3D conhecidos, em escaneamentos incompletos e desordenados. Esta abordagem adota um algoritmo de seleção evolucionária que procura a concordância global entre pontos da superfície, enquanto operando em nível local. Logo, a abordagem efetivamente estende o escopo dos descritores locais através da seleção ativa de correspondência que satisfazem restrições de consistência global. O pipeline do *framework* proposto consiste em uma etapa de pré-processamento e dois jogos, da teoria de jogos, não-cooperativos. Esta abordagem foi utilizada também em um trabalho anterior dos autores, onde eles utilizam um jogo não cooperativo para a detecção de objetos 3D em cenas desordenadas [282]. A etapa de pré-processamento é realizada em ambos, modelo e cena. Esta etapa envolve uma seleção inicial de pontos relevantes nas respectivas superfícies, sendo o critério de relevância estipulado para evitar remendos da superfície que não são significantes para a parte de correspondência, tais como áreas planas. Todos os pontos de interesse do modelo são armazenados, enquanto os na cena são sub-amostrados uniformemente. Esta diferença

faz sentido devido algumas razões: em muitas aplicações o conjunto de modelos não muda durante o tempo, e portanto os descritores precisam ser calculados apenas uma vez; Em adição, a direção de correspondência é da cena para o modelo e tendo menos pontos na fonte do que no alvo permite proceder mais rapidamente sem comprometer a acurácia; Por fim, o modelo tende a ser medido com maior acurácia (ou porque tem mais tempo a ser gasto no processo ou porque vem de um modelo CAD). Um descritor é computado para todos os pontos mantidos, utilizou-se o SHOT [167], e estes são utilizados para construir os candidatos iniciais que, em turnos, são alimentados no jogo de correspondência. Em geral, um jogo de correspondência [283] pode ser construído definindo apenas quatro entidades básicas: um conjunto de pontos do modelo, um conjunto de dados de pontos, um conjunto de candidatos correspondentes e uma função de compatibilidade por pares entre eles. O objetivo do jogo é a de operar uma seleção entre os elementos do conjunto inicial de correspondências, com base nas entidades, estratégias e nas recompensas vinculadas as estratégias de forma a visar um *equilibrium* [283]. Com a finalidade de avaliar a performance do pipeline proposto, executou-se alguns experimentos comparando a performance da abordagem proposta contra outras técnicas [237] [81] [138][16]. O conjunto de dados utilizado foi o mesmo adotado em [237] [81] e os resultados foram apresentados com base nos valores de precisão e lembrança, taxa de reconhecimento em relação a quantidade de oclusão e desordem e resistência ao ruído.

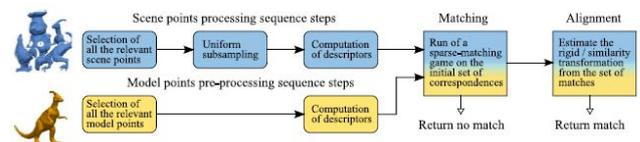


Fig. 85. Representação do método proposto por Rodola. Figura extraída de [281].

Keaikite [284] compara dois sistemas de reconhecimentos ativos. Ambos os sistemas [285], [286] utilizam características SIFT para o reconhecimento de objetos, ou seja, utilizam as mesmas características de baixo nível, mas fazem uso de modelos contrastantes de estratégias de atualização e seleção de ponto de visão. Em [286] todas as características SIFT extraídas de todas as imagens de treinamento são clusterizadas utilizando uma árvore de vocabulário [287]. Em [285] um modelo pseudo-3D para cada objeto de treinamento é gerado pela retenção de características que são visíveis de qualquer dois pontos de vista adjacentes. Outro fator diferenciando os dois sistemas é a estratégia de atualização. Enquanto [286] utiliza uma estratégia de atualização Bayesiana, e um mecanismo de seleção a priori para o próximo ponto de vista, em [285] é utilizado um modelo discriminativo, o qual é atualizado aditivamente e utiliza um mecanismo de seleção online. A base de dados utilizada na comparação foi criada pelo autor do [286] e dois experimentos foram realizados. No primeiro experimento, os dois sistemas foram executados com seus formatos originais e no segundo experimento, a estratégia de seleção do melhor próximo ponto selecionado

por Govender[286] foi substituído no sistema de Kootstra[285] para determinar se haveria uma melhoria ou um declínio de performance. Isto também ajuda a isolar a contribuição do modelo e da estratégia de seleção de ponto de visão nos resultados. Os resultados obtidos foram demonstrados em termos de taxa de reconhecimento dos objetos.

Oleari [288] apresenta um sistema de visão estéreo de baixo custo elaborado para o reconhecimento de objetos utilizando Fast Point Feature Histogram (FPFH) [137]. O sistema de visão estéreo de baixo custo, composto de duas câmeras Logitech C270 UVC (USB Video Class) fixadas em uma barra rígida, associado a um software para sincronização aproximada, fornece uma imagem de disparidade precisa e densa, a qual é transformada em uma nuvem de pontos para a realização do reconhecimento de objetos. A calibração das câmeras e o cálculo da imagem de disparidade é realizado utilizando os pacotes e bibliotecas fornecidas pelo *framework* Robotic Operating System. Dado que segmentação da cena não é o foco do trabalho, assume-se que o objeto a ser reconhecido está em uma superfície plana e dentro de uma região delimitada. Por fim, parte-se para o reconhecimento via cluster, o qual tem como objetivo computar a correspondência de um cluster selecionado com uma entrada no conjunto de dados de modelos conhecidos. Cada modelo é obtido pelo acúmulo de pontos de múltiplos frames para preencher as falhas da nuvem de pontos produzidas pela visão estéreo. Então um filtro de grid de voxel é aplicado para alcançar uma nuvem de pontos amostrada uniformemente. O algoritmo de reconhecimento tem como base o alinhamento de nuvem de pontos. As nuvens de pontos do modelo P_{mod} e do objeto de entrada P_{obj} no espaço 3D necessitam ser registradas ou alinhadas afim de serem comparadas. O processo de registro computa uma transformação geométrica rígida que deve ser aplicada ao P_{mod} para alinhá-la a P_{obj} . O registro é realizado em 3 passos: Remoção da dependência com um *frame* de referência externo, onde P_{mod} e P_{obj} são inicialmente expressados em referência aos respectivos centroides. Realização do alinhamento inicial entre as nuvens de pontos utilizando o RANSAC com base nos descritores FPFH como parâmetros para a função de consenso. E refinamento do alinhamento através do ICP, o qual minimiza a *mean square distance* entre os pontos. O reconhecimento é então realizado pela computação do valor de aptidão, o qual avalia a qualidade geral do alinhamento. Para cada ponto de P_{obj} o algoritmo calcula a *square mean distance* para o ponto de P_{mod} e recupera a porcentagem dos pontos cuja a distância é menor que um limiar fixado. O de reconhecimento itera sobre cada modelo no conjunto de modelos conhecidos e retorna o modelo com o maior valor de aptidão. O conjunto de dados de modelos conhecidos utilizada para os testes consiste em 8 objetos observados de 8 pontos de vista e os resultados dos testes foram demonstrados em termos de curva Receiver Operating Characteristics (ROC) e matriz de confusão.

Bennamoun [289] apresenta um sistema de reconhecimento de objetos 3D de formato livre com base em um descritor de características de superfície. O sistema segue os seguintes passos: Para um ponto de característica selecionado aleatoriamente, define-se um Local Reference Frame (LRF) através do cálculo dos autovetores de uma matriz de covariância de

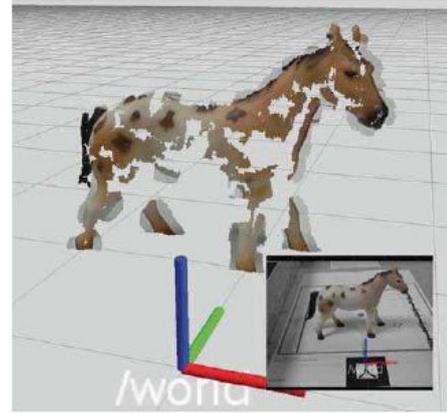


Fig. 86. Representação do método proposto por Oleari. Figura extraída de [288].

uma superfície local. Depois, um descritor de características denominado Rotational Projection Statistics (RoPS) é construído através do cálculo das estatísticas da distribuição do ponto no plano 2D definido pelo LRF. Por fim, apresenta-se o algoritmo de reconhecimento com base nas características RoPS: os modelos candidatos e as hipóteses de transformações são geradas através da correspondência de características da cena contra as características do modelo na biblioteca por meio de um cálculo de distância entre as características modelo-cena, utilizando uma árvore k-d, e um sistema de votação, no qual cada correspondência modelo-cena gera um voto para um modelo. Estas hipóteses são testadas e verificadas através do alinhamento do modelo a cena. Os experimentos com o sistema proposto foram realizados em duas bases de dados, Bologna [150] e UWA [81], e a performance foi avaliada através do *Global matching index* (GB) [290] e comparado com outros métodos de descritores de característica estado da arte (Spin Image [16], Normal Histogram (NormHist) [291], Local Surface Patches (LSP) [82], THRIFT [292] e SHOT [150]).

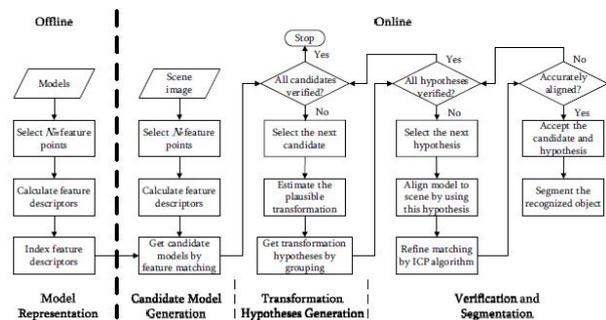


Fig. 87. Representação do método proposto por Bennamoun. Figura extraída de [289].

Yabushita [293] propõe uma técnica de reconhecimento de objetos 3D que permite ao usuário realizar buscas visuais moveis de objetos 3D. Está técnica com base no *framework* proposto anteriormente pelo autor [238], inicia pela extração de pontos chave de todos os frames do vídeo capturado. Depois, estes pontos chave são rastreados entre duas imagens

contíguas pela comparação de seus descritores. Pares de pontos chave que possuem os descritores mais similares através de uma série de duas ou mais imagens significa que eles são extraídos dos mesmos pontos no objeto. Utilizando esta computação, dois passos são realizados: reúnem-se pontos chave de múltiplas visões e estima-se as coordenadas 3D de cada ponto-chave reunido, utilizando a correlação da quantidade de movimento transversal e a profundidade. O resultado dos processos de agrupamento e estimativa são utilizados no processo de correspondência. Neste processo tem-se como foco a correspondência dos pontos chave entre a imagem de consulta e cada imagem na base de dados através da comparação de seus descritores. A técnica utilizada considera não só cada ponto chave isolado, mas também a similaridade de seus pontos chave vizinhos, os quais são ponderados pela distância 3D computada da coordenada 3D estimada. A técnica de correspondência está detalhadamente descrita em [238]. Os experimentos foram realizados com base em 7 imagens estáticas do objeto alvo em várias posições e um vídeo capturando o objeto alvo de várias posições.

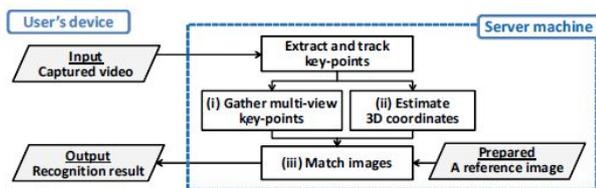


Fig. 88. Representação do método proposto por Yabushita. Figura extraída de [293].

Byung-soo Kim [294] propõe um *framework* que explora a compatibilidade entre hipóteses de segmentação do objeto na imagem e o mapa 3D correspondente para determinar a localização de objetos em 3D. O *framework* inicia com base em métodos de detecção [295], [296], [297], os quais identificam objetos em imagens por meio de caixas delimitadoras. Com base nestas caixas delimitadoras, explora-se a compatibilidade entre hipóteses dos objetos dentro da caixa e o mapa 3D associado aos *pixels* dentro da caixa. Estas hipóteses de objetos, denominadas Hypotheses object Foreground Masks (HFM), são geradas a partir da hipótese plano de fundo e plano principal de segmentação do objeto dentro da caixa delimitadora. Os modelos de objetos são aprendidos utilizando uma formulação de margem-máxima latente, através da qual as variáveis latentes são as localizações das partes do objeto em 3D (Latent SVM [298]). As características são extraídas de pistas de aparências dentro da HFM e os descritores 3D são computados na nuvem de pontos associada. Os custos de deformação, ou custos de penalização, para a distância relativa entre as partes do objeto e as posições raiz do objeto, são calculados no espaço 3D, onde uma estratégia de correspondência é utilizada. Este procedimento de correspondência 3D envolve os seguintes passos: projeção dos mapas de resposta dos filtros na nuvem de pontos 3D pela associação de um valor de confiança do pixel ao seu ponto 3D correspondente. Então, define-se uma função de pontuação a qual é obtida como a soma das repostas da raiz e da parte, com relação aos seus

custos de deformação. Esta pontuação fornece o maior valor na sua localização máxima. Uma vez obtida a localização da raiz do objeto em 3D, a localização das partes do objeto podem ser encontradas olhando o deslocamento ótimo, similar ao caso 2D [295]. O *framework* proposto foi avaliado em duas bases de dados, Washington RGBD (WRGBD) [299] e Berkeley 3D Object (B3DO) [300], e os resultados comparados com outros métodos (DPM com diversas combinações e com dois métodos propostos em [300]) em termos de precisão média.

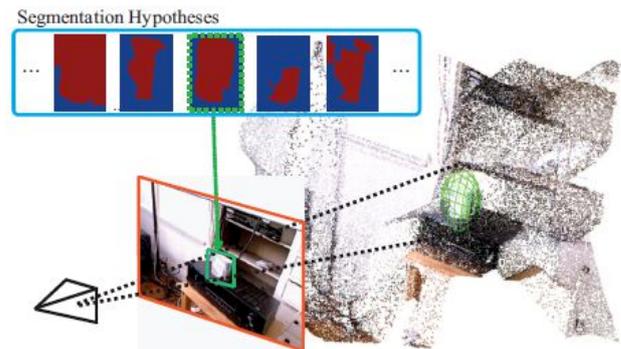


Fig. 89. Representação do método proposto por Byung-soo Kim. Figura extraída de [294].

Uma versão melhorada do descritor spin image (SI)[16], a qual aprimora a performance de discriminação do formato, é proposta por Choi [301]. A versão aprimorada, denominada angular-partitioned spin images (APSI), melhora o poder discriminativo e a acurácia na detecção através da geração de sub-spin images para espaços cilíndricos azimuthally particionados. Com base neste espaço, características angulares únicas das superfícies locais podem ser codificadas como APSI. Como resultado, diferentes superfícies da face com diferenças sutis podem ser distinguidas com êxito. Experimentos foram conduzidos com dados de face 3D da base de dados GavabDB [302]. Primeiro, SI *templates* para pontos de referência da face foram gerados. Para a representação APSI, *templates* para os pontos de referência da face, com exceção da ponta do nariz, foram produzidos. Um SI consistindo de 16 x 16 bins em ambos os intervalos, vertical e horizontal, de cada bin (tamanho do bin) foram setados para 2 mm. Como resultado pontos dentro do espaço cilíndrico, os quais a largura e altura são 3.2 cm são codificados em cada SI. Para o APSI, o histograma 3D formado de 8 x 8 x 4 bins, e tamanho de bin foi setado para 4 x 4 mm afim de ter o mesmo número de dimensões em ambos os descritores. Para medir a similaridade entre os descritores, utilizou-se normalised linear correlation (NLC). Para a avaliação da performance dos descritores SI e do APSI, descritor proposto, um ponto tendo o maior NLC foi escolhido como ponto de referência e o erro de distância entre a posição detectada como ponto de referência e o ponto de referência manualmente anotado foi calculado, mostrando o valor de erro médio e o desvio padrão alcançado para o SI e para APSI.

Palossi [303] explora a otimização de descritores locais 3D através da programação voltada para GPU. Esta otimização por GPU tem como objetivo alcançar processamento em tempo

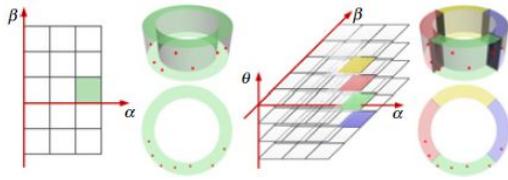


Fig. 90. Representação do método proposto por Choi. Figura extraída de [301].

real em conjuntos de dados 3D típicos de descritores 3D. Para alcançar este objetivo, o descritor SHOT [150] foi dividido em várias passos elementares, analisando os requerimentos de tempo de computação de cada passos e propondo uma otimização por GPU para cada um deles. Inicialmente, foi realizado uma análise precisa, afim de determinar quais operações elementares envolvidas na computação do descritor são computacionalmente relevantes e adequadas para paralelização. Depois deste perfil e identificação das partes elementares relevantes, propõe-se uma otimização em GPU para cada uma das seguintes etapas: Computação da normal, Radius Neighbor Search, computação do Local Reference Frame, computação do histograma, Interpolação Quadrilinear e normalização. Para os experimentos, testou-se a implementação paralela contra a implementação sequencial do SHOT comparando-as em termos de tempo de execução para cada etapa. Depois, realizou-se um experimento adicional com o GPU-SHOT em um *pipeline* para reconhecimento de objetos em imagens desordenadas [304], demonstrando os resultados em termos de curvas de precisão e lembrança.

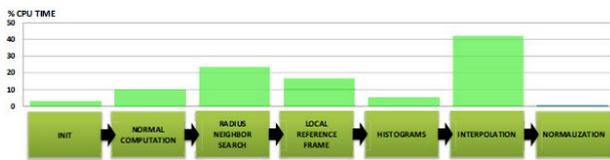


Fig. 91. Representação do método proposto por Palossi. Figura extraída de [303].

Uma abordagem para o reconhecimento de faces 3D com base nos contornos frontais da propagação do calor sob a superfície da face é proposto por Abdelrahman [305]. Os contornos frontais são extraídos automaticamente à medida que o calor se propaga (*heat kernel*) a partir de um conjunto pontos de referência detectados automaticamente, utilizando para a detecção dos pontos de referência o método proposto em [306]. A abordagem codifica as características locais da face bem como a distância de difusão sobre a superfície em torno destes pontos de referência. As propagações dos contornos são utilizadas para discriminar com êxito as várias faces. Após calcular os *heat kernels* em cada ponto, os contornos 3D são extraídos no ponto 3D da superfície da face que possui os mesmos valores de *heat*. Utiliza-se um número pré-determinado de contornos ao redor de cada ponto. Então, cada contorno é amostrado com um número fixo de pontos. Esta representação fornece um conjunto ordenado e finito de pontos 3D por face. Para a correspondência entre duas faces, utiliza-se o ICP para estimar parâmetros rígidos de

transformação entre o conjunto de pontos correspondentes aos contornos encontrados para as duas faces. A distância L2-norm entre os pontos de contorno da face adquirida e das faces da galeria, após o registro, é usada como medida de distância e as faces da galeria são ranqueadas com base nesta medida de distância. A abordagem proposta é avaliada na base de dados Texas 3D Face Recognition Database [307] e comparada com outros métodos de reconhecimento 3D de faces (eigensurfaces [67], fishersurfaces [308], ICP algorithms [309], e o algoritmo anthropface 3D em duas versões [307]) em termos de taxa de reconhecimento.

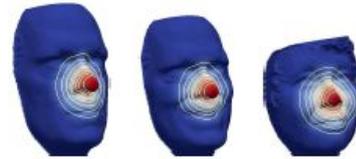


Fig. 92. Representação do método proposto por Abdelrahman. Figura extraída de [305].

Jincheng Yu [310] apresenta um sistema robótico com base em visão, que pode não apenas reconhecer diferentes objetos, mas também estimar sua pose através do uso de um modelo de *deep learning*. O modelo de *deep learning* utilizado é o Max-pooling Convolutional Neural Network (MPCNN). O autor afirma que que o modelo de *deep learning* não funciona tão bem para detecção do objeto. E propõe, para contornar esse problema, o uso de um método de detecção do objeto para segmentar o objeto do plano de fundo. Logo o fluxo de dados para o sistema robótico inicia pela aquisição por uma câmera seguido da detecção dos objetos por um método de aprendizagem com base em dicionário. Então, utiliza-se o MPCNN para reconhecer os objetos e estimar a suas poses. Por fim, o controlador do robô move o braço robótico, pega o objeto e o move para uma posição pré-destinada. Cada etapa será resumidamente explicada a seguir. A primeira etapa é a segmentação do objeto do plano de fundo, na qual utiliza-se um método de aprendizagem com base em dicionário. O dicionário é construído utilizando uma clusterização por K-means para as componentes de cor das imagens com o plano de fundo e o objeto. No caso, utiliza-se o K-means com distância euclidiana para agrupar os *pixels* das imagens de 5 objetos, 30 imagens por objeto, em m *clusters*. Depois, com esse dicionário, pode-se pegar as características chave pela comparação das estatísticas de projeção do plano de fundo e do plano principal. Empiricamente o número de *clusters* foi setado para 30. Após a extração das características, utiliza-se o MPCNN para estimar a pose e reconhecer os objetos. Métodos por *deep learning* possuem a capacidade de reconhecer ou prever um grande conjunto de características pelo aprendizado de características esparsas de um pequeno conjunto de padrões. Com essa vantagem pode-se utilizar um pequeno conjunto de poses para treinar o modelo de *deep learning* e então utilizá-lo para prever um conjunto maior de poses. No caso atribuiu-se diferentes poses dos objetos como diferentes classes. Conseqüentemente, uma classe representa um objeto no MPCNN. Por exemplo, as poses de uma caixa

são divididas em 6 classes de pose cada 30 graus, e numeradas classe 1 até 6. Estas 6 classes correspondem aos primeiros 6 neurônios da camada de saída no MPCNN. Deste modo, pode-se não apenas reconhecer o objeto como também a pose pela saída dos neurônios na última camada. Para os testes com o sistema proposto, construiu-se uma base de dados com 5 objetos capturados em várias poses, distâncias e iluminações. A parte de reconhecimento de pose e do objeto do sistema proposto foi avaliado em termos de taxa de reconhecimento.

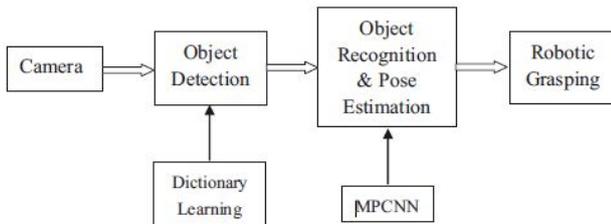


Fig. 93. Representação do método proposto por Jincheng Yu. Figura extraída de [310].

Joseph Lam [311] apresenta um sistema para reconhecimento de objetos 3D com base em registro de segmentos. Este sistema possui dois estágios (offline e online). No estágio offline, realiza-se o treinamento, no qual modelos são pré-processados em segmentos de interesse e em pares de pontos quantizados. No estágio online três fases são executadas: extração do segmento de interesse, registro do segmento e verificação por reprojeção. Na extração do segmento de interesse o algoritmo utilizado pode ser dividido em 3 partes. Primeiro, pontos de interesse são extraídos por meio de operadores como *difference of normal operator*. Então, a linha RANSAC é utilizada para estimar os limites dos segmentos como curvas lineares por pedaços e por fim, o completo segmento de interesse é finalmente extraído utilizando um algoritmo de crescimento por regiões. Na fase de registro do segmento, utiliza-se um método de emparelhamento tal como o 4PCS [312], o qual dado segmentos do modelo e da cena realiza uma comparação por pares, onde cada comparação resultará em uma pose candidata. Por fim na parte de verificação por reprojeção cada pose candidata é verificada através da reprojeção dos pontos do modelo na cena. De tal forma que se as instâncias do modelo na cena não estiverem fortemente oclusas, então a qualidade da correspondência que fornece a maior pontuação de sobreposição será naturalmente selecionada como a melhor estimativa de pose. O método foi testado utilizando nuvem de pontos adquiridas de duas formas, Lidar e Kinect, e os resultados obtidos foram avaliados em termos de taxa de reconhecimento.

Ekekrantz [313] propõe um método, denominado Adaptive Iterative Closest Keypoint (AICK), para registro de dados RGB-D, baseando-se no princípio do algoritmo Iterative Closest Point (ICP). A entrada do AICK é um conjunto de pontos chave invariantes detectados em cada *frame* RGB-D. Cada ponto chave é associado com uma posição 3D em um Local Reference Frame (LRF) e um descritor de característica. Os pontos chave são geralmente selecionados de um grande conjunto de pontos possíveis. O AICK requer que estes pontos

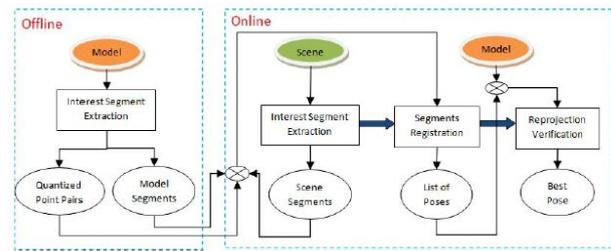


Fig. 94. Representação do método proposto por Lam. Figura extraída de [311].

chave sejam estáveis, o que significa que o mesmo ponto chave deve ser detectado como ponto chave em vários *frames* consecutivos. O algoritmo AICK não faz nenhuma suposição sobre os algoritmos de detecção de pontos chave e descrição empregados, no entanto é desejável que o detector de pontos chave seja invariante a rotação e escala. Para tanto, foram avaliados os pontos chave do SURF [143] e ORB [314]. No ICP original utiliza-se a distância euclidiana associada ao pares de pontos entre duas nuvens. No AICK, substitui-se a distância euclidiana pela soma ponderada da distância euclidiana e a distância entre dois descritores de pontos chave, onde esta distância é calculada *squared L2 norm* da diferença em vetores descritores dimensionados por uma constante dependendo do tipo de recurso descritor usado. Introduzindo uma medida ponderada é de importância fundamental para a adaptabilidade do algoritmo. Por escolha dos autores, o fator de distância euclidiana é completamente negligenciado para a correspondência inicial. Como resultado, o algoritmo não requer um palpite inicial para a transformação entre as nuvens de pontos. Começando da segunda iteração, permitem-se correspondências as quais são geometricamente próximas, mas não tão próximas em aparência. Esse processo continua até que não seja mais considerado a aparência. Neste ponto o ICP original é realizado para um registro mais fino. Para fazer o algoritmo mais eficiente, utiliza-se um método de aprendizagem de vocabulário de palavras como no método de *bag of words*. A aprendizagem é essencialmente uma *clusterização* dos descritores, de todas as imagens de treinamento, em um pré-determinado número de *clusters*. Logo, as palavras do vocabulário consistem da média dos descritores para cada *cluster*. Então, associa-se a cada ponto chave uma lista de suas palavras mais próximas. Quando procura-se pela correspondência mais próxima para um ponto chave em particular em um *frame*, utiliza-se apenas a palavra mais próxima. Em seguida, procura-se o outro *frame* para todas as ocorrências daquela mesma palavra e compara-se com o ponto chave correspondente. O resultado é que ao invés de ter que corresponder todos os pontos com todos os pontos, tem-se apenas que corresponder cada ponto com um pequeno subconjunto de pontos do outro *frame*. Nos experimentos, utilizou-se um conjunto de dados público [315] e um estabelecido procedimento de avaliação e medida de performance[315], comparando o algoritmo proposto com outros métodos (generalized ICP (GICP) [316], 3D normal distribution transform (3D-NDT) [317] e um método com base em RANSAC[318], nomeado pelo autor como 3-

point RANSAC).

O trabalho apresentado por Gomes [319] propõe o uso de uma abordagem de fovea móvel para diminuir a amostragem dos dados 3D e reduzir o processamento do sistema de recuperação do objeto de uma nuvem de pontos. O modelo de nuvem de pontos com fovea móvel diminui a amostragem da nuvem de pontos original através do uso de caixas concêntricas, cada uma representando um nível. Cada caixa representa um corte da nuvem de pontos com uma diferente densidade da nuvem de pontos. A caixa externa tem o seu canto posicionado em uma coordenada 3D específica, definindo o sistema de coordenadas. A menor caixa é centralizada no parâmetro denominado fovea, uma coordenada 3D onde a densidade da nuvem de pontos é máxima, e um esquema de diminuição de amostragem é aplicado nas caixas reduzindo a densidade da nuvem. Esta abordagem está associada a um sistema de reconhecimento de objetos com base em pontos chave. O sistema como um todo pode ser descrito nos seguintes passos: Inicialmente, aplica-se a abordagem de fovea móvel para diminuir a densidade da nuvem de pontos. Depois, ocorre a computação dos descritores dos pontos chave, Signature of Histograms of Orientations (SHOT), utilizando um algoritmo de redução de amostragem adaptado para utilizar a nuvem de pontos foveada e calculando o Local Reference Frame (LRF) para cada ponto chave, por meio do algoritmo de Petrelli and Di Stefano [320], tanto para o objeto modelo quanto para a cena. Por fim, para cada ponto chave, uma correspondência na cena é procurado através da procura do ponto mais próximo na cena, no espaço n -dimensional, contendo o descritor SHOT. O procedimento de procura é realizado em uma árvore-KD e se o quadrado da distância entre os descritores SHOT do modelo e da cena for menor que um limiar, uma correspondência de pontos chave é estabelecida. Como cada ponto chave tem um LRF associado, pode-se estimar a transformação associada a cada correspondência entre pontos chave. Consequentemente, um 3D Hough Space é utilizado para juntar evidências da presença do objeto através de um processo de voto. Os resultados obtidos para o método proposto foram avaliados com base nas taxas de verdadeiros positivos e falsos positivos e no tempo de execução.

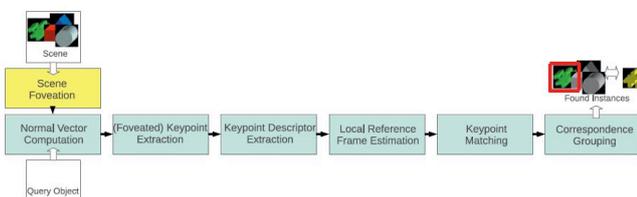


Fig. 95. Representação do método proposto por Gomes. Figura extraída de [319].

Yulan Guo [321] apresenta algoritmos de reconhecimento de objetos 3D que utilizam não apenas informação de formato, mas também informações de cor. Primeiro, estende-se o descritor apresentado anteriormente pelo autor, Shape only Rotational Projection Statistics (S-RoPS) [322], para obter o descritor de característica RoPS com cor apenas (C-RoPS). Se

propõe também um algoritmo para reconhecimento de objetos 3D com base na fusão do S-RoPS e do C-RoPS. Para gerar o C-RoPS utiliza-se a mesma estrutura do S-RoPS substituindo as informações espaciais pelas informações de cor. Dado uma superfície local, primeiro constrói-se um Local Reference Frame (LRF) da superfície utilizando a mesma abordagem utilizada no S-RoPS. Então, utiliza-se os parâmetros de cor para substituir as coordenadas e segue-se uma abordagem similar ao do S-RoPS para gerar o C-RoPS. O algoritmo de reconhecimento inclui quatro módulos: representação do modelo, geração do modelo candidato, geração da transformação hipóteses, verificação e segmentação [289], [323]. No módulo de representação do modelo, pontos de características são selecionados uniformemente de cada modelo. Então, extrai-se o S-RoPS e o C-RoPS em cada ponto de característica, armazenando-os na biblioteca para correspondência de características. No módulo de geração de modelo candidato, inicialmente, selecionam-se pontos de característica da cena de entrada. Então, extrai-se os descritores de forma similar ao módulo anterior e corresponde-se todas as características da cena com todas as características do modelo para gerar um conjunto de correspondências. Estas correspondências são utilizadas para votar para um modelo candidato. No módulo de geração de hipóteses de transformação, calcula-se uma estimativa de transformação para cada correspondência de características. Estas estimativas de transformação são então *clusterizadas* em um pequeno número de hipóteses de transformação. No módulo de verificação e segmentação, alinha-se os modelos candidatos à cena de entrada utilizando as hipóteses de transformação. O processo de alinhamento produz uma pontuação de confiança para cada modelo. Se um modelo candidato alinha bem com uma porção da cena, o modelo candidato e a hipótese de transformação são aceitas. Caso contrário, são rejeitadas. Para a parte de fusão das características espaciais e de cor, S-Rops e C-RoPS, duas estratégias de fusão são adotadas: fusão a nível de característica, a qual concatena o S-RoPS e o C-RoPS em cada ponto de característica para gerar uma característica combinada, e fusão a nível de decisão, na qual primeiro realiza-se o reconhecimento do objeto utilizando o S-RoPS e o C-RoPS separadamente, resultado em uma pontuação de confiança para cada modelo para cada um dos dois métodos. E então o método de fusão seleciona a pontuação de confiança de maior valor entre todos os resultados dos dois modelos separados. Os algoritmos propostos foram testados em duas base de dados, Bologna Dataset III [150] e Bologna Dataset VI [324], e os resultados foram demonstrados em termos de taxa de reconhecimento em relação ao número de pontos de características.

Shuang Ma [325] apresenta o desenvolvimento de um manipulador móvel e customizado para a estimativa de pose e aperto de trava de torção. Para a parte de percepção, propõe-se uma abordagem para o reconhecimento de objetos 3D utilizando Kernel Principal Component Analysis (KPCA) com base apenas nas informações de profundidade. Com base nos últimos desenvolvimentos em aprendizagem de características com base em kernel [326], desenvolveu-se um conjunto de *kernel principal features* em imagens de profundidade, para o reconhecimento do tipo de trava de torção e pose por

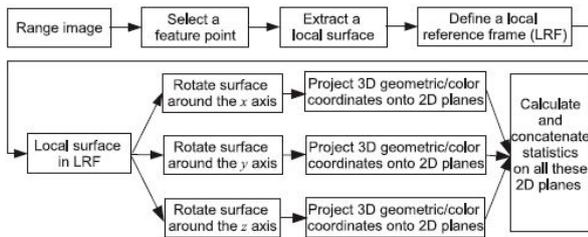


Fig. 96. Representação do método proposto por Yulan Guo. Figura extraída de [321].

meio do KPCA de acordo com a distância do vizinho mais próximo de forma hierárquica [327]. O processo de percepção está dividido em duas partes: análise offline, composta de amostragem dos dados, extração de características e treinamento com base nas características extraídas; Processamento em tempo real: com base no treinamento de características obtido na parte anterior, detecta-se a trava de torção, identifica-se o tipo e a pose. Na parte de detecção do objeto e extração de características utiliza-se uma combinação de técnicas para remover o objeto do plano de fundo e extrair características do objeto. Inicialmente, aplica-se o método de subtração do plano de fundo para remover o objeto do plano de fundo com base em um limiar de diferença para definir a área alvo. Depois, aplica-se um filtro mediano para remover o ruído seguido de um operador *sobel* para a detecção da borda. A posição do objeto pode ser estimada de forma grosseira após a detecção da borda. Após localizar o objeto, um conjunto de *kernel principal features* das imagens de profundidade é utilizado para a descrição e reconhecimento dos objetos 3D. Nos experimentos são distinguidos a categoria do objeto e a pose entre três tipos de objeto em três poses diferentes. Uma parte do conjunto de imagens foi utilizada para o treinamento, construção da matriz kernel e seleção de parâmetros. Após a computação da matriz de projeção na parte de treinamento, todas as imagens foram projetadas para o espaço de características e o reconhecimento foi realizado com base em um classificador de coeficiente de correlação, demonstrando os resultados com base na acurácia alcançada.

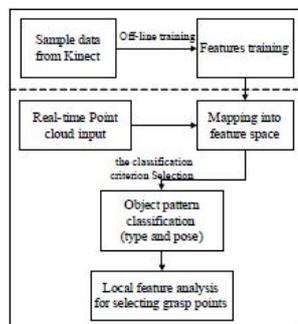


Fig. 97. Representação do método proposto por Shuang Ma. Figura extraída de [325].

Um método para classificação de objetos 3D com base em *Local Keywords* e Hidden Markov Model (HMM) é apresentado por Guo Jing [328]. Na abordagem proposta um

vetor de características geométricas com base na *Relative-Angle Context Distribution* (RAC) dos pontos da superfície é extraído. As *Local Keywords* são geradas a partir de *clusters* de histogramas de RAC. Para a parte de clusterização dos histogramas de RAC (HRAC) utiliza-se o algoritmo k-means. Então, cada objeto é separado pelo modelo combinado e em cada *bin* pode-se adquirir uma *Local Keyword*. As *Local Keyword* são caracterizadas pelo centro dos *cluster*. O *cluster* pode ser associado também a algum tipo de significado semântico, como por exemplo, a cabeça do tigre ou a asa da abelha, etc, cada *cluster* representa uma única *Local Keyword*. No processo de classificação um HMM de primeira ordem foi treinado para cada classe de objeto. Os experimentos realizados com o método proposto utilizaram a base de dados Princeton Shape Benchmark [39], comparando os resultados obtidos com outros dois métodos, ADHMM[329] e SD[73] em termos de acurácia na classificação.

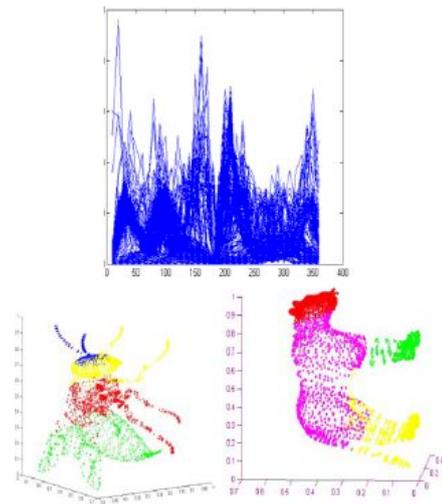


Fig. 98. Representação do método proposto por Guo Jing. Figura extraída de [328].

Um paradigma para localização e mapeamento em tempo real, o qual utiliza reconhecimento de objetos 3D para pular sobre o processamento de geometria de baixo nível e produzir mapas incrementalmente construídos diretamente em um nível orientado a objetos é proposto por Moreno [330]. Este paradigma denominado SLAM++ pode ser descrito pelos seguintes passos: inicialmente cria-se uma base de dados de objetos, de modo offline, com um escaneamento utilizando o KinectFusion em um ambiente controlado, extraíndo uma malha de pontos por meio do algoritmo de Marching Cubes. Depois, utiliza-se uma representação do mundo com grafo, onde cada nodo armazena ou a pose estimada do objeto ou o histórico da pose da câmera em um determinado estante de tempo. Cada nodo de objeto é anotado com um tipo de um objeto da base de dados. Cada medida de pose de um objeto de uma câmera é armazenado no grafo como um fator o qual liga uma pose de câmera e uma pose de objeto. Para o reconhecimento da pose do objeto, o 6DoF do objeto 3D, utiliza-se uma abordagem similar a utilizada em Drost [136], a qual utiliza a representação de malha dos objetos

para, em uma imagem de profundidade, definir a pose do objeto. Na abordagem de Drost, um objeto é detectado e simultaneamente localizado através do acúmulo de votos no espaço de parâmetros. A base de votação é a correspondência entre Point-Pair Features (PPFs), descritores 4-dimensional da posição relativa e das normais dos pares dos pontos orientados na superfície do objeto. Pontos com a normal estimada, são amostrados aleatoriamente em uma imagem bilateral filtrada da câmera de profundidade. Estas amostras são pareadas em todas as possíveis combinações para gerar PPFs os quais votam para uma configuração de um modelo 6DoF contendo PPF similares. Por fim, a correspondência entre os objetos da cena e os objetos do da base de dados de modelos pode ser eficientemente realizada em paralelo via uma busca binária vetorizada, produzindo um voto para cada correspondência. O resultado da detecção dos objetos é frequentemente múltiplos picos de *cluster* e localizações quantificadas. Estes devem passar por um refinamento por Iterative Closest Point (ICP), o qual habilita a estimativa da câmera ao vivo. Para os testes, foi realizado o mapeamento de um ambiente, mostrando as estatísticas do processo de mapeamento (número de objetos, memória utilizada, número de classes entre outros).

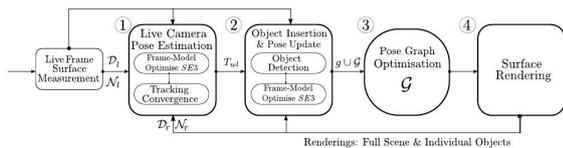


Fig. 99. Representação do método proposto por Moreno. Figura extraída de [330].

Guan Pang [331] descreve um método de reconhecimento de objetos 3D que combina procedimentos de aprendizagem de máquina com características 3D locais. O método proposto é dividido em dois módulos, treinamento e detecção. No módulo de treinamento, treina-se um detector para cada classe de objeto, utilizando o procedimento de treinamento do Adaboost [332], com amostras de treinamento geradas de uma biblioteca de objetos pré-rotulados. Os detectores escaneiam e avaliam exaustivamente a nuvem de pontos, retornando o máximo de respostas positivas como posições de objetos detectados. Os detectores de objetos consistem em N classificadores fracos, treinados com base em características 3D da imagem (*3D Haar feature* [333]), cada um com um peso. Cada classificador fraco avalia um subconjunto da região candidata e retorna uma decisão binária. O detector do objeto, ou classificador forte, é uma combinação de todos os pesos dos classificadores fracos, a qual é comparada com um limiar pré-determinado para determinar se a região candidata é uma possível correspondência. A entrada do módulo de detecção é uma região da nuvem de pontos 3D, a qual é pré-processada em uma tabela de área 3D somada para uma computação mais eficiente. Uma janela 3D de detecção é movida para procurar através da imagem 3D, avaliando a correspondência entre cada *cluster* de nuvens de pontos dentro da janela de detecção e o objeto alvo com o qual o detector foi treinado durante a fase de treinamento. Após o escaneamento exaustivo da nuvem de pontos de entrada, todas as instâncias positivas detectadas são processadas através de

non-maximum suppression para identificar o objeto alvo com a melhor correspondência e confiança acima de um limiar. Para demonstrar a performance do método proposto foram utilizados dados de engenharia, de ruas e de uma base de dados pública para o reconhecimento de objetos 3D [81]. Os resultados obtidos foram demonstrados em termos de precisão e lembrança e foram comparados com outros métodos em termos de taxa de reconhecimento.

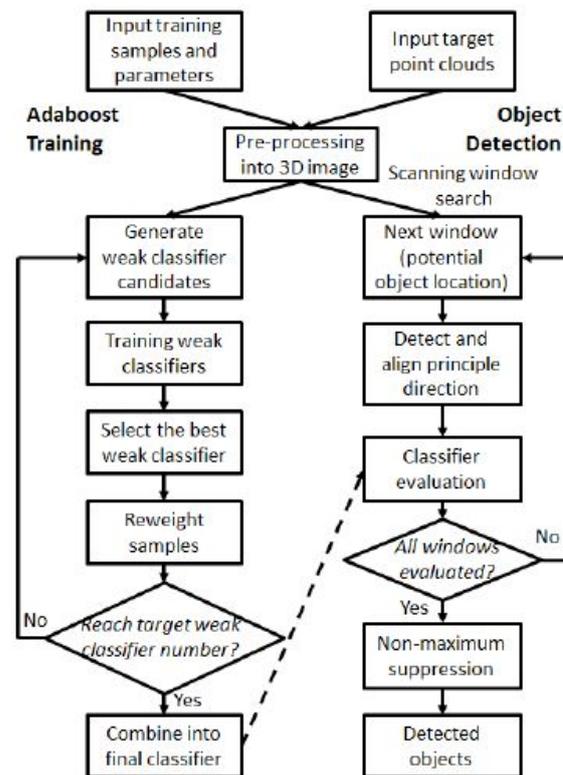


Fig. 100. Representação do método proposto por Guan Pang. Figura extraída de [331].

Nakashika [334], apresenta uma abordagem para o reconhecimento de objetos 3D em imagens RGB-D. Primeiro, a imagem de profundidade e as imagens RGB de cada canal são montadas a partir das informações de cor e profundidade. Depois, pontos de características de cada imagem são localizados por uma amostragem em grade e características são extraídas de cada ponto de característica, o Histograms of Oriented Normal Vectors (HONV) é extraído da imagem de profundidade e o scale-invariant feature transform SIFT é extraído das imagens RGB. As características extraídas são codificadas utilizando o Locality-constrained Linear Coding (LLC), o qual utiliza restrições locais para projetar cada descritor em seus sistemas de coordenadas locais. Então, múltiplos códigos dentro de cada sub-região da *spatial pyramid* (o método de *spatial pyramid* particiona a imagem em sub-regiões espaciais hierárquicas e computa o histograma de características locais para cada sub-região.) são agrupados. O agrupamento da *spatial pyramid* de profundidade é processada adicionalmente para a imagem de profundidade. Por fim, as características agrupadas de todas as sub-regiões são concate-

nadas juntas para classificação. O classificador é treinado, Support Vector machine (SVM) foi o classificador utilizado, por este vetor concatenado das imagens de treinamento, sendo posteriormente utilizado para o reconhecimento de objetos. Para os experimentos foi utilizada a base de dados RGB-D Object Dataset [299], a qual é composta de mais de 300 objetos em 51 categorias com aproximadamente 42000 imagens contendo informações de cor e profundidade. A performance da abordagem proposta foi demonstrada em termos de acurácia média e desvio padrão em 10 testes randômicos para o reconhecimento da categoria do objeto.

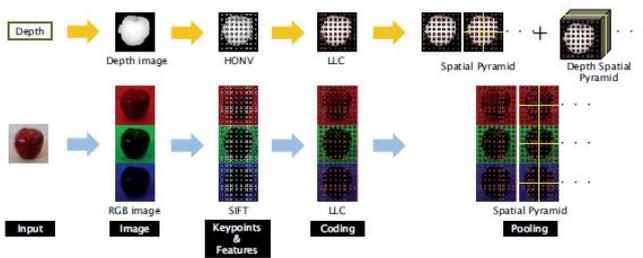


Fig. 101. Representação do método proposto por Nakashika. Figura extraída de [334].

Takei [335] propõe um método para a localização e reconhecimento de pose de objetos 3D em cenas com desordem. Este método é um método de seleção de características com base na avaliação da performance de discriminação em um espaço de característica. A abordagem proposta extrai características da cena utilizando 3D Computer Graphics e utiliza as características extraídas para alcançar uma melhor performance discriminatória. Em um esforço para melhorar a performance discriminatória, o método Diverse Density [336] avalia as características pela utilização da distância em um espaço de características. A eficácia das características é avaliada utilizando um conjunto de amostras positivas e a separabilidade das amostras negativas. Este método pode procurar por pontos correspondentes pelo simples cálculo da norma no espaço de características no processo de correspondência. Portanto, foi utilizada a ideia do Diverse Density como meio de permitir ao método proposto avaliar a performance discriminatória da característica. O método proposto utiliza pares de vetores com alta performance discriminatória para o processo de correspondência. A ideia tem como base o Vector Pair Matching (VPM) method [337]. A abordagem proposta, primeiro, sintetiza cenas incluindo objetos aleatoriamente empilhados e extrai *vector pairs* nas cenas. Depois, esses *vector pairs* são classificados como amostras positivas ou negativas pela utilização de um *vector pair* do objeto modelo. Amostras positivas são representadas como *vector pairs* que possuem características verdadeiras e amostras negativas como aqueles que possuem características falsas. O *vector pair* extraído é então avaliado pelo grau de concentração de amostras positivas e o grau de separabilidade de amostras negativas com base na distância no espaço de características. Por fim, a performance discriminatória é avaliada pela integração do grau de concentração e grau de separabilidade. O método proposto avalia a performance discriminatória de todos os

vector pairs no objeto modelo e seleciona *vector pairs* que possuem os maiores valores de avaliação. Para avaliar a performance de reconhecimento, utilizou-se quatro tipos de conjunto de dados, 3 cenas sintetizadas com a presença de um objeto modelo e uma cena real com a presença de um objeto modelo. O método proposto foi comparado com outros métodos de reconhecimento de objetos 3D (Correspondence Grouping Method [167] e VPM Method [337]) em termos de taxa de sucesso de reconhecimento e tempo de processamento.

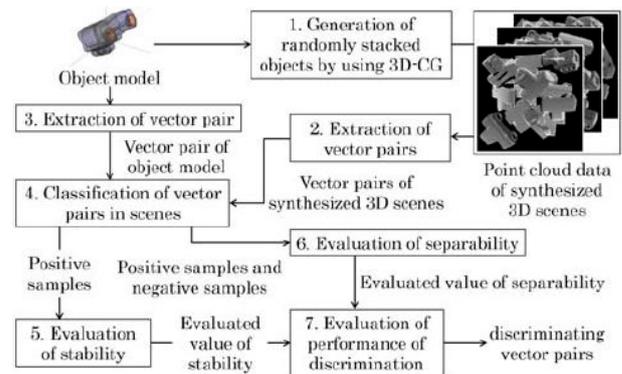


Fig. 102. Representação do método proposto por Takei. Figura extraída de [335].

Uma abordagem da viabilidade de um *codebook* para a classificação automática de ameaças em imagens pré-segmentadas de tomografia computadorizada de bagagens é investigado por Mounton [338]. Para tanto, seguiu-se o *framework* de classificação Bag of (Visual) Words (BoW) [339]. Com base neste *framework* comparou-se a performance de 5 modelos de *codebook* utilizando uma variedade de combinações de estratégias de amostragem (escassas, por meio do detector de pontos de interesse 3D SIFT, e densas, como recomendado em [339]), técnicas de codificação de características (k-means clustering e Extremely Randomised Clustering (ERC) forests) e classificadores (Support Vector Machine e Random forests) com a abordagem utilizando *visual cortex* [340]. Demonstrou-se uma melhoria da abordagem *visual cortex* em termos de acurácia, assim como em tempo de processamento, ao utilizar um *codebook* construído via Extremely Randomised Clustering (ERC) forests [341] em conjunto com uma estratégia de amostragem de características densa e o classificador SVM. A combinação de técnicas propostas foram avaliadas no contexto de classificação de dois objetos alvos em imagens de tomográfica computadorizada de bagagens. Os objetos foram escaneados em poses randômicas para obter invariância rotacional e foram isolados manualmente para a extração a priori de características. Todos os demais objetos foram considerados como parte do meio desordenado. Um teste de validação cruzada de 10 vezes foi realizado utilizando dados e cortes de dados iguais aos realizados em [340], permitindo uma comparação direta da performance entre os métodos.

Xiangfei Qian [342] propõe um método de reconhecimento de objetos 3D, o qual segmenta um conjunto de pontos 3D em um número de segmentos planares e extrai a Inter-Plane

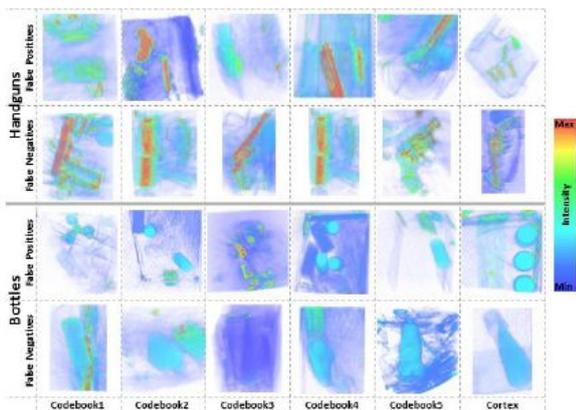


Fig. 103. Representação do método proposto por Mouton. Figura extraída de [338].

Relationships (IPRs) para todos os segmentos. Com base nos IPRs, o método determina o High Level Feature (HLF) para cada segmento. Então, um classificador de planos com base em Gaussian-Mixture-Model (GMM) é utilizado para classificar cada segmento em um plano pertencente a um certo objeto modelo. Por fim, um procedimento recursivo de clusterização de planos é realizado para clusterizar os planos classificados no modelo do objeto. O método de reconhecimento consiste de 5 procedimentos principais: Aquisição dos dados 3D de profundidade, extração dos planos dos dados de profundidade, extração de características, projeto e treinamento de um classificador de planos GMM e clusterização dos planos. No primeiro procedimento, os dados de profundidade são capturados em várias posições da câmera. Em cada posição, o sistema de estimativa de pose computa a pose da câmera. Utilizando a informação de pose, os dados de profundidade são registrados para formar uma larga visão da cena, na forma de nuvem de pontos 3D. Os dados 3D são então segmentados em N segmentos planares pelo método NCC-RANSAC [343]. Para cada um dos N segmentos um vetor de características é construído com base em três classes de categorias, Basic Feature (BF), Low Level Feature (LLF), e High Level Feature (HLF). As BFs são características locais que descrevem atributos intrínsecos dos segmentos. As LLFs são matrizes formadas com base em 9 Inter-Plane Relationship definidas para um segmento em referência a outro, estabelecendo a relação entre os planos. São elas: *plane-distance*, *plane-angle*, *projection-distance*, *is-parallel*, *is-perpendicular*, *is-coplanar* e *is-adjacent*. Como o objetivo do reconhecimento do objeto é identificar um entre oito objetos de uma cena, cada segmento planar é classificado a um plano pertencente a um dos oito objetos (modelos). Para isto, definem-se seis mutuamente exclusivas High Level Features (HLFs), as quais cada HLF representa um conjunto de IRPs que existem em um modelo. A extração da HLF é um processo de identificar a HLF para cada um dos N segmentos planares e atribuir um vetor HLF para cada segmento. Após a atribuição de cada vetor HLF, os vetores são submetidos para o Classificador de planos GMM. O classificador de planos GMM consiste de 8 GMMs cada uma das quais deve ser treinada utilizando os dados capturados para

um tipo particular de objetos e é assim capaz de identificar um plano relacionado com esse tipo de objecto quando o vector HLF relevante é apresentado. Os experimentos com o método proposto foram realizados com um conjunto de dados 500 com diferentes poses da câmera, utilizados para treinar o classificador GMM, e após o treinamento foram coletados 60 conjuntos de dados de cada tipo de cena para o processo de classificação. Os resultados foram demonstrados em termos de taxa de reconhecimento e tempo de processamento.

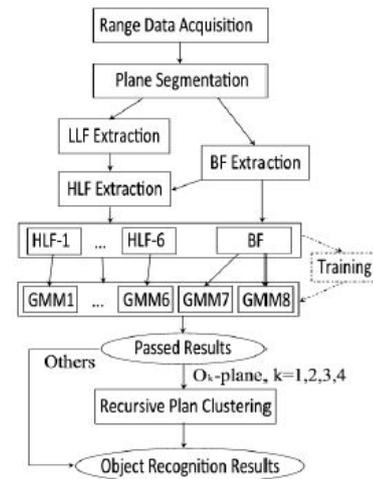


Fig. 104. Representação do método proposto por Xiangfei Qian. Figura extraída de [342].

Mateo [344] descreve um estudo analisando descritores 3D com base em normais da superfície para o reconhecimento de objetos 3D. Mais especificamente, avaliou-se o comportamento dos descritores no processo de reconhecimento utilizando modelos virtuais de objetos criados por CAD softwares e em cenas reais utilizando objetos criados com uma impressora 3D com base nos modelos virtuais. Em ambos os casos, os mesmos modelos virtuais foram utilizados no processo de correspondência para encontrar similaridade. A análise avaliou 4 aspectos: a eficácia do descritor 3D dependendo do ponto de visão da câmera, a complexidade geométrica do modelo e o tempo utilizado no processo de reconhecimento e a taxa de sucesso para reconhecer a visão do objeto entre os modelos armazenados na base de dados. Os descritores analisados foram 5: PFH (Point Feature Histogram [345]), FPFH (Fast Point Feature Histogram[137]), VFH (Viewpoint Feature Histogram [145]), CVFH (Clustered Viewpoint Feature Histogram [205]) e SHOT (Signature of Histograms of Orientations [150]). Os descritores foram sempre computados sobre uma malha de uma nuvem de pontos e eles incluem apenas informações geométricas com base no formato da superfície local sem incluir outro tipo de informação visual. O autor fez um breve resumo de cada um dos descritores, apresentou as 4 medidas de similaridade, com base em distância (Manhattan, Euclidean, Chi-squared e Hellinger), utilizadas para associar o modelo CAD e a visão do objeto, e apresentou a parte experimental realizada. Nos experimentos, os descritores foram submetidos a experimentos com dados reais e sintéticos analisando os resultados com base em matrizes de distância, utilizando as

métricas de similaridade apresentadas com base na utilização de conjunto de modelo vs conjunto de modelo e conjunto de modelo vs conjunto de teste, com base no tempo de processamento e na acurácia alcançada.

Shang-Hung Lee[346] propõe um método de registro de formatos 3D com base em momentos. O método proposto é uma das etapas do sistema proposto de reconhecimento de ações 3D do corpo humano, o qual é composto por dois estágios, treinamento e teste. O estágio de treinamento é composta por 6 etapas que são: clusterização de formatos 3D através de registro, o qual gera um *codebook* de poses; Extração de características locais, a qual para cada ponto no formato da ação 3D, extraí-se um descritor de características e assim o formato é representado por um conjunto de descritores ponderados; Geração do *codebook* de características, o qual é utilizado para codificar todos os formatos com um histograma bag-of-words (BoW) [347]; Detecção das poses chave; Representação das poses em movimento, a qual é formada ao aplicar a técnica de registro de formato 3D. Desta aplicação, uma sequência dinâmica de formatos 3D é representada como uma sequência de poses chave através da correspondência de cada formato da ação 3D de entrada com o *codebook* de poses, para encontrar o *template* de pose mais próximo; Por fim, ocorre o treinamento do Suport Vector Machine com base nestas representações de poses chave em movimento. Para a parte de testes, extraí-se as características locais e classifica-se as ações 3D utilizando o classificador SVM previamente treinado. A performance do método proposto foi verificada em uma base de dados de gestos composta de 10 diferentes ações realizadas por 5 diferentes atores e os resultados de classificação foram demonstrados em termos de taxa de classificação nas dez diferentes classes existentes.

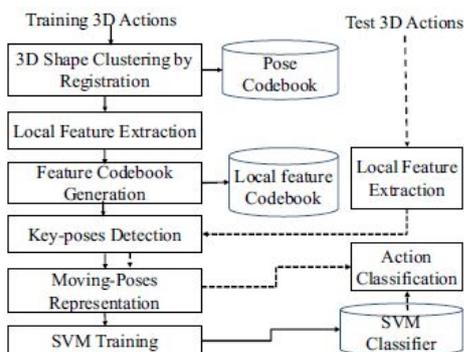


Fig. 105. Representação do método proposto por Yaniv Gur. Figura extraída de [346].

Yulan Guo [348] propõe um algoritmo para registro de imagens de profundidade de múltiplas visões. Neste algoritmo, primeiro extrai-se um conjunto de características Rotational Projection Statistics (RoPS)[323] de um par de imagens de profundidade e realiza-se a correspondência entre elas. As duas imagens de profundidade são então registradas utilizando um método de estimativa de transformação (Consistent Correspondences Verification (CCV) [323]) e uma variante do algoritmo Iterative closest Point (ICP)[349], para um registro mais refinado. Com base no algoritmo de registro aos pares,

propõe-se um algoritmo de registro por múltiplas visões com base em crescimento de formato. A semente do formato é inicializada com uma imagem de profundidade selecionada, a qual é sequencialmente atualizada pela realização de um registro aos pares entre ela e uma imagem de profundidade de entrada. Todas as imagens de profundidade de entrada são iterativamente registradas durante o processo de crescimento do formato. Uma vez que as malhas correspondentes a um formato em particular foram registradas grosseiramente, estes registros são refinados com um algoritmo de registro de múltiplas visões [350]. Por fim, um modelo 3D contínuo e sem emenda é reconstruído para cada formato através do uso de um algoritmo de integração e reconstrução de superfície [351]. Os experimentos realizados com o algoritmo proposto foram realizados na base de dados UWA 3D Modeling Dataset [81] testando inicialmente o registro aos pares e depois o registro com base em múltiplas visões. Nos testes do registro aos pares testaram-se diferentes métodos de estimativa de transformação (CCV, LSF, RANSAC-500, RANSAC-1000, RANSAC-5000, RANSAC-10000 e RANSAC [352]) e diferentes extratores de características (Spin images, LSP, THRIFT, USC, FPFH). Nos testes de registro múltiplas visões de um único objeto demonstrou-se a acurácia do algoritmo proposto em termos de porcentagem de imagens de profundidade registradas e do erro médio de registro para cada objeto. No último experimento, modelagem de objetos 3D, testou-se o *framework* em imagens de profundidade com alta (UWA 3D Modeling Dataset [81] e Stanford 3D Scanning Repository [351]) e baixa resolução (Bologna Reconstruction Dataset [150]) comparando os resultados dos modelos reconstruídos com o *Ground truth* do modelo.

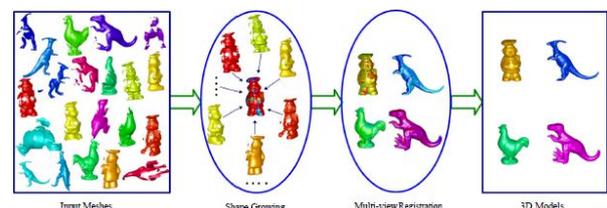


Fig. 106. Representação do método proposto por Yulan Guo. Figura extraída de [348].

Um versão comprimida do viewpoint feature histogram descriptor (VFH)[145] para classificação de objetos com base no reconhecimento do formato é proposto por Salih [353]. O descritor proposto emprega o *eigenvalue decomposition* para extrair a características de orientação dominante de uma nuvem de pontos e utilizá-las como descritores ao invés do histograma esparsa existente no VFH. O uso de características dominantes, mantém as propriedades principais do objeto com um descritor de comprimento mínimo que é aplicável para reconhecer a classe geométrica de objetos 3D. O processo de construção do descritor proposto inicia pela reamostragem dos pontos 3D da nuvem de pontos para ter o mesmo tamanho que os dados utilizados para o treinamento. Depois, 5 vetores de características são computados para a nuvem de pontos como descritores, os quais correspondem a três ângulos de orientação entre o centroide e cada ponto

da nuvem, a distância normalizada de cada ponto 3D em relação ao ponto centroide da nuvem de pontos e a direção de visão de cada ponto em relação a direção de visão média. Cada uma desses vetores de característica é subtraído utilizando o correspondente vetor médio computado durante a etapa de treinamento. O descritor comprimido é construído através da projeção do vetor de características médio subtraído em um espaço dimensional menor utilizando os autovetores que correspondem aos top r -autovalores. O número desses autovalores determina o tamanho do descritor. Se os top 5 autovalores forem selecionados, cada característica será representada com 5 *bins* e o descritor geral 25 bins de tamanho. O descritor proposto foi utilizado para construir um sistema de classificação utilizando o classificador Support Vector Machine. Os dados de classificação foram criados através da seleção de 1000 pontos da nuvem de pontos para cada objeto 3D com diferentes formatos. Cada um desses objetos contém pelo menos 1000 pontos e para cada nuvem de pontos, o descritor proposto foi calculado com tamanhos variados. Os resultados de classificação foram computados utilizando 3 métricas, precisão, lembrança e acurácia em dois experimentos, um utilizando apenas duas classes e outro utilizando várias classes.

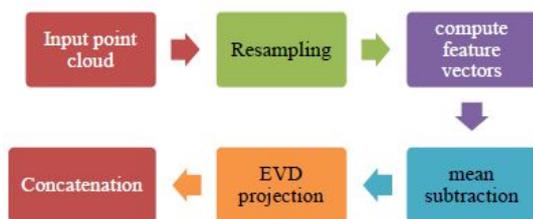


Fig. 107. Representação do método proposto por Salih. Figura extraída de [353].

Xun Yu [354] apresenta uma abordagem para o reconhecimento de faces utilizando 3D direcional corner points (3D DCPs), o qual combina a informação da conectividade estrutural com informações espaciais das faces 3D. Nesta abordagem, as superfícies 3D são representadas pelo 3D DCPs derivados das curvas de cume e vale. Após a detecção das curvas de vale e cume na superfície 3D, um processo de detecção de *corner point*, o qual tem como base o algoritmo de Douglas-Peucker [355], é aplicado para gerar os 3D DCPs. Então, após representar a superfície com o 3D DCP, desenvolveu-se um processo de conversão ponto-a-ponto para calcular a diferença entre dois 3D DCPs, da face no banco de dados de pesquisa e a face na base de dados da galeria. A dissimilaridade entre as duas faces é então calculada através de um processo de conversão global entre dois conjuntos de 3D DCP. Tanto a face a ser correspondida quanto a face na galeria de faces passam pelo mesmo processo de normalização, detecção de curvas de cumes e vales e descrição através do 3D DCP, os quais são por fim correspondidos. Para avaliar a performance do método para reconhecimento de faces, experimentos foram realizados na Grand Challenge v2.0 database (FRGC v2.0) [10] e os resultados foram avaliados em termos de resultado de reconhecimento, comparando-os com diferentes configurações

do 3D DCP e com os resultados obtidos em [356].

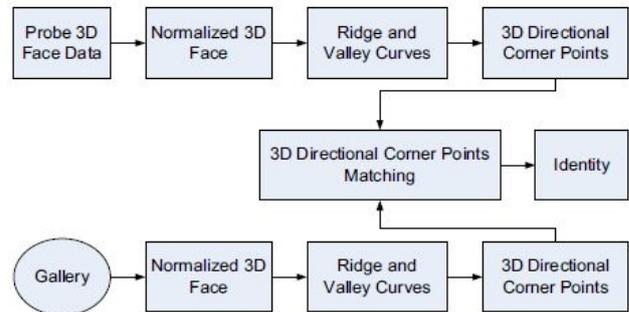


Fig. 108. Representação do método proposto por Xun Yu. Figura extraída de [354].

Ejima [357] propõe um método de reconhecimento de objetos 3D com base em *reference point ensemble*, o qual é uma extensão natural do *generalized Hough transform*. O *reference point ensemble* consiste em vários pontos de referência, cada um dos quais é codificado por cores verde ou vermelho, onde os pontos de referência vermelhos são utilizados para verificar a hipótese e os pontos de referência verde são utilizados para a votação. A configuração dos pontos de referência é projetada dependendo do formato do modelo. No método proposto, um conjunto de *reference point ensembles* é gerado pelas características locais de uma cena 3D dada. Estas características locais são descritas através do Labeled-Surflet-Pair (L-Surflet-Pair), o qual é derivado do Surflet-Pair [136], e o conjunto de *reference point ensembles* é gerado desta descrição. Cada ponto de referência gerado é uma pose 3D hipotética de um dado objeto na cena. Hipóteses passando pela verificação pelo pontos de referência vermelhos são utilizados na votação. A votação *Hough* é realizada independentemente em cada espaço dos pontos verdes, o que reduz o espaço de votação para 3 dimensões. Apesar de um espaço de 6 dimensões é geralmente necessário para o reconhecimento 3D, no método proposto, o espaço de votação de 6 dimensões é decomposto em alguns espaços de 3 dimensões. Em outras palavras, o reconhecimento efetivo do objeto é alcançado pela mudança em dois diferentes modos: o modo individual, no qual é realizada a votação das hipóteses independentemente em cada espaço Hough dos pontos de referência verdes e a verificação da hipótese com os pontos de referência vermelho; e o modo *Ensemble*, no qual ocorre a verificação do registro em uma lista de hipóteses promissoras e a agregação do total de votos. O método proposto foi aplicado em dois diferentes tipos de componentes industriais um tubo com formato em Y e um suporte em forma de placa. No primeiro experimento um único tubo foi usado como objeto para reconhecimento 3D. O método proposto foi aplicado em 41 cenas, cada uma das quais inclui um único tubo em uma pose única. No segundo experimento uma pilha de tubos é utilizado como objeto 3D para reconhecimento. No terceiro experimento, o método proposto foi utilizado junto com o ICP de modo a recolher cada tubo presente na imagem. Adicionalmente, o método proposto foi aplicado para pegar parafusos. Os resultados dos experimentos foram demonstrados em termos de tempo de

processamento e taxa de sucesso no reconhecimento.

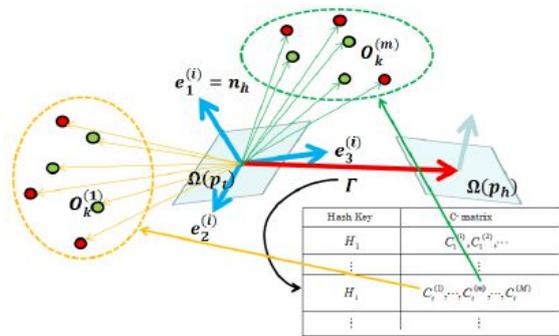


Fig. 109. Representação do método proposto por Ejima. Figura extraída de [357].

Dong Liang [358] apresenta um método de reconhecimento de objetos 3D e estimativa de pose com modelo de *deep learning*. Inicialmente, treina-se duas Deep Belief Networks (DBN) separadamente antes de conectar as últimas camadas para treinar um classificador. Para superar falha na detecção de objetos do modelo de *deep learning*, um método de detecção de objetos com base na clusterização por K-means é utilizado, extraindo assim o objeto do plano de fundo antes do reconhecimento. Comparado com métodos tradicionais com base em modelos *deep learning*, o foco do trabalho é não só o reconhecimento do objeto, mas também sua estimativa de pose. Métodos com base em *deep learning* possuem a habilidade de reconhecer ou prever um grande conjunto de padrões através da aprendizagem de características esparsas de um conjunto pequeno de padrões. Por essa razão, pode-se treinar o modelo de *deep learning* com um pequeno número de poses e então prever um grande conjunto de poses com o modelo. No sistema proposto, diferentes poses dos objetos significam diferentes classes, o que significa que uma classe representa uma pose de um objeto no modelo DNB. Logo, ocorre uma simplificação do problema de estimativa de pose para um problema de classificação. Desta forma, pode-se não apenas reconhecer o objeto, mas também estimar a pose com base na saída das últimas camadas. A base de dados utilizada é composta por 4 objetos incluindo 2310 imagens para o conjunto de treinamento e 3150 imagens para o conjunto de teste com o mesmo plano de fundo sob diferentes iluminações. Com base nesta base, comparou-se o método proposto com outro método para reconhecimento e estimativa de pose apresentado anteriormente [310] em termos de taxa de reconhecimento.

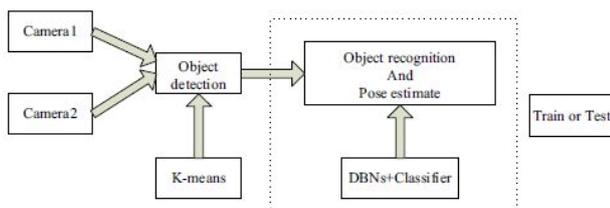


Fig. 110. Representação do método proposto por Dong Liang. Figura extraída de [358].

Carrer [359] propõe um algoritmo para detecção e classificação de objetos 3D em Imagens de radar de micro-ondas de alta resolução para propósitos de triagem de segurança. A ideia por trás da abordagem proposta tem como base a aplicação de um algoritmo de correspondência óptica na imagem (SIFT) de radar de micro-ondas de alta resolução. Para reduzir o número de falsos alarmes causados pelo SIFT a ideia principal é combinar diferentes projeções de dados escalares volumétricos 3D junto com segmentação e *histogram thresholding* em combinação com um filtro de correspondência. A tarefa da unidade de segmentação é a de dividir a imagem em partes menores de acordo com uma sobreposição parcial entre si do *grid* primário e secundário. Cada segmento, antes da detecção com o SIFT, tem seu contraste melhorado através da realização de equalização de histograma adaptável limitada ao contraste [360]. Para cada item armazenado na biblioteca de objetos, os descritores SIFT foram previamente computados e armazenados junto com a imagem de micro-ondas do objeto. Os descritores de cada objeto são então comparados com os descritores gerados pela aplicação do SIFT aos segmentos. O número de correspondências para cada segmento, o qual é determinado por uma abordagem de vizinho mais próximo com base na distância euclidiana, é denotado por α . Segmentos candidatos, para os quais $\alpha > 0$, são enviados para a unidade de *histogram thresholding* para investigação adicional. Muitos segmentos para os quais o SIFT identifica um objeto infelizmente não o contém, tendo a presença apenas do plano de fundo. O *histogram thresholding* permite descartar esses segmentos de plano de fundo que causam falsos alarmes. A saída do filtro de correspondência é um coeficiente de correlação β definido como a soma da matriz de correlação do segmento. Para cada um dos oito *grids* de segmentação, uma matriz de ranqueamento é construída, na qual os pesos sob o coeficiente de correlação e α de cada segmento em potencial são formados pela computação da pontuação S definida como $S = \beta(1 + \alpha)$. A matriz de ranqueamento global coleta a informação de pontuação de todos os *grids* de segmentação e decide qual segmento, de um *grid* em particular, contém o objeto. Se α é igual a zero em todos os segmentos significa que não existe nenhum objeto na imagem ou não foi detectado. A validade e confiabilidade do algoritmo proposto foi verificada experimentalmente, utilizando um manequim com vários objetos escondidos sob o casaco e avaliando o potencial de detecção e classificação do algoritmo.

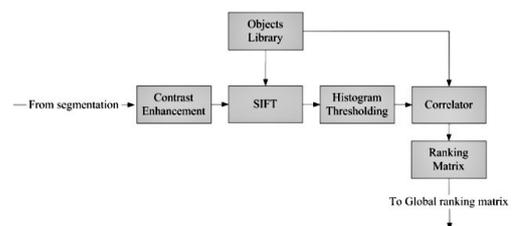


Fig. 111. Representação do método proposto por Carrer. Figura extraída de [359].

Geetha [361] propõe um algoritmo o qual reconhece o objeto 3D utilizando as características de formato 3D da

superfície, características de fronteira do formato 2D e características de cor. Para a extração do formato 3D do objeto o primeiro passo é a detecção de pontos de referência. Para identificar tais pontos dois métodos são utilizados pela abordagem proposta. O primeiro utiliza as derivadas de primeira e segunda ordem da função de distância de cada ponto na superfície 3D para um plano de projeção. O segundo, procura algumas distâncias de pontos uniformes na imagem RGBD. Após detectar os pontos de referência, vetores de características correspondendo aos pontos de referência, são computados. Para isto utiliza-se o conceito de curvaturas principais, as quais são calculadas utilizando o Principal Component Analysis (PCA). Para a extração da fronteira do formato 2D, descritores HOG de fronteiras são utilizados e para a extração das características de cor, histogramas normalizados são empregados. Para avaliar a abordagem proposta foram realizados testes utilizando a base de dados RGBD object dataset [257]. O vetor de características, de tamanho 60, de cada objeto no conjunto de dados foi armazenado na base de dados. Para obter as correspondências mais próximas utilizou-se a distância euclidiana. O primeiro teste avaliou a acurácia do método proposto. O segundo teste, conduziu experimentos comparando o método proposto com um método utilizando a correspondência apenas com as características de formato 2D. O terceiro e último teste comparou o método proposto contra outros dois métodos 3D de Content-Based Image Retrieval (Closed sharp edge detection [362] e Basic shape detection [363]).

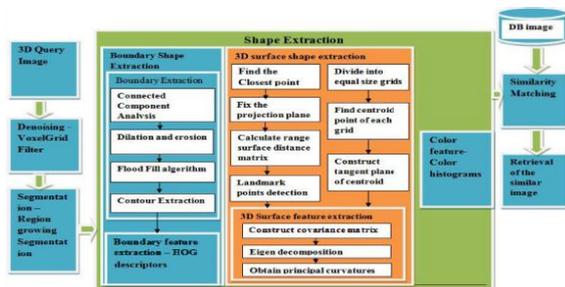


Fig. 112. Representação do método proposto por Geetha. Figura extraída de [361].

O trabalho apresentado por Mhamdi [6] utiliza um conjunto de funções de medidas para o reconhecimento de objetos 3D. O princípio das funções de tamanho é descrever o objeto 3D através da codificação das mudanças topológicas fornecidas por seus pontos críticos e pela ligação entre eles. Um ponto crítico pode ser um máximo ou um mínimo local ou um ponto de sela da função de medida definida no objeto 3D e a ideia é descrever o objeto 3D por uma característica como uma função. No caso em estudo cada objeto 3D é descrito por 18 funções de medida referente a 18 porções do objeto 3D buscando aproveitar as diferentes informações presentes em cada eixo. A comparação entre dois objetos 3D descritos pelas funções de medida é realizada através da medida de similaridade entre as 18 funções associadas a cada objeto, a qual poderia ser expressa através da mínima distância entre as distâncias correspondentes das 18 funções

associadas ao objeto. Entretanto, essa forma de comparação pode gerar confusões entre objetos de diferentes categorias, devido ao fato que pode-se encontrar objetos 3D de diferentes categorias que possuem uma ou mais porções similares. Para solucionar esse problema, o autor propôs ramificar a média das distâncias mínimas entre as 18 funções em duas partes. A primeira parte expressa a média das distâncias mínimas entre as funções de medida associadas com as porções que foram extraídas confiando em uma restrição apenas e a segunda parte expressa a média das médias mínimas entre as 12 funções de medidas restantes que foram extraídas confiando em duas restrições. Essas restrições fazem referência às porções do objeto selecionado. Antes de executar todo o processo de separação em porções e descrição através das 18 funções de medida ocorre uma etapa de pré-processamento, a qual é dividida em processamento de baixo nível, responsável na reparação de possíveis falhas nas malhas a fim de assegurar a conectividade entre os pontos e aumentar a qualidade das malhas para uso futuro, e normalização dos dados com a finalidade de dotar de invariância a transformações, torções e articulações às características a serem extraídas. O método proposto foi testado, inicialmente, experimentalmente com uma coleção de 457 modelos 3D em 19 categorias sendo esta coleção uma base de dados padrão para validação de descritores 3D. Depois foi testado em uma composição de base de dados 3D, a qual resultou em 2902 modelos e 121 categorias. A performance do processo de reconhecimento foi medida em termos de precisão e lembrança e comparada com demais métodos.

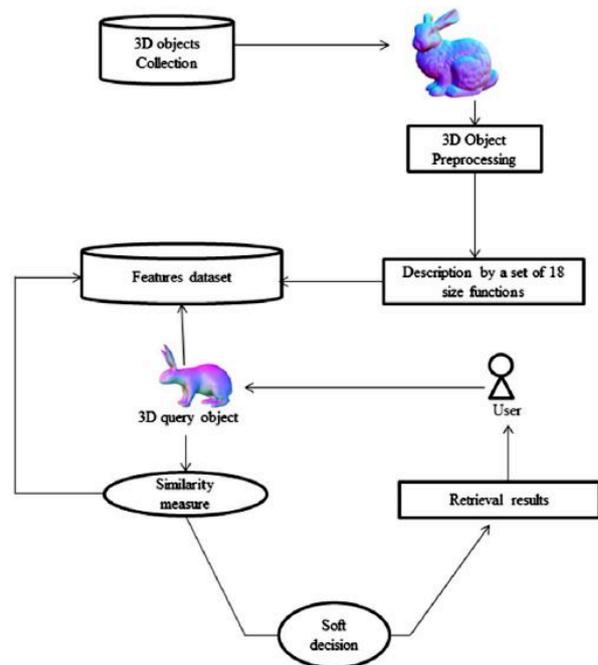


Fig. 113. Representação do método proposto por Mhamdi. Figura extraída de [6].

Yulan Guo [364] apresenta uma descrição de características locais de superfície denominado Tri-Spin-Image (TriSI) utilizado para o reconhecimento de objetos 3D na presença de

occlusão e desordem. O esquema montado para reconhecimento de objetos 3D consiste de 4 partes: pré-processamento, geração de características, comparação das características e verificação de hipóteses. No pré-processamento, pontos de característica são uniformemente selecionados de cada modelo e características do modelo são obtidas através do cálculo das características TriSI em cada ponto de característica. TriSI são obtidas através da concatenação de três assinaturas responsáveis pela codificação de características geométricas globais. Estas características de todos os modelos são utilizadas para construir um subespaço derivado da aplicação do método de Principal component Analysis (PCA). Depois, cada característica do modelo é projetado no subespaço gerado, obtendo uma característica comprimida. Estas características comprimidas são indexadas e armazenadas. O pré-processamento é realizado antes como um processo offline para a construção da base de consulta. A segunda parte é a geração de características, a qual dado uma cena, realiza o mesmo processo de geração de características projetadas no subespaço gerado pelo PCA. A terceira parte utiliza as características geradas na etapa anterior e calcula, para cada uma das características, a primeira e a segunda distâncias mais próximas entre a característica da cena e as características do modelo armazenadas. Se a característica do modelo f_m é o vizinho mais próximo da característica da cena f_c e a taxa da distância mais próxima para a segunda distância mais próxima é menor que um *threshold*, então a característica da cena e a característica do modelo ($f_c e f_m$) são consideradas características correspondentes. Cada correspondência entre características da um voto ao modelo e também proporciona uma estimativa de transformação entre a cena e modelo através do alinhamento de seus Local Reference Frame (LRF). LRF é composto pela posição de um ponto de características e de 3 vetores não ambíguos derivados do cálculo de autovalores e autovetores e são construídos para obter a invariância com relação a transformações rígidas. Na última parte, os modelos são ordenados conforme o número de votos recebidos e são verificados um a um com base nas transformações estimadas entre o candidato e o modelo. Nesta verificação é atribuído uma pontuação de confiança e a transformação com maior pontuação é utilizada para alinhar o modelo ao candidato e medir o alinhamento entre ambos. Se estiverem corretamente alinhados o modelo candidato é aceito caso contrário rejeita-se a transformação e confere-se a próxima transformação. Se todas as transformações entre o modelo e a cena foram testados o próximo modelo é testado seguindo a mesma abordagem previamente realizada. A abordagem proposta foi avaliada em duas base de dados populares (BoD1 e U3OR) em termos de lembrança e precisão. O TriSi foi comparado contra outros algoritmos com respeito a inserção de ruídos tais como: Gaussian noise, Laplacian noise, shot noise, variações na resolução da malha, oclusões e desordem.

Figueredo [365], propõe um algoritmo para reconhecimento de objetos 3D partindo de nuvens de pontos 3D de objetos simétricos rotacionados. A base do trabalho é um método anteriormente desenvolvido que representa o objeto utilizando uma tabela hash de características de formato e sobre esse algoritmo foi proposto uma forma de economizar tempo de

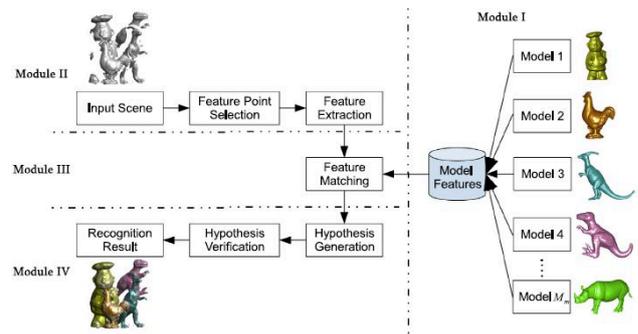


Fig. 114. Representação do método proposto por Yulan Guo. Figura extraída de [364].

execução e armazenamento. O algoritmo pode ser descrito em duas etapas: descrição do objeto em termos de surflets e reconhecimento do objeto. Surflets são as unidades básicas utilizadas para descrever o formato da superfície e são representadas por pontos de amostragem da superfície e pelas normais da superfície associadas. O processo de reconhecimento consiste em um processo de correspondência entre pares de surflets na cena e pares de surflets no modelo. Com base em um par de surflets pode-se calcular o Point Pair Feature (PPF), o qual é definido por quatro tuplas referentes aos dois pares de surflets associados. Buscando rapidez no processo de correspondência entre o modelo e a cena, a estrutura de dados utilizado para representar a descrição do modelo foi uma tabela Hash, na qual o valor de chave é dado por um PPF discreto enquanto o valor mapeado é o respectivo par de surflet. Com base na descrição através de surflet todo o processo de correspondência entre modelo e cena, cálculo de transformações e posterior estimativa de posição é realizado. No caso de objetos simétricos, o ângulo de rotação sobre o eixo de simetria não fornece nenhuma informação, logo a tabela hash contém informações redundantes. Neste caso é proposto uma forma de remover as características redundantes através da adição de um fator de peso para cada conjunto de objetos simétricos, o qual tem como base a detecção, por meio do cálculo de transformação, de pares de surflets que podem ser representados por pares homólogos. Vários experimentos foram realizados de forma a validar e avaliar o método proposto em termos de ganhos de performance e armazenamento e também em termos de robustez perante a ruído.

Yan Zhuang *YanZhuang2015* apresenta um *framework* de reconhecimento de objetos 3D, para um robô de serviço, para eliminar detecções falsas em ambientes desordenados, onde os objetos estão em uma grande diversidade de formatos e difíceis de serem representados pelos modelos exatos. Inicialmente, as nuvens de pontos de laser são convertidas para *bearing angle images* e uma abordagem com base em *Gentleboost* [366] é então empregada para detecção de objetos de múltiplas classes. Com o intuito de resolver o problema de escalas variáveis dos objetos na detecção dos objetos, uma técnica de coordenação de escala é adotada em cada sub-cena segmentada da cena total de acordo com a distribuição espacial dos pontos 3D do laser.

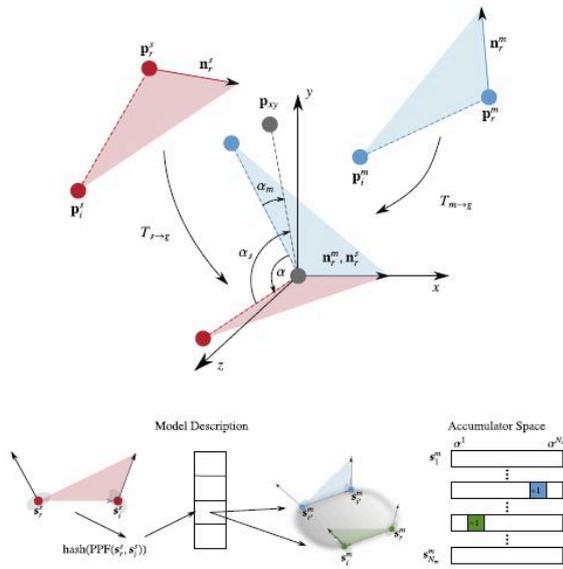


Fig. 115. Representação do método proposto por Figueiredo. Figura extraída de [365].

Além disso, informação semântica extraída dos pontos 3D do laser é utilizada para eliminar resultados falsos de detecção de objetos. Por fim, a clusterização por K-means e a distância de Mahalanobis são empregados para realizar a segmentação do objeto em uma nuvem de pontos de laser. O *framework* proposto foi avaliado em um conjunto de dados de laser, denominado DUT3-D, o qual contém 349 grupos de dados de escaneamento a laser de diferentes cenas e quatro tipos de objetos a serem detectados. Os resultados da detecção foram demonstrados em termos de curvas de precisão-lembrança.

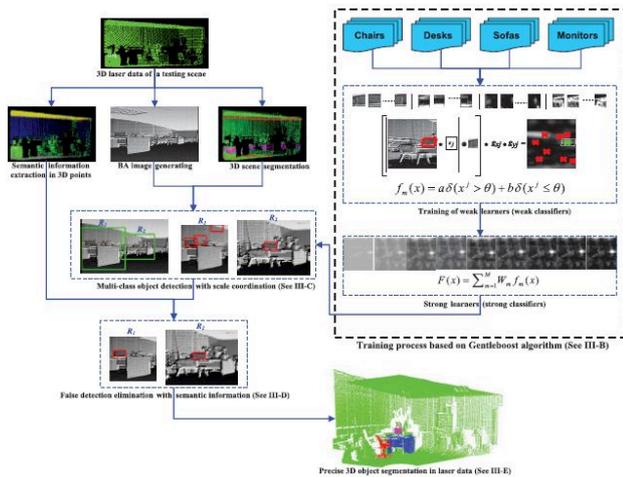


Fig. 116. Representação do método proposto por Yan Zhuang. Figura extraída de [367].

Sanguino [368] apresenta uma abordagem para detectar e classificar objetos 3D utilizando o método Hough generalizado e o sensor Kinect. O algoritmo considera pontos de característica e o espectro de cor como dois processos entrelaçados para cooperativamente reconhecer o objeto em

uma forma 2.5D. Com esta estratégia, o algoritmo automatiza as operações de pré-processamento independentemente das cenas e reduz a carga de processamento sobre a nuvem de pontos do objeto para classificação de objetos 3D. O processo de integração do formato e da informação de cor é realizado através da simplificação do Generalized Hough Transform (GHT) para seções de objetos 2D e do uso do espectro de cor como critério de decisão. A ideia básica do Hough Transform (HT) é extrair características geométricas das figuras pelo mapeamento do espaço de entrada (espaço da imagem) em um espaço de parâmetro mais adequado (espaço Hough). A seguinte sequência de passos é utilizada para a segmentação do objeto da cena: Primeiro, a imagem RGB é capturada pela câmera Kinect e convertida para o espaço HSI. Depois, a informação de profundidade é coletada pelo sensor do Kinect e mostrado em tons de cinza. A informação de distância é aplicada a imagem de entrada devido a correspondência similar e os objetos na cena são então discriminados. O filtro Canny é aplicado para detectar os contornos ao redor dos objetos e por fim obtém-se o resultado após a remoção da cena. Após a segmentação, utiliza-se a combinação do espectro Hough e do espectro de cor para o reconhecimento e classificação dos objetos, combinando e comparando os espectros da imagem de entrada e da base de dados, com base no valor de aptidão, classificando o objeto. Os experimentos foram realizados com base em quatro casos de estudo: reconhecimento entre objetos diferentes, reconhecimento entre objetos similares, reconhecimento sob oclusão e reconhecimento sob mudança de perspectiva. Os resultados dos experimentos foram demonstrados em termos do resultado da pontuação de aptidão, tempo de processamento e Receiver Operating Characteristics (ROC) em um conjunto de 5 objetos.

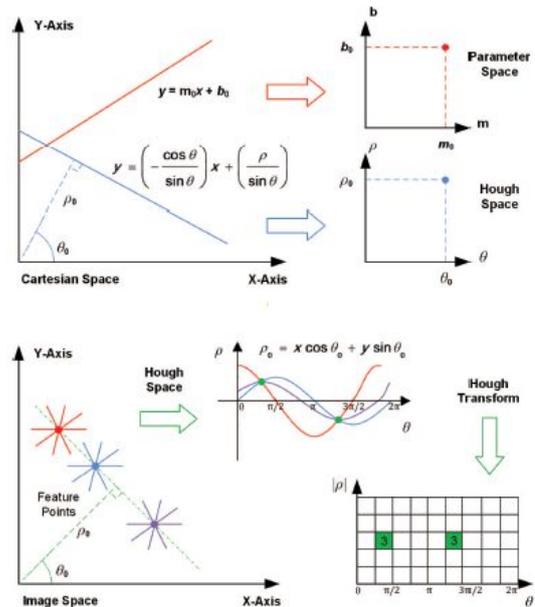


Fig. 117. Representação do método proposto por Sanguino. Figura extraída de [368].

Ribeiro [369] apresenta uma técnica efetiva de localização global utilizando *soft* reconhecimento de objetos 3D para

estimar a pose em relação aos pontos de referência de um dado mapa. Um sensor de profundidade adquire visão parcial para cada objeto observado, da qual o algoritmo proposto extrai a pose do robô relativa ao objeto, com base em uma biblioteca de Partial View Heat Kernel (PVHK) descritores [370]. O PVHK permite uma comparação fácil entre visões parciais sem requerer registro contra modelos 3D completo. O PVHK foi usado pela primeira vez no contexto de reconhecimento a partir de uma sequência de ângulos de visão [371], onde um filtro de partícula foi também utilizado. No algoritmo proposto, utiliza-se a mesma abordagem para comparar os descritores e estimar as probabilidades. Isto é, se estabelece a probabilidade que um descritor d corresponde a uma classe de objeto c e um ângulo de visão v pela computação da distância entre d e o descritor de referência no modelo do objeto. Então, assume-se uma distribuição exponencial nas distâncias, dado que ela tem uma distância de corte mais suave do que, por exemplo, uma distribuição gaussiana. Também utilizou-se a função de distância como proposto em [371], isto é, compara-se os descritores pela comparação do formato das curvas definidas pelo gráfico, onde a temperatura é plotada como função do comprimento limite. Deste modo, primeiro converte-se cada vetor dos descritores para uma curva e então, utiliza-se uma distância modificada de Hausdorff para estabelecer as distâncias entre os conjuntos. O algoritmo de localização foi validado em um conjunto diversificado de experimentos, em um ambiente fechado utilizando objetos do dia-a-dia. Os experimentos visam mostrar que o algoritmo estima uma posição final correta através da redução do erro com novas observações, como a inclusão de múltiplos objetos direciona para uma melhor estimativa e o impacto do *soft* reconhecimento de objetos contra o uso de pontos de referência sozinhos. Os resultados dos experimentos foram demonstrados em termos de erros da localização e da posição final.

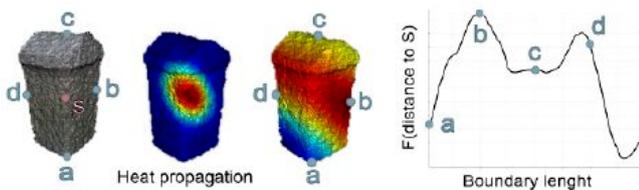


Fig. 118. Representação do método proposto por Ribeiro. Figura extraída de [369].

Sánchez[372] introduz um descritor projetado para a representação de classes de objetos. O descritor denominado, SCurV, explora informações do formato 3D computando e incorporando a curvatura das superfícies e as distribuições dos pontos projetados da superfície local, que representam planura, concavidade e convexidade, num descritor 3D centrado no objecto e dependente da visão. Estas diferentes fontes de informação são combinadas de uma forma simples mas efetiva de combinação de diferentes características para aprimorar os resultados de classificação. Logo o descritor proposto é resultado da computação dos seguintes quantidades: Uma representação global centrada no objeto com base na curvatura da superfície; Uma representação local centrada em

um ponto de visão fornecendo graus de planura, concavidade e convexidade; O descritor final, o qual é resultado da computação do produto tensor entre as duas representações previamente calculadas. Esse produto tensor fornece todas as possíveis combinações dos vetores das duas representações. Para testar o descritor 3D proposto utilizou-se a base de dados The Princeton Shape Benchmark [39] e comparou-se os resultados em duas formas de classificação (hierárquica e plana), utilizando para todos os descritores (ESF[166], VFH [145], CVFH [205], OUR-CVFH [373], PFH [374], FPFH[137], SHOT[150], USC[149], 3DSC [375], SI[16]) o mesmo classificador (margin based regression (MMR)[376], o qual é uma extensão do Support Vector Machine) em termos de valores de acurácia, precisão, lembrança e medida F1.

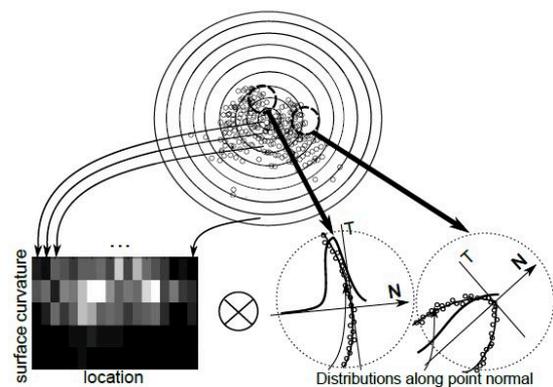


Fig. 119. Representação do método proposto por Sánchez. Figura extraída de [372].

Um método de descrição de características denominado SHORT (Shell Histograms and Occupancy from Radial Transform) é apresentado por Takei [377]. O SHORT consiste em um detector de pontos chave que não necessita calcular estatísticas e um descritor de características que utiliza uma pequena quantidade de pontos nas regiões locais restritas. O detector de pontos chave, avalia o formato local através do uso da ocupação, a qual é a proporção do número de pontos para o tamanho da região local. Na região local de um objeto a ocupação difere de formato para formato porque a extensibilidade espacial da nuvem de pontos local difere de um formato para o outro. Deste modo, os pontos chave são detectados através do uso da ocupação estimada como sendo o valor de avaliação do formato em cada ponto. O descritor de características descreve características de conchas de múltiplas escalas. Ele utiliza também os pontos existentes na região das conchas mais externas na esfera de múltiplas escalas e a ocupação estimada com a detecção dos pontos chave. Ele seta, com antecedência, as regiões de conchas esféricas que diferem em escala em um ponto chave e estima um vetor de direções dominantes pelo uso do valor de ocupação e da nuvem de pontos nas regiões de configuração para o ponto-chave. Então o descritor computa o histograma dos produtos internos de um vetor de direção dominante e os vetores direcionais para os pontos nas regiões de conchas para os pontos chave para cada escala. Por fim, integra-se um histograma dos produtos internos em cada escala como

sendo a característica de conchas de múltiplas escalas. Nos experimentos realizados comparou-se o descritor proposto, SHORT, com outros descritores (um método combinacional composto pelo método ISS [378] e o método SHOT [379]) em termos de reconhecimento de objetos 3D, utilizando a base de dados Descriptor Matching Dataset 1 [380], a qual foi utilizada como referência em [379], e empregando o método 3D Hough Voting[167] como algoritmo de reconhecimento.

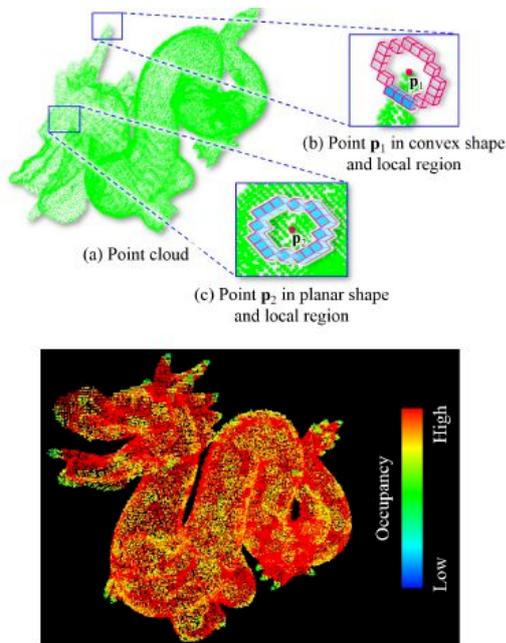


Fig. 120. Representação do método proposto por Takei. Figura extraída de [377].

Rocha [381] apresenta um *framework* para o reconhecimento de objetos 3D em um contexto industrial. O reconhecimento de objetos é realizado em duas fase diferentes, aprendizagem e classificação. Na parte de aprendizagem, o objetivo é treinar um algoritmo de classificação. Para este propósito, e pelo uso do sistema de luz estruturada, diferentes amostras de partes necessitam ser coletadas. O *framework* desenvolvido aceita diferentes fontes de dados: amostras reais extraídas da tabela Human Robot Interface (HRI) ou um arquivo CAD do objeto [382]. Após a realização da amostragem, uma segmentação necessita ser realizada. Tipicamente, este procedimento necessita ser aplicado em casos de amostras reais. No caso em questão, os dados da tabela HRI foi segmentada. O próximo passo é a extração de características, onde 5 características diferentes foram selecionadas: Dimensões (altura, largura e comprimento); Casco convexo da área da superfície da parte; e diâmetro médio dos buracos. Com base nessas características, organizadas em um vetor de transformação, treinou-se um Support Vector Machine (disponível na LibSVM [383]). Na fase de classificação, tendo o classificador do modelo treinado, é possível reconhecer os objetos quando o robô está em modo de produção. A mesma cadeia de processos empregada na fase de treinamento é utilizada nesta etapa, que consiste em: amostragem por meio da tabela HRI, segmentação dos dados HRI, extração de características e

por fim a classificação do objeto com base no classificador treinado. Os experimentos foram realizados utilizando o caso industrial em questão e extraindo 8 amostras diferentes de cada um dos quatro acessórios de viga considerados. Os resultados foram demonstrados em termos de taxa de classificação e comparados com os resultados obtidos utilizando uma abordagem com base no descritor View Point Feature Histogram (VFH), associado ao uso de uma heurística do K-nearest Neighbor [145], para classificação das mesmas partes de objetos.

Shimamura [384] propõe um método de verificação geométrica para lidar com mudanças de ponto de visão 3D em cenas com desordem para o reconhecimento robusto de objetos. Para superar a violação de consistência ocasionada pela mudança de ponto de visão, propõe-se verificação geométrica com base em votação utilizando novas restrições denominadas restrição de consistência *view-directional*. Tal restrição requer que o ângulo 3D entre as direções observadas de todas características de pontos correspondidas em duas imagens fornecidas, seja consistente com a pose relativa entre duas câmeras. O algoritmo de recuperação de objetos que incorpora o método de verificação proposto, inicia com a extração das características locais e seus descritores das imagens inqueridas, com parâmetros de ângulos observados incorporados. Subsequentemente, a correspondência de características locais é realizada entre a imagem inquerida e uma imagem candidata na base de dados. Todas as características locais extraídas da imagem inquerida são comparadas com todas as características locais extraídas das imagens candidatas através de seus descritores [47]. Então, a verificação geométrica proposta com base em restrições de consistência *view-directional* é conduzida. Na base das diferenças dos parâmetros das características locais, cada correspondência vota para o *grid* correspondente de uma mapa de votação. Este procedimento de voto identifica todos os *clusters* com ao menos sete entradas, o qual é um número suficiente para resolver a geometria *epipolar*. Para alcançar uma verificação mais rigorosa, cada um desses *clusters* passa por um procedimento de verificação geométrica completa, como uma hipótese. Utiliza-se o algoritmo LO-PROSC para estimar a geometria *epipolar*, então descarta-se *outliers* que escorregaram dentro de um *cluster* devido ao grande tamanho do *grid* ou outros erros. Se menos de sete correspondências permanecem depois do descarte dos *outliers*, o *cluster* é rejeitado. A decisão final para aceitar ou rejeitar a hipótese e a localização do objeto é realizada com base do modelo probabilístico de Lowe's presente em [44]. Por fim, o sistema fornece o número total de *inliers* para cada imagem candidata como sendo a pontuação. Para testar o método proposto, mediu-se a acurácia de recuperação de objetos na base de dados Stanford Mobile Visual Search com uma acurácia de correspondência de imagem como em [385]. A performance da acurácia para outras bases de dados foi medida com *mean average precision* (mAP).

O trabalho proposto por Fuhao Zou [386], busca explorar um conjunto de classificadores com a finalidade de construir um método de extração de características para classificação de objetos 3D. O método proposto segue a seguinte ideia: inicialmente um classificador é treinado para cada classe e então as saídas de todos os classificadores são combinadas

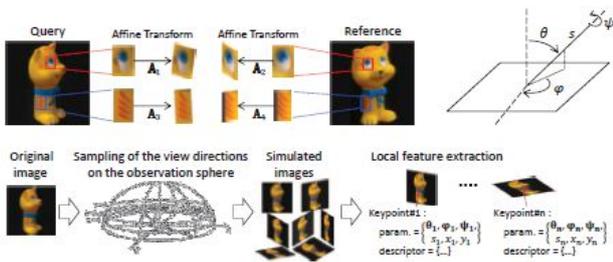


Fig. 121. Representação do método proposto por Shimamura. Figura extraída de [384].

como uma característica do objeto. Para construir o método de extração foi utilizado o l2-norm regularized logistic regression devido a facilidade de desenvolver a regra de atualização do l2-norm na forma de uma subida de gradiente estocástica, a qual torna o método proposto escalável para treinamento sobre um grande volume de dados. O método proposto foi comparado com outros 3 métodos de extração de características (SIFT-based BOF, sparse coding e deep belief networks). Vários experimentos foram conduzidos com a finalidade de verificar a habilidade de representação semântica, eficiência, escalabilidade e sensibilidade a variação. Os experimentos foram realizados em 4 base de dados (COIL-100[23], 3Ddata [387], ETH-80[105] e RGB-D[257]) e, com a finalidade de avaliar o método de extração de característica proposto, foram extraídos os vetores de característica para as 4 base de dados selecionadas e calculada a média da precisão média utilizando k-nearest neighbour search. Desta forma, os vetores de característica com as mesmas classes possuem menores distâncias e vice e versa. Logo, a acurácia de reconhecimento do objeto pode representar a extensão de quão bom é a característica extraída, ou seja, quão bom é o extrator de característica sendo utilizado.

Flitton [388] avalia a performance do modelo de classificação de objetos, Bag of Words (BoW), como uma abordagem para detecção de automática de objetos perigosos em imagens 3D de tomografia computadorizada. Nesse contexto, o modelo BoW é estendido de sua aplicação convencional 2D para imagens 3D. Para tanto, explora-se a combinação de 4 descritores de características 3D (Density Histogram(DH), Density Gradient Histogram(DGH), Scale Invariant Feature Transform(SIFT) e Rotation Invariant Feature Transform(RIFT)), três metodologias de atribuição de codebook (hard, kernel e uncertainty) e sete tamanhos de codebooks dentro de um framework de aprendizagem de máquina supervisionado baseado no classificador Support Vector Machine (SVM). O modelo BoW é composto de 3 passos: detecção e descrição de características, geração do codebook e classificação. Uma estratégia similar a esta foi utilizada para combinar os descritores de característica 3D, a metodologia de atribuição e o treinamento e classificação via SVM. Os experimentos realizados consideraram a classificação de dois objetos alvo (armas de mão e garrafas) em uma base de dados e as melhores configurações para cada configuração experimental apresentadas foram escolhidas com base nas configurações que maximizaram a taxa de verdadeiros positivos e portanto

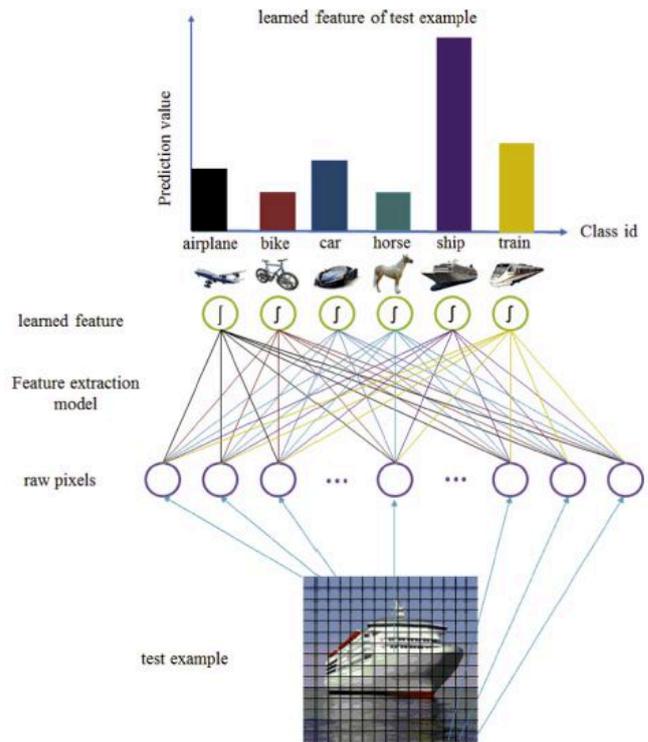


Fig. 122. Representação do método proposto por Fuhao Zou. Figura extraída de [386].

minimizaram a taxa de falsos negativos.

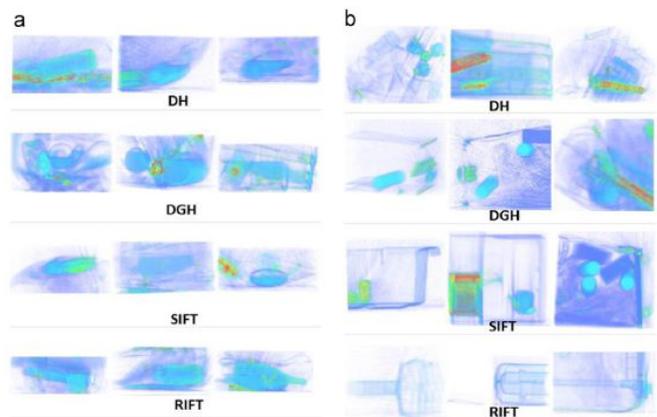


Fig. 123. Representação do método proposto por Flitton. Figura extraída de [388].

Qingtong Xu [389] apresenta um método para reconhecimento de objetos 3D, o qual inclui estimativa da normal, seleção de pontos de característica, extração de descritores de característica, correspondência entre características da cena e do modelo na biblioteca de modelos e geração e verificação de hipóteses. Entretanto no artigo o autor foca apenas nas primeiras três partes (estimativa de normal, extração de pontos chave e computação do descritor de característica local). Para a estimativa da normal, o autor analisa os autovalores e autovetores de uma matriz de covariância criada a partir dos vizinhos mais próximos do ponto inquerido. Para a seleção

de pontos característicos, utiliza-se um método de extração denominado NARF (Normal Aligned Radial Feature), o qual é utilizado para extrair os pontos de fronteira. O algoritmo é composto pelos seguintes passos: entrada de uma nuvem de pontos, e transversão de cada ponto para detecção de borda através da análise da mudança da posição de profundidade na vizinhança; Transversão de cada ponto, para computar o coeficiente de variação da superfície e a orientação, de acordo com as mudanças da superfície na área da superfície da vizinhança; Calculo dos pontos de interesse com base na direção principal do passo anterior, rotule as mudanças na superfície e a diferença entre esta e a outra direção, isso é, a estabilidade do ponto. Suavize e filtre valores de interesse; Por fim, encontre os pontos chave finais através de compressão de não máximos. Para o descritor de características utilizou-se o Fast Point Feature Histograms (FPFH) [137]. Após extrair o descritor de característica da cena e do modelo, os próximos passos a serem implementados são: o uso do KNN para a correspondência das características, votação por Hough para a geração de hipóteses e verificação de hipóteses globais para a verificação de hipóteses. Para os experimentos foram adquiridas nuvens de pontos com um escâner e testadas as três partes desenvolvidas e explanadas no trabalho.

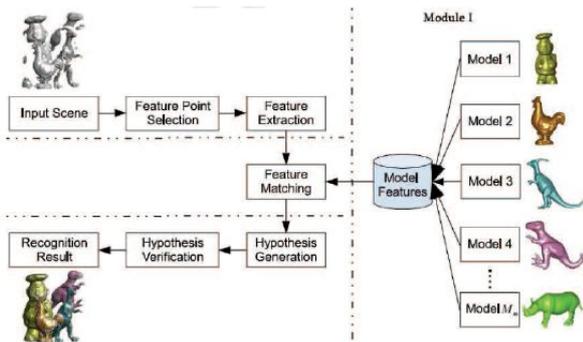


Fig. 124. Representação do método proposto por Qingtong Xu. Figura extraída de [389].

Um método para o reconhecimento de objetos 3D que utiliza características locais de forma eficiente é proposto por Soysal [390]. A abordagem proposta pode ser vista como a integração da invariante com base em uma abordagem geométrica detalhada em [391] com métodos modernos com base em aparência local, como por exemplo SIFT. Esta mesma abordagem está detalhada em [392]. O método proposto está dividido em duas partes, modelagem e reconhecimento. Entretanto alguns elementos principais estão presentes nas duas partes. Estes elementos são: as ferramentas de descrição e detecção de características locais invariantes, as quais fornecem um número factível trechos de regiões repetíveis bem localizadas e uma robusta descrição de suas aparências. Nesta pesquisa utilizaram-se os detectores de características locais Difference-of-Gaussian (DoG) e Harris-Affine (Har-Aff) como as duas alternativas representativas para detecção de características de baixo nível. As características locais da região são discriminadas pelo descritor SIFT [47]; as invariantes geométricas de transformações 2D-2D e 3D-3D e suas relações invariantes; e um método que emprega os pontos fortes da aparência e

geometria para corresponder grupos de características locais detectadas em imagens de objetos 3D, de forma similar ao método proposto em [393]. Analisando as duas partes separadamente, a parte de modelagem pode ser descrita em três passos principais: construção do grafo de adjacência entre as imagens do modelo; aplicação iterativa do procedimento central, *invariant-based Modeling Algorithm*, para identificação de características robustas do modelo; e criação da biblioteca de modelos utilizando as relações robustas de correspondência entre as imagens do modelo, as quais são determinadas no passo anterior. A parte de reconhecimento, também pode ser descrita em três passos principais: determinação da correspondência, com base na aparência, entre as características da imagem de teste e cada um dos conjuntos de características refinados das imagens modelo; Identificação de características geométricas 3D consistentes; e avaliação do nível de confiança para as características da imagem de teste que foram correspondidas anteriormente. O método proposto foi avaliado em uma base de dados pública [394] criada para análise comparativa de um conjunto grande de algoritmos para reconhecimento de objetos [393] e os resultados comparados com outros métodos da literatura que foram testados nesta mesma base de dados em termos de quantidade de acertos.

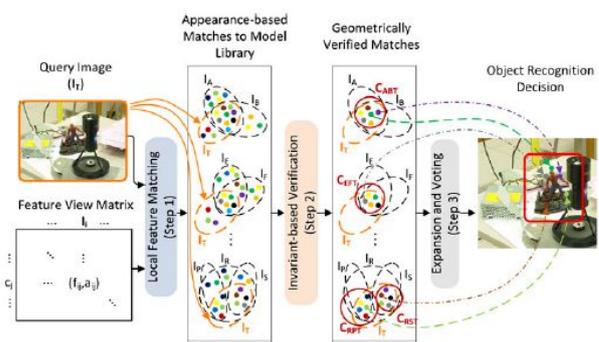


Fig. 125. Representação do método proposto por Soysal. Figura extraída de [390].

Beksi [395] apresenta um *framework* de aprendizagem de dicionário utilizando descritores de covariância RGB-D em nuvem de pontos para realizar classificação de objetos. Aprendizagem de dicionário em combinação com descritores de covariância RGB-D fornecem uma descrição compacta e flexível dos dados da nuvem de pontos. Descritores de covariância encapsulam características (posição, cor, normais, etc) sobre a nuvem de pontos do objeto por meio de uma única positiva definida matriz, a qual caracteriza o objeto. Estes descritores, no trabalho apresentado, são derivados de dois vetores de características compostos de invariantes de formato e visão. Estes descritores de covariância são utilizados para criar um dicionário representando o objeto. O dicionário é uma coleção de átomos que representam uma aproximação esparsa do descritor de covariância. Um conjunto destes dicionários para uma classe de objetos pode ser utilizado para classificar uma nova nuvem de pontos. Para cada nuvem de pontos do objeto, classifica-se a instância e a categoria do objeto, onde a classificação da instância consiste em determinar se um objeto é fisicamente o mesmo objeto, o qual já foi anteriormente

visto, e a classificação da categoria consiste em determinar se um objeto anteriormente não visto pertence a mesma categoria que objetos que já foram vistos. Para tanto, dois descritores de covariância são computados junto com suas respectivas representações esparsas. Para corresponder a categoria, itera-se sobre os dicionários da categoria computando o erro para cada dicionário. Então, se prevê a categoria com base no dicionário que produz o menor erro entre os conjuntos. De forma similar realiza-se a classificação da instância. Para avaliar a o método proposto empregou-se a base de dados RGB-D object database[299] e três experimentos foram conduzidos em um subconjunto dessa base buscando a classificação da categoria e da instância de um objeto. Os resultados foram comparados com outros métodos (Linear SVM [299], Nonlinear SVM [299], Random Forest[299], IDL [396], HKDES [327], Kernel Desc.[397], CKM Desc. [398], Upgraded HMP [399], Cov Desc. SVM [400]) em termos de acurácia de classificação e dimensionalidade.

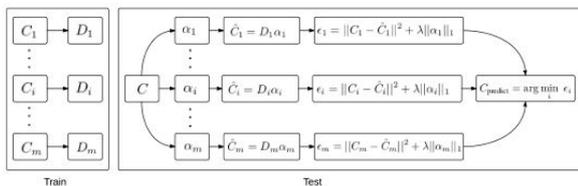


Fig. 126. Representação do método proposto por Beksi. Figura extraída de [395].

Filipe [401] apresenta um método para detecção de pontos chave 3D em nuvem de pontos e realiza uma avaliação comparativa entre cada par de detectores de pontos 3D e descritores 3D para avaliar suas performances no reconhecimento de objetos e categorias. O Método de detecção de pontos chave 3D proposto, denominado The Biologically Inspired 3D Keypoint based on Bottom-Up Saliency (BIK-BUS), é um detector de pontos chave com base em mapas de saliência, os quais são determinados pela computação de mapas de conspicuidade da intensidade das características e pela orientação de uma forma ascendente e baseada em dados. Estes mapas de conspicuidade são fundidos em um mapa de saliência e o foco da atenção é sequencialmente direcionado para os pontos mais salientes do mapa. Utilizando esta teoria e as seguintes etapas apresentadas em [402] e [403], os seguintes passos são realizados (filtragem linear, *Center-Surround Differences*, normalização, combinação através de escalas, combinação linear e *Inhibition-of-Return*) para apresentar o detector de pontos chave 3D, BIK-BUS. O *pipeline* utilizado no processo de avaliação de performance é composto das seguintes etapas: como entrada utiliza-se uma nuvem de pontos previamente segmentada do conjunto de dados da RGB-D Object Dataset [299]. Estas nuvens de pontos são alimentadas em um processo de detecção de pontos chave, os quais são utilizados para reduzir o custo computacional do sistema de reconhecimento. Então, computa-se um conjunto de descritores de características do objeto de entrada, o qual é utilizado para comparação com aqueles computados anteriormente e que estão na base de dados dos objetos. O modelo de objeto armazenado na base

de dados que contém a menor distância ao objeto de entrada é considerado o objeto correspondente. Na parte de detecção de pontos chave e computação dos descritores 3D, avaliou-se um conjunto de métodos (Detectores de ponto chave disponíveis na Point Cloud Library (PCL) [404] e alguns previamente avaliados pelo autor em [405]. Descritores avaliados: 3D Shape Context (3DSC)[98], Point Feature Histograms (PFH) [406], Fast Point Feature Histograms (FPFH) [137], Viewpoint Feature Histogram (VFH) [145], Clustered Viewpoint Feature Histogram (CVFH) [205], Oriented, Unique Repeatable Clustered Viewpoint Feature Histogram (OUR-CVFH) [373], PFHRGB, PPFHRGB, The Signature of Histograms of Orientations (SHOT)[150], SHOTCOLOR e SHOTLRF [324], Ensemble of Shape Functions (ESF)[166], The Point Curvature Estimation (PCE)) em termos de tempo de computação e tamanho. Depois, com base no *pipeline* apresentado, testou-se uma combinação de detectores e descritores e os resultados foram avaliados em termos de Receiver Operator Characteristic (ROC) Curve, a Area Under the ROC Curve (AUC) e a decidability(DEC) [407].

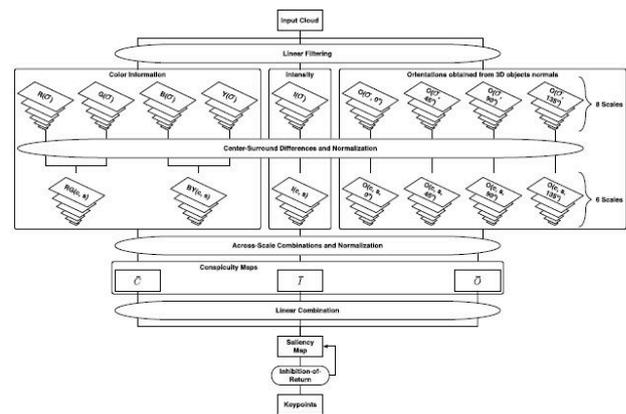


Fig. 127. Representação do método proposto por Filipe. Figura extraída de [401].

Garstka [408] propõe uma abordagem adaptativa para classificação de objetos 3D. Nesta abordagem, algoritmos apropriados de descritores de característica 3D para nuvem de pontos são selecionados via reforço de aprendizagem dependendo das propriedades dos objetos a serem classificados. O objeto principal da abordagem proposta é o aprendizado autônomo de uma combinação e aplicação otimizada de vários algoritmos de descrição de características 3D com o propósito de aumentar a taxa de classificação geral das nuvens de pontos 3D. Além disso, dependendo de algumas propriedades básicas da nuvem de pontos, a combinação desses algoritmos deve variar. Os passos principais para classificação são: Dado uma nuvem de pontos 3D inicia-se com uma coleção de propriedades globais. Essas propriedades são o número total de pontos, a resolução da nuvem, os autovalores e as dimensões da nuvem de pontos ao longo dos autovetores. Estes valores serão utilizados pelo agente de reforço de aprendizagem para selecionar o primeiro algoritmo. No segundo passo, o algoritmo de detecção de pontos chave intrinsic shape signature [378] é utilizado para determinar os pontos de interesse.

Durante o terceiro passo, um dos algoritmos de descrição de características locais 3D será aplicado. Como resultado obtém-se um conjunto de descrições de características 3D locais. Cada uma das descrições de características determinadas é quantizada para ser *binned* em um histograma. No último passo, os valores do histograma são utilizados como vetor de entrada para um classificador SVM para identificar uma apropriada classe de objeto. Este *pipeline* de classificação é aprimorado por meio de um agente de reforço de aprendizagem. No caso, os estados consistem no conjunto de algoritmos de descritores de características locais 3D que ainda não foram utilizados e das propriedades globais mencionadas anteriormente. As ações correspondem a seleção dos algoritmos e a aplicação dos três últimos passos do *pipeline* básico. Usualmente a classe de objeto com maior valor seria utilizada como um único resultado. Na abordagem proposta, utiliza-se os SVMs como classificadores binários com respostas -1 e 1 e rejeitam-se as classes com saídas correspondentes menores que 0 e mantém-se as demais classes como classes candidatas do objeto. Sem nenhuma restrição o agente de reforço de aprendizagem iria parar se não existir mais classes de objetos remanescentes ou algoritmos de descrição de características. Mas essa forma de termino natural não é desejada, dado que se propõe um tempo limite máximo para quanto tempo uma única classificação de objeto deveria levar. Portanto, o processo de aprendizagem se divide em tarefas episódicas que terminam em quatro possíveis estados finais. Estes estados são alcançados se todos os algoritmos foram utilizados uma vez, se não há mais classes de objetos remanescentes, se existe apenas uma classe de objetos remanescente ou se o tempo computacional acumulado excede o tempo limite. Apenas na situação em que existe mais de uma classe candidata de objetos, o agente continua com a próxima ação. A recompensa para o agente para todos os estados, exceto para os estados finais, é 0. Quando o tempo computacional acumulado é alcançado ou não há mais algoritmos ou classes de objeto restantes a recompensa também é 0. Para os casos restantes, quando existe apenas uma classe de objeto restante, a recompensa depende da fase do sistema de reforço de aprendizagem. Para avaliar os resultados obtidos, será determinado a taxa de reconhecimento, individualmente, para cada descritor de características local utilizando o *pipeline proposto* e subsequentemente comparado o resultado com os resultados obtidos pelo esquema adaptativo proposto.

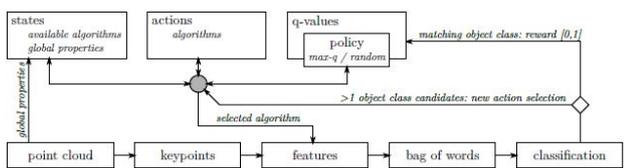


Fig. 128. Representação do método proposto por Garstka. Figura extraída de [408].

Naguib [409] propõe um classificador com base em Tree-Augmented Naïve (TAN) Bayesian Network. O espaço de características foi separado em regiões verdadeiro/falso o que permite dirigir a inferência Bayesiana das probabilidades condi-

cionais a priori de uma base de dados estatística. Utilizou-se também as regiões verdadeira/falsa para estimar probabilidades posteriores esperadas de cada objeto sob condições específicas ativas. Estas expectativas são então utilizadas para selecionar um conjunto de características ótimas sob esse ambiente e, de forma autônoma, reconstruir a rede Bayesiana. O sistema como um todo pode ser descrito pelos seguintes passos: treinamento do sistema; aquisição do objeto alvo; segmentação por octree; atualização das distribuições de probabilidade de cada característica de todos os objetos na base de dados; construção de uma tabela de força de discriminação e cálculo da probabilidade de discriminação; escolha de um conjunto ótimo de características (altura, largura média e SIFT) e construção de uma tabela de probabilidade condicional; computação da confiabilidade associada com estas condições suficientes; medição do conjunto ótimo de características e cálculo da probabilidade das características corresponderem ao objeto alvo; O sistema proposto foi treinado com 10 objetos diferentes, cada um com múltiplos alinhamentos e orientações. Para fins de comparação desenvolveu-se um adaptative Naïve Bayesian classifier [410] e os resultados dos experimentos foram apresentados de forma qualitativa, mostrando como o sistema se comporta sob várias condições, e de forma estatística e comparativa, comparando com o adaptative Naïve Bayesian classifier em termos de falsos positivos, falsos negativos e probabilidade média.

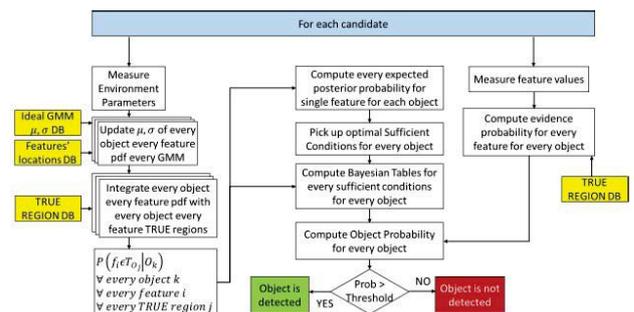


Fig. 129. Representação do método proposto por Naguib. Figura extraída de [409].

Kasaei [411], [412] apresenta uma abordagem eficiente capaz de aprender e reconhecer categorias de objetos de forma iterativa e sem a necessidade de conhecer os objetos antecipadamente. As instancias de treinamento são extraídas da experiência online de um robô, e assim se torna gradualmente disponível ao longo do tempo, ao invés de disponível no começo do processo de treinamento. O ciclo de processamento inicia quando o robô captura uma imagem da cena. O primeiro passo é a detecção do objeto, a qual envolve filtragem de distância, sub-amostragem e clusterização da nuvem de pontos dos objeto. O detector de objeto, o qual tem como base o algoritmo RANSAC [318], periodicamente requisita a outro módulo do sistema, denominado *tabletop segmenter*, uma lista de todos os objetos atualmente no topo da tabela. O módulo de detecção de objetos cria um novo *pipeline* de percepção para cada objeto novo detectado. Cada *pipeline* inclui módulos para o rastreamento de objetos, a extração de características e o reconhecimento de objetos. O rastreador de

objetos trabalha com base em um filtro de partículas [413], o qual utiliza informações geométricas assim como dados de cor e as normais da superfície para prever a provável próxima pose do objeto. O rastreador do objeto envia a nuvem de pontos do objeto para o módulo de extração de características, o qual computa as características para a visão do objeto fornecida por meio de um descritor de formato 3D (spin-images[16] associada a extração de pontos chave). As características dos objetos são mantidas em memória e o usuário pode fornecer rótulos de categorias para estes objetos. A rotulação de objetos, manipulada pelo módulo de interface com o usuário, dispara o módulo de conceituação de objetos. Em tal situação, o conceituador de objetos lê a atual categoria do objeto da memória, assim como o conjunto de características descrevendo o objeto rotulado, e cria ou aprimora a categoria do objeto. Durante o reconhecimento, uma regra de classificação por *nearest-neighbor* é utilizada para estimar o rótulo da categoria do objeto detectado. Todos os módulos retratados estão detalhadamente explicados pelo autor. Três experimentos foram realizados afim de avaliar o sistema proposto. O primeiro experimento avalia o sistema em termos de precisão e lembrança utilizando dados do mundo real. O segundo experimento conduz uma análise qualitativa do sistema como um todo. Por fim, o terceiro experimento avalia a performance da parte de aprendizagem através de um professor simulado, desenvolvido para acessar a performance de sistemas de aprendizagem de categorias *open-ended* de forma sistemática e reproduzível [414].

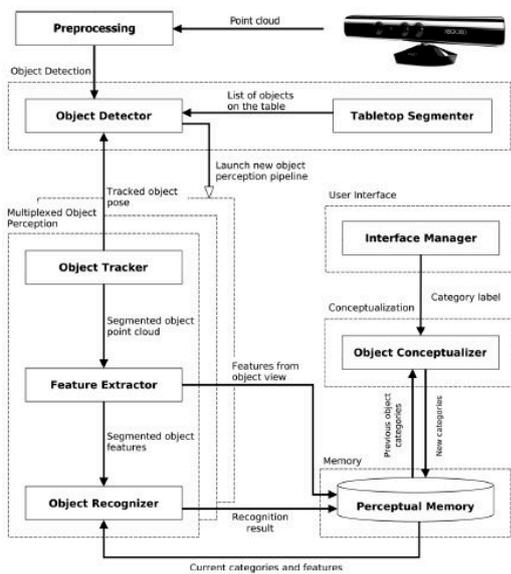


Fig. 130. Representação do método proposto por Kasaei. Figura extraída de [411].

Jianhua Wang [415] utiliza um modelo de Convolutional Neural Network (CNN) para aprender características de um conjunto de dados RGB-D, as quais são entregues para um classificador SVM linear (LinSVM) para classificar objetos. No trabalho em questão utilizou-se o *framework* de código aberto Caffe [416] com um classificador SVM para classificar o conjunto de dados RGB-D. Para fazer isso, primeiro ajustam-

se separadamente duas redes Caffe usando todas as imagens RGB e de profundidade. A técnica de ajuste é utilizada porque ela pode fazer pequenos ajustes aos pesos e as influências de forma que a rede final seja mais adequada para o conjunto de dados alvo. Para fazer a rede treinada em imagens RGB aplicável à imagens de profundidade, primeiro preenche-se os valores de profundidade faltantes em cada imagem de profundidade e converte-se a imagem de profundidade preenchida em três canais utilizando o método proposto em [417]. Então, utiliza-se a rede ajustada para extrair características das duas modalidades de imagem e concatena-se as características para treinar e testar o classificador SVM linear com rótulos de *ground truth*. A abordagem proposta foi testada no reconhecimento de categorias, classificando objetos não vistos sob as categorias nos dados de treinamento. Inicialmente, testou-se algumas combinações de dados de profundidade, dados RGB e ambos os dados para o treinamento e classificação utilizando o LinSVM e Nearest Neighbor (NN) retirando as características para treinamento da primeira e da segunda camada da CNN. Depois, compararam-se os resultados obtidos contra outras abordagens estado da arte (LinSVM [299], ksSVM [299], RF [299], SVM [397], CKM [398], CNN-RNN [254] e SP+HMP [399]) em termos de acurácia na classificação com as características RGB e profundidade isoladas e combinadas.



Fig. 131. Representação do método proposto por Jianhua Wang. Figura extraída de [415].

Ren C. Luo [418] [419] apresenta um método de reconhecimento de objetos 3D com base em modelo e busca através um robô 7-DoF com prevenção de obstáculos online para automação de fábrica. O robô pode pegar tipos randômicos de objetos 3D com uma pose arbitrária utilizando uma câmera 3D. *Pipelines* de reconhecimento de objetos com base em diferentes descritores foram comparados afim de decidir qual a melhor escolha para este cenário. A forma correta de pegar o objeto é ensinada por um operador humano utilizando programação por toque. O sistema como um todo é projetado utilizando uma máquina de estados dividido em três estágios: estágio de ensino, estágio funcionando e estágio Ocioso. Cada estágio é dividido em alguns sub-estados como mostrado na imagem 132. Como foco é na classificação de objetos e estimativa de pose, será analisado esse sub-estado em particular. Primeiro, de forma offline, gera-se uma base de dados de reconhecimento utilizando modelo CAD. Seguindo o *pipeline* de reconhecimento utilizando descritores globais, para cada *frame* obtido da câmera 3D realiza-se uma etapa de pré-processamento para segmentar o *cluster* do objeto. O pré-processamento inclui: sub-amostragem, remoção de pontos externos, segmentação de uma região de interesse; segmentação de planos RANSAC e clusterização do objeto com distância euclidiana. O resultado final é um *cluster* da nuvem de pontos do objeto. Então, computa-se o descritor global sobre esse *cluster*. No caso foram implementados o seguintes descritores: VFH, CVFH, OUR-CVFH, e ESF. Depois disso, uma procura

aproximada pelo vizinho mais próximo utilizando o FLANN [146] foi realizado na base de dados. Os k melhores candidatos são selecionados. Estes candidatos passam por alguns passos de pós-processamento para estimar a pose 6-DoF. Nesta etapa utiliza-se uma estimativa de pose com base no RANSAC e a correspondência dos pontos é realizada com base nos descritores locais. A transformação computada serve como um chute inicial para o ICP. Uma vez que um bom chute inicial foi dado, ICP realiza um refinamento da posição o que aumenta consideravelmente a precisão da estimativa de pose 6-DoF. Então, escolhe-se o modelo com o maior número de *inlier*, sobre os k candidatos, para ser o resultado de reconhecimento. O primeiro experimento foi realizado na categorização do tipo de objeto e estimativa de pose. Seis objetos e os seus modelos CAD foram utilizados, utilizando o método descrito em [205] para renderizar visões parciais para a base de dados de reconhecimento. Foram testados e comparados o uso dos descritores ESF, VFH, CVFH, e OUR-CVFH no *pipeline* proposto, demonstrando os resultados em termos de taxa de sucesso em relação ao número de candidatos utilizados e o tempo de computação necessário.

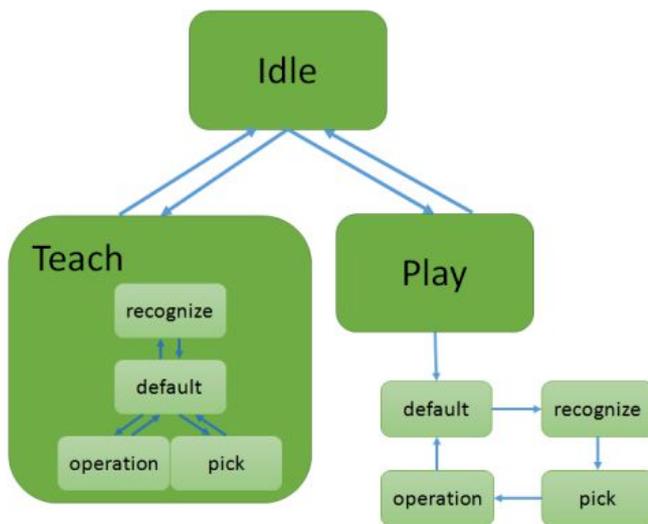


Fig. 132. Representação do método proposto por Ren C. Luo. Figura extraída de [418].

Rangel [420] propõe o uso de uma Growing Neural Gas (GNG) [421] para representar e reduzir os pontos da nuvem de pontos visando a redução do ruído das nuvens de pontos 3D e a melhoria no processo de reconhecimento de objetos 3D. Essa representação compacta é utilizada em um pipeline de reconhecimento de objetos afim de demonstrar a melhoria alcançada e comparar com outros método de filtragem/redução. O *pipeline* de reconhecimento tem como base o proposto em [229]. No entanto, foram introduzidas algumas modificações: a primeira modificação foi o uso do método de amostragem uniforme [404] como detector de pontos chave e o Unique Signatures of Histograms for Local Surface Description (SHOT) [150] como descritor uma vez que ele proporciona uma taxa de reconhecimento maior quando comparados com outros métodos. O *pipeline* de reconhecimento é descrito pelos seguintes passos: Inicia-se com uma nuvem de pontos 3D de

um modelo e uma cena; Estima-se as normais dos pontos para cada ponto da nuvem; Detecta-se os pontos chave para cada nuvem utilizando a amostragem uniforme; Depois, utilizando os pontos chave identificados, os pontos originais e a normal da nuvem, descritores são calculados pelo descritor SHOT; Utilizando os descritores computados, as correspondências são estimadas pelo método KdTreeFLANN [422]; Uma vez que as correspondências foram estimadas, o próximo passo é um processo de clusterização utilizando consistência geométrica [82]; Depois disso, todas as possíveis instâncias do modelo passam por um processo de refinamento utilizando o método Iterative Closest Point (ICP) [15]; Por fim, a verificação das hipóteses decide se as instância reconhecidas são falsos positivos ou verdadeiros positivos. No caso de verdadeiro positivo, o sistema mostrará um visualizador com a posição da instância do modelo reconhecido na cena; Para testar as diferentes abordagens, utilizou-se a base de dados da universidade de Bologna [150]. O experimento realizado, consiste em procurar um modelo selecionado, na cena com o *pipeline* descrito anteriormente. Primeiramente, utilizando as nuvens de pontos originais e depois as nuvens de pontos filtradas pelos métodos GNG e Voxel Grid [423]. Os resultados foram demonstrados em termos do número de verdadeiros positivos encontrados.

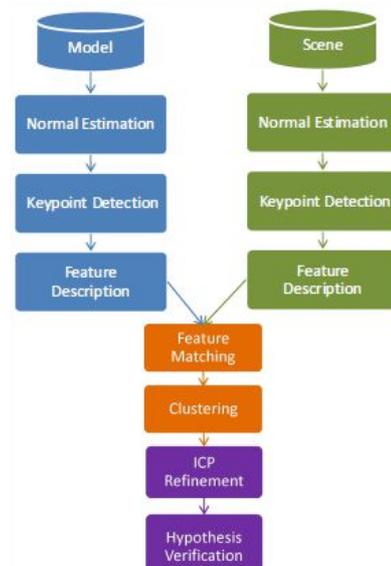


Fig. 133. Representação do método proposto por Rangel. Figura extraída de [420].

Um método para reconhecimento eficiente de objetos 3D com oclusão é apresentado por Yingjie Xia [424]. A proposta de utilizar um método com base em *deep learning* inicia com a construção de um modelo de formato de múltiplas visões com base em objetos 3D. Este modelo é construído através do uso de uma rede *deep learning* codificadora/decodificadora para representar as características. A rede utilizada para a aprendizagem das características é uma composição de blocos de Restricted Boltzmann Machine (RBM), a qual é uma rede neural que pode aprender a distribuição de probabilidade com base no conjunto de entrada, associados a uma Deep Belief Networks (DBN), a qual é um tipo de *deep neural*

network que comprimi múltiplas camadas de unidades ocultas que possuem conexões entre as camadas, mas não entre as unidades dentro da camada. O conjunto de características dos objetos de treinamento são aprendidos pela DNB e depois utilizados como entrada pelo Random Forest para classificar de acordo com os rótulos que representam várias classes de objeto, podendo ser formulado como $L = g(f(i), V)$, onde g refere-se ao classificador por Random forest, $f(i)$ é o vetor de características aprendido pelo DBN e V é o vetor de configuração dos parâmetros para os nodos utilizados para construir as árvores de decisão no Random forest. Nos experimentos realizados foram avaliados a acurácia do método proposto utilizando diferentes números de camadas do DBN e de unidades para o treinamento da característica. A abordagem proposta foi comparada com outros métodos (Support Vector Machine (SVM) e Random Forest sem o processo de deep learning) em termos de taxa de reconhecimento.

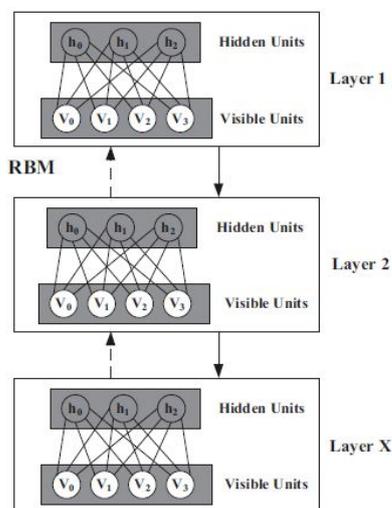


Fig. 134. Representação do método proposto por Yingjie Xia. Figura extraída de [424].

O trabalho apresentado por Chaoqun Hong [425], propõe um método de reconhecimento de objetos 3D com base em fusão de dados de múltiplas visões. Denominado Multi-view Ensemble Manifold Regularization (MEMR), pode ser dividido em duas partes. A primeira parte é o treinamento, o qual pode ser dividido nos seguintes passos: Inicialmente, as características da imagem, distribuição e informações de localidade são descritas e representadas através de Locality-constrained Linear Coding, o qual é similar ao tradicional bag-of-features, extraídas de diferentes visões do objeto, representando a imagem com um vetor de características. Um PCA também foi aplicado visando a redução da dimensionalidade do vetor de características. Depois, o kernel do Support Vector Machine e a matriz de penalidade são computadas para cada visão. Por fim, um processo de otimização alternada é aplicado para obter uma combinação refinada dos kernels e das matrizes de penalidade, obtendo desta forma o classificador com base em múltiplas visões. Para a parte de classificação é utilizado o mesmo processo de extração de características as quais são submetidas ao classificador treinado. Os experimentos realiza-

dos com o método proposto utilizaram duas base de dados (ETH-80[105] e COIL-100[23]) e os resultados obtidos foram avaliados em termos de taxa de acurácia e comparados com outros métodos (Support Vector Machine (SVM), Transductive Support Vector Machine (TSVM), Laplacian Support Vector Machine (LapSVM), Ensemble Regularized Manifold Support Vector Machine (EMR) e Multi-view Hypergraph Analysis (MHA)).

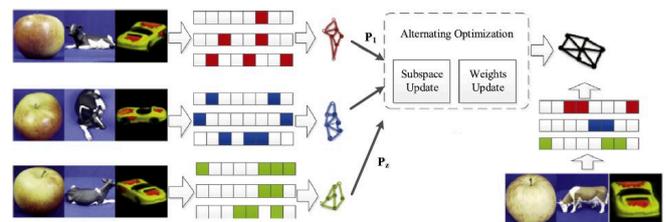


Fig. 135. Representação do método proposto por Chaoqun Hong. Figura extraída de [425].

Syed Shah [426] apresenta uma técnica de descrição local de superfícies para o reconhecimento de objetos 3D. A técnica proposta inicia com a fase de detecção de pontos chave, os quais são pontos que satisfazem às seguintes restrições: alta repetibilidade entre diferentes visões do mesmo objeto e a sua vizinhança deve incluir suficiente informação descritiva para caracterizar unicamente o ponto. Como vários pontos podem satisfazer essas duas restrições é proposto uma etapa de poda de pontos chave, a qual detecta locais na superfície 3D que violam o princípio de conservação de energia, e uma medida de saliência de pontos chave (utilizada para ranquear os pontos chave e selecionar os melhores para a etapa subsequente de correspondência). Uma vez que um pontos chave foi detectado, a informação predominante da superfície próxima ao mesmo pode ser extraída e codificada em descritores de característica local. O descritor utilizado é o Local Reference Frame (LRF), o qual é representado pelo próprio pontos chave, representando a origem e três vetores não ambíguos, calculados com base na área esférica do pontos chave, representando os eixos de coordenada. Como próximo passo, o vetor de campo normalizado da superfície local é alinhado com os vetores dos LRF para construir um descritor de superfície local invariante a rotação. Esse descritor, denominado 3D-Vor descriptor, é derivado das informações calculadas associadas a vorticidade da superfície local. Uma abordagem similar foi utilizada pelo autor, anteriormente, ao apresentar o descritor 3D-Div [427] para o reconhecimento de objetos 3D em cenas de baixa resolução. O descritor apresentado foi avaliado, inicialmente, com base na variação do parâmetro crítico, o raio. O poder descritivo e a robustez do 3D-Vor foram avaliados na base de dados de Bologna [150], para a tarefa de reconhecimento de objeto, com resolução variada de malha e diferentes níveis de ruído. A robustez do 3D-Vor também foi testada na base de dados de CaFLFoscari. Nestes experimentos, a performance do 3D-Vor foi comparada com outros descritores de características. Para a parte de reconhecimento de objetos, tem-se o processo de representação da cena, a correspondência das características, geração de hipóteses de

transformações, verificação das hipóteses e segmentação. A representação da cena passa pelo mesmo processo de geração do 3D-vor. A etapa de correspondência de características entre a cena e a biblioteca de modelos, previamente construída e descrita utilizando o 3D-Vor, ocorre com base em um cálculo de sobreposição, o qual tem seu resultado comparado com um limiar para dizer se as características do modelo e da cena são correspondentes. Cada correspondência gera um voto para o modelo e no final os modelos candidatos são ranqueados com base no número de votos. No próximo passo, calcula-se a distância de todos os pontos 3D do modelo com o ponto 3D mais próximo da cena e se a distância máxima entre as duas visões sobrepostas é menor ou igual a um limiar e o número de pontos correspondentes entre as duas visões é maior ou igual a uma taxa (razão entre número de pontos correspondentes pelo número total de pontos no modelo), então a transformação entre a cena e o modelo é aceita e passa-se para o próximo passo. Para a verificação da hipótese previamente gerada, utiliza-se um refinamento através do Iterative Closest Point (ICP), o qual resulta em um erro residual que é comparado com um limiar para a aceitar ou não a hipótese de transformação gerada no passo anterior. Uma aceitação na hipótese resulta no reconhecimento, estimativa de pose e segmentação do objeto modelo na cena. Para as análises quantitativas e qualitativas do método proposto foram utilizados três bases de dados (Bologna, UWA e CaFLFrosca) e comparado contra outros métodos de reconhecimento 3D (Spin image, TriSI, 3DTensor, key point based recognition, VD-LSD, EM, integral invariant based geometry descriptor e RoPs based algorithm).

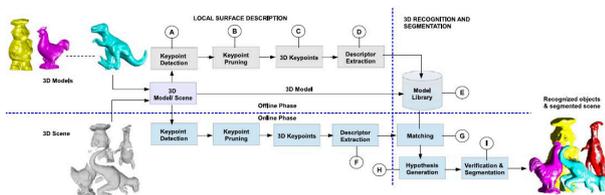


Fig. 136. Representação do método proposto por Syed Shah. Figura extraída de [426].

Kechagias-Stamatis [428] propõe um descritor 3D, o qual remove a necessidade de um Local Reference Frame/Axis (LRF/A) reduzindo o tempo de processamento necessário. O descritor proposto, denominado Histogram of Distances (HOD) tem como base múltiplas métricas L2-norm de remendos locais, fornecendo um descritor simples e rápido de ser computado. A computação do descritor é inspirado no shape distributions [209]. A principal diferença entre o descritor proposto e o shape distributions é a extensão da função D1 para uma base local e a substituição do ponto de referência do centroide dos pontos envolvidos para a borda. A última modificação aumenta a robustez do descritor HoD através da extensão de sua variabilidade L2-norm. Com base nesse ponto de referência, a L2-norm foi calculada para todos os vértices em cada área local, que foi devidamente normalizada e discretizada para um número pré-definido de bins. Depois disso, as distâncias normalizadas foram codificadas em um

histograma denominado Histogram of Distances (HoD). Para aumentar o poder descritivo do HoD, concatenaram-se as distâncias grosseira e a normalizada selecionando bins de tamanho adequado. A invariância a resolução da nuvem de pontos foi realizada pela normalização do HoD [150]. A invariância a resolução da malha foi estendida através da substituição da support radius metric por um múltiplo de cada resolução da malha da cena. As características da cena foram correspondidas com todas as características do modelo com base em suas distâncias euclidianas e o critério da taxa de distância do Nearest Neighbour. O descritor proposto, HOD, foi avaliado em duas bases de dados, Bologna and SpaceTime [150], analisando sua robustez diante de ruído e variação da resolução da malha. Também foi realizada a comparação do descritor com outros descritores (SHOT, RoPS, FPFH e Local D1) em termos da curva de precisão e lembrança.

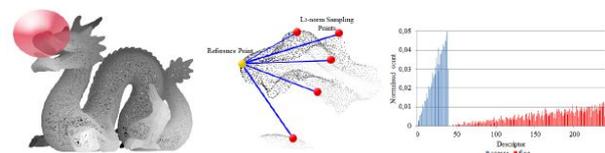


Fig. 137. Representação do método proposto por Kechagias-Stamatis. Figura extraída de [428].

Fan Zhu [429] apresenta um *framework* boosted cross-domain categorization (BCDC) que utiliza dados rotulados de outros domínios como dados auxiliares para expandir a diversidade intra-classe do sistema de aprendizagem original. O *framework* de classificação apresentado trabalha em conjunto com um método de aprendizagem de dicionário de *cross-domain* [430]. Pela atualização iterativa da representação dos dados do domínio auxiliar e de suas distribuições, as instâncias de treinamento do domínio auxiliar podem ser otimizadas e reproduzidas e assim podem ajudar a melhorar as tarefas de categorização visual no domínio de destino. O *framework* de aprendizagem proposto compartilha os mesmos princípios básicos de atualização sequencial dos impactos das instâncias de treinamento, mas ele tenta atualizar e reproduzir, sequencialmente, a representação dos dados das amostras dissimilares ao em vez de atribuir um peso menor a elas. Desse modo, o BCDC transfere conhecimento através de ambos níveis, representação dos dados e classificação. O BCDC tem como objetivo melhorar a performance de um classificador fraco, por meio do ajuste dos pesos das instâncias de treinamento, fornecendo assim um classificador forte. O BCDC tem como entrada os dados rotulados do domínio alvo, os dados do domínio auxiliar, o número máximo de iterações e o classificador fraco. O algoritmo inicializa a distribuição dos dados como uniforme, ou seja, os pesos iniciais possuem o mesmo valor. Depois, aplica-se o método de aprendizagem de dicionário de *cross-domain* para os dados do domínio auxiliar e do domínio alvo inicializados com a distribuição uniforme. Então para o número de iterações fornecido como entrada, seta-se a distribuição de dados, se reproduz os dados do dicionário aprendido como um conjunto adicional de dados auxiliares com base no domínio de dados auxiliares sob a distribuição de dados setada anteriormente, com a aprendiza-

gem de dicionário de *cross-domain*. Computa-se a hipótese, calcula-se o erro da hipótese, setam-se fatores e atualiza-se o novo vetor de peso. Como saída desse processo tem-se um classificador forte e representações atualizadas das instâncias do domínio auxiliar. Os experimentos realizados utilizaram seis diferentes fontes de dados, UCF YouTube dataset [431], Caltech-101 dataset [432] e SHREC 2014 benchmark [433] como domínio alvo e tratou-se as fontes de dados HMDB51 dataset [434], NYU Depth V2 dataset [435], e um conjunto de imagens da internet indexadas pelo google como domínio fonte. Para avaliar o BCDC método para reconhecimento ativo, classificação de imagens e tarefas de reconhecimento de objetos 3D utilizou-se as categorias que existem em ambos os domínios. Para a tarefa de classificação de imagens utilizou-se o SIFT+ modelo SPM para a representação inicial da imagem. Para a tarefa de reconhecimento ativo, projetaram-se trajetórias locais densas das características para um *codebook* aprendido (utilizando a clusterização por k-means) utilizando locality constrained linear coding (LLC) [436] de modo que o comprimento da característica global seja igual ao tamanho do *codebook*. Para a tarefa de reconhecimento de objetos 3D utilizou-se imagens de profundidade do NYU Depth V2 como dados de domínio auxiliar para melhorar o desempenho do reconhecimento de objeto 3D no conjunto de dados SHREC 2014. Os resultados dos experimentos foram demonstrados em termos de taxa de erro e classificação, comparando os resultados com outros métodos (ScSPM [437], K-SVD [438], Label Consistent-Singular Value Decomposition (LC-KSVD),[439] AdaBoost [332], Weakly Supervised Cross-Domain Dictionary Learning (WSCDDL)[430] e TrAdaBoost [440]).

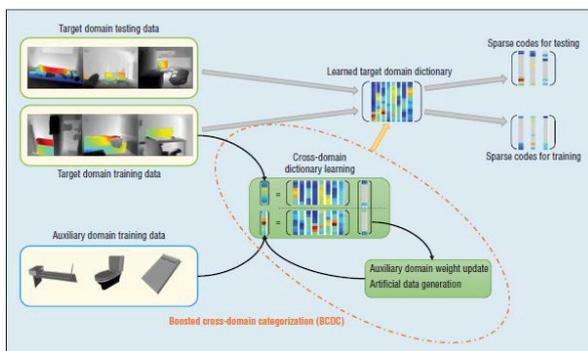


Fig. 138. Representação do método proposto por Fan Zhu. Figura extraída de [429].

Tateno [441] propõe um *framework* que é capaz de conduzir uma segmentação incremental, em tempo real, de uma cena 3D enquanto sendo reconstruída via Simultaneous Localization And Mapping (SLAM), e simultaneamente e incrementalmente conduzir o reconhecimento de objetos 3D e a estimativa de pose nas representações 3D reconstruídas e segmentadas. O *framework* proposto inicia com a reconstrução 3D, a qual tem como base uma abordagem Kinect Fusion [442], um método em tempo real que estima a pose de câmera de um sensor móvel e depende de uma representação volumétrica da superfície denominada Truncated Signed Distance Function (TSDF). Sucessivamente, o algoritmo de segmentação

incremental tem como base uma abordagem recente proposta em [443]. Contrária a abordagem em [443], a qual fundi os segmentos dentro de uma nuvem de pontos (the Global Segmentation Map), dado que a abordagem de reconstrução implementa a Kinect Fusion e uma representação com base em TSDF, propõe-se fundir incrementalmente os segmentos 3D dentro de uma representação específica com base em voxel, a qual denomina-se Label Volume. Por fim, a parte de reconhecimento é inspirada no *pipeline* do descritor global proposto em [444] para fornecer a correspondência 3D entre um mapa de profundidade da cena e um conjunto de visões renderizadas de cada modelo 3D. Diferentemente do proposto em [444], o *framework* proposto computa o descritor 3D diretamente de cada segmento 3D derivado do estágio de segmentação incremental, e corresponde-o com um único descritor 3D computado no modelo completo do objeto 3D. Dado que a representação 3D do segmento atual é um conjunto de vértices 3D, pode-se empregar diretamente os descritores 3D propostos para nuvem de pontos, tal como VFH [145], CVFH [205] e OUR-CVFH [373]. Para aumentar a robustez deste estágio, propõe-se um método denominado verificação geométrica, a qual é conduzida após a correspondência do descritor e tem a capacidade de reconhecer o modelo correto com um pequeno número de pontos de visão. O estágio offline consiste em computar, para cada modelo na base de dados de objetos, um descritor 3D da nuvem de pontos do modelo, a qual pode ser obtida diretamente de um modelo CAD do objeto ou via reconstrução SLAM densa. Durante o estágio offline, armazena-se também os descritores dos modelos em uma árvore-KD, fornecendo a importante propriedade de escalabilidade da correspondência entre descritores em log-linear em relação ao número de modelos na base de dados. Para os experimentos, o *framework* proposto foi comparado com o *pipeline* de reconhecimento de visão única proposto em [444], utilizando a base de dados RGB-D Scenes Dataset v2 [445]. Os resultados foram comparados em termos de curva de precisão e lembrança utilizando 3 descritores 3D previamente citados. Adicionalmente o *framework* foi testado utilizando um número diferente de pontos de visão e o descritor OUR-CVFH, com a utilização ou não da etapa de verificação geométrica. Os resultados deste teste também foram demonstrados em termos de curva de precisão e lembrança.

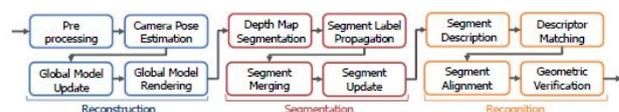


Fig. 139. Representação do método proposto por Tateno. Figura extraída de [441].

Sukhan Lee [446] apresenta uma abordagem para reconhecer com precisão objetos industriais e estimar suas poses com base em um *framework* Bayesiano com ótima seleção de características e correspondência de nuvem de pontos com base em modelo. O *framework* consiste principalmente de duas partes: reconhecimento de objetos 3D com múltiplas evidências e estimativa de pose do objeto 3D com base em modelo. No processo de reconhecimento, evidências candi-

datas incluem características globais e locais e são extraídas dos dados RGB-D, ou seja, extraem-se da nuvem de pontos o 3D SIFT [42], CLB[447] e descritores de forma, que descrevem a textura o formato geral do objeto [448]. Após coletar as características candidatas, aplica-se a seleção de características com base no teste-t para escolher características suficientes para diferentes objetos alvo como evidências de suporte. Por fim, as características medidas são comparadas com cada objeto na base de dados utilizando uma rede Bayesiana. O objeto com maior probabilidade é considerado como o resultado do reconhecimento. Para o objeto reconhecido, registro 3D de linhas e 3D SIFT são utilizados para a estimativa inicial da pose. Se ambos estiverem disponíveis, o com maior valor de confiança seria adotado com base na acurácia do registro. Então, emprega-se a Octree-ICP para obter uma estimativa refinada da pose. Com a finalidade de validar a abordagem proposta, testou-se o sistema de reconhecimento com 10 objetos industriais em 2 testes. O primeiro de reconhecimento, no qual calculou-se a taxa de verdadeiros positivos e o segundo de estimativa de pose, no qual comparou-se o resultado de pose obtido com base em um *Ground truth* fornecido por um braço mecânico.

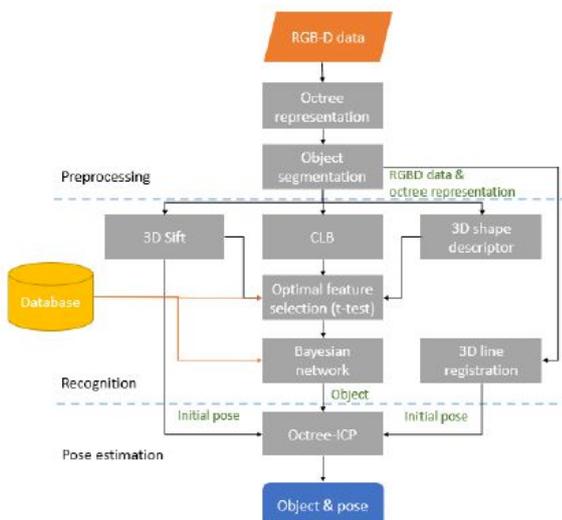


Fig. 140. Representação do método proposto por Sukhan Lee. Figura extraída de [446].

Iyappan[449] propõe uma metodologia para o reconhecimento de orelhas 3D utilizando spin images [16] e removendo características mapeadas erroneamente utilizando propriedades geométricas da superfície do formato 3D. Inicialmente, spin images são criadas para todos os vértices dos dados de orelha alvo e de teste e os pontos de correspondência são encontrados utilizando a correspondência entre spin images. Entre estas correspondências, as correspondências erradas são removidas usando a normal da superfície e a vizinhança estendida dos pontos correspondidos. Na parte de reconhecimento da orelha, os spin images da orelha alvo e orelha de teste são transformadas em matrizes X,Y. Utilizando o teste Chi-squared ambas as matrizes são comparadas para achar o vizinho mais próximo, o qual fornece menor valor de divergência, para cada linha de Y em X (comparando spin images). Os vizinhos

mais próximos são considerados como pontos correspondentes para a imagem de teste na imagem alvo. Experimentos foram realizados em um pequeno conjunto de 11 elementos com cada um contendo em média 4 imagens e os resultados demonstraram um fraco reconhecimento. Visto que os valores de reconhecimento foram ruins, escolheu-se a utilização de características fundidas para o reconhecimento. Desta forma, foram escolhidas duas imagens de treinamento, extraídos seus spin images e duas matrizes foram construídas, Z1 e Z2, com base nos spin images de cada imagem. Para a fusão, monta-se uma matriz única com base nas linhas únicas de cada matriz. A caracterização de linha única é alcançada com base na comparação de distância Chi square das matrizes Z1 e Z2, na qual se duas linhas são próximas uma da outra, mantém-se uma como característica única e remove-se a outra. Outra remoção executada tem como base a detecção de correspondências cruzadas. Uma correspondência cruzada ocorre quando existe uma correspondência de um ponto da imagem alvo, tal como um ponto em uma região plana da parte de cima da imagem, com um ponto da imagem de teste, tal como um ponto de uma região plana na parte inferior da imagem. Estas correspondências cruzadas devem ser eliminadas e isso é alcançado utilizando propriedades geométricas básicas da superfície, normal e curvatura principal. Experimentos foram realizados utilizando as duas melhorias propostas e os resultados demonstraram uma melhoria na detecção de orelhas 3D quando comparados com os resultados anteriormente obtidos.

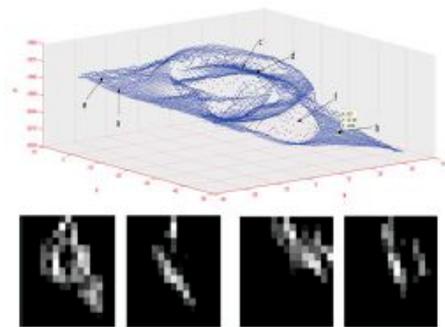


Fig. 141. Representação do método proposto por Iyappan. Figura extraída de [449].

Bedkowski [450] apresenta uma aplicação móvel inteligente para o suporte ao mapeamento espacial para a área da segurança. Para tanto, é apresentado o Complex Shape Histogram (CSH), um componente central do motor da inteligência artificial da aplicação, utilizado para classificar nuvens de pontos 3D com um Support Vector Machine. Utiliza-se também o CSH para o fechamento do laço de detecção no algoritmo de localização e mapeamento. O sistema inteligente móvel é construído em cima do *framework* Qualitative Spatio-Temporal Representation and Reasoning. Este *framework* define uma ontologia e um modelo semântico, os quais são utilizados para construir a interface com o usuário. Inicialmente, robôs adquirem dados 3D e registram eles utilizando uma versão melhorada do algoritmo 6D SLAM [451]. O 6D SLAM foi aprimorado através da

utilização da classificação semântica, fechamento de laço com o CSH e implementações em paralelo. A classificação semântica é utilizada para discriminação de pontos, deste modo o procedimento de procura por *nearest neighborhood* na correspondência semântica do escaneamento 3D utiliza esta informação para encontrar pares de pontos próximos tendo o mesmo rótulo semântico. Rótulos semânticos denotam formatos ao redor dos pontos inqueridos e estão relacionados com o contexto de computação dos CSHs. No CSH utiliza-se a classificação do formato em torno de cada ponto de consulta em duas classes (plano ou não plano). Na correspondência semântica de escaneamento 3D o número de classes é estendido para mais duas classes (telhado e chão). A parte principal da correspondência semântica de escaneamento 3D é uma versão modificada do Iterative Closest Point (ICP). O aspecto chave do ICP pode ser resumido em dois passos: computação da correspondência entre dois escaneamento (procura por *nearest neighborhood*) e computação da transformação a qual minimiza a distância entre dois pontos correspondentes. Repetindo iterativamente essas duas etapas resulta na convergência da transformação desejada. O Reconhecimento de objetos 3D é composto na detecção e identificação de Objects of Potential Interest (OPI). Para detectar pontos 3D de OPI utiliza-se um procedimento de procura por *nearest neighborhood* para cada ponto inquerido na medição 3D atual. Executa-se o alinhamento por ICP do escaneamento 3D atual com o modelo de referência global, minimizando o número de falsas detecções. Para identificar objetos 3D constrói-se uma base de conhecimento, deste modo prepara-se o conjunto de dados de treinamento composto de objetos com rótulos semânticos atribuídos. O reconhecimento dos objetos 3D é realizado pela classificação dos objetos observados em rótulos semânticos com base na base de conhecimento. Nesta aplicação existem dois rótulos semânticos OPI e not OPI. Esses rótulos são atribuídos manualmente, deste modo a vantagem do sistema é a capacidade de treinar para o reconhecimento de diferentes objetos. Para caracterizar os objetos 3D utiliza-se o CHS, o qual é uma extensão do Point Pair Feature (PPF) [136] que reduz a dimensão do espaço de características e utiliza a caracterização do formato atribuído pelo rótulo semântico ao redor do ponto inquerido. Para classificar os objetos identificados utiliza-se o SVM com base na nuvem de pontos adquirida com o escaneamento 3D e o conjunto de treinamento composto de exemplos positivos e negativos dos objetos. Os experimentos realizados visaram mostrar a capacidade do sistema proposto em construir modelos 3D do ambiente e procurar e identificar mudanças no ambiente, com base nos modelos construídos anteriormente em um estudo de caso do usuário final.

Kasaei [452] apresenta um descritor de objetos denominado Global Orthographic Object Descriptor (GOOD) construído para ser robusto, descritivo e eficiente para computar e utilizar. O descritor proposto é construído com os seguintes passos: Inicialmente, utilizando uma nuvem de pontos do objeto, aplica-se o método de Principal Component Analysis (PCA) e determinam-se os três eixos principais do objeto, aplicando um método de desambiguação para definir a direções dos três eixos principais e calcular um Local Reference Frame (LRF).

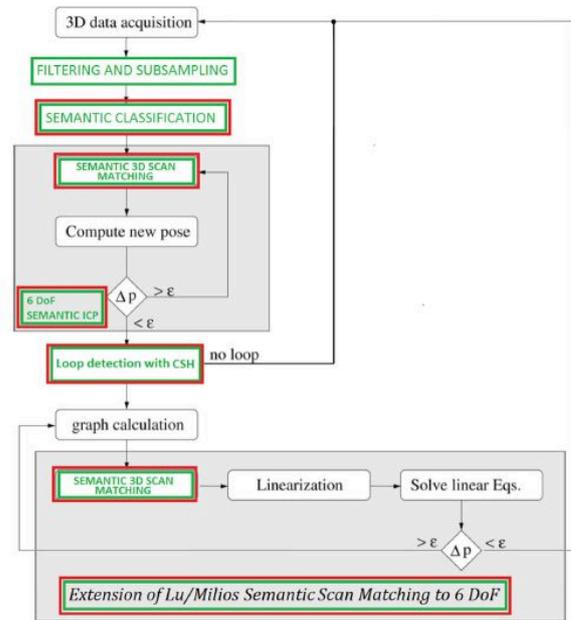


Fig. 142. Representação do método proposto por Bedkowski. Figura extraída de [450].

Com o LRF calculado, o próximo passo consiste em fazer a concatenação das projeções ortográficas do objeto nos três planos ortogonais. Cada projeção é descrita por uma matriz de distribuição com número de bins iguais e de mesmo tamanho, para garantir a comparação correta entre diferentes formatos de objetos. Para garantir invariância com a densidade da nuvem de pontos, a matriz é normalizada e depois convertida em um vetor. Os três vetores de projeção são concatenados produzindo assim o descritor. Vários experimentos foram realizados para avaliar a performance do descritor do objeto em relação ao poder de descrição, escalabilidade, robustez e eficiência. A performance e a escalabilidade do descritor foi avaliada na base de dados Washington RGB-D Object. Depois, vários testes foram conduzidos para medir a robustez, em relação a diferentes níveis de ruído e variações na resolução da malha de pontos, na base de dados Restaurant Object. Então, duas avaliações de eficiência relacionadas a eficiência computacional e uso de memória foram realizadas. Por fim, uma demonstração real foi realizada para mostrar todas as características do descritor proposto. Em todos os experimentos, uma abordagem de aprendizagem com base em instâncias foi utilizada para avaliar as representações. No caso, foi utilizada uma abordagem simples de Nearest Neighbour com distância euclidiana, comparando os resultados obtidos com outros 4 descritores disponíveis na biblioteca Point-Cloud.(Ensemble of Shape Functions (ESF) , Global Fast Point Feature Histogram , Viewpoint Feature Histogram (VFH) e Global Radius-based Surface Descriptor (GRSD)).

Fuhai Chen [453] propõe o uso do Support vector machine associado com 3 modalidades de representação de características para classificação de objetos 3D. O framework proposto inicia com a extração das características através do Sift, do Outline Fourier e do Zernike Moments de uma base

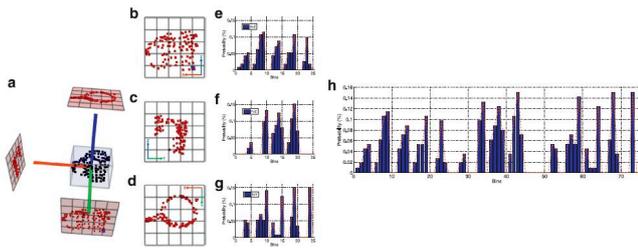


Fig. 143. Representação do método proposto por Kasaei. Figura extraída de [452].

de dados. As características extraídas por cada modalidade de representação muitas vezes são redundantes ou até mesmo irrelevantes. Com a finalidade de fazer uma representação discriminativa, um subconjunto de características relevantes é selecionado do conjunto de características compartilhadas pelas modalidades de representação. Com base nas características selecionadas o próximo passo é treinar o SVM com base em uma abordagem de multi-kernel SVM, a qual mantém a independência de cada modalidade de representação ao mesmo tempo que utiliza as características de cada modalidade no processo classificatório. O passo final é utilizar o SVM para classificar o novo objeto sendo pesquisado. Os experimentos realizados com o framework proposto foram conduzidos sobre uma base de dados contendo 8 categorias, 10 exemplares por categoria e 41 visões por exemplar e os resultados foram comparados com outros 10 métodos em termos de taxa de precisão, lembrança referente a taxa de verdadeiro positivos e F-score calculado com base na taxa de precisão e lembrança.

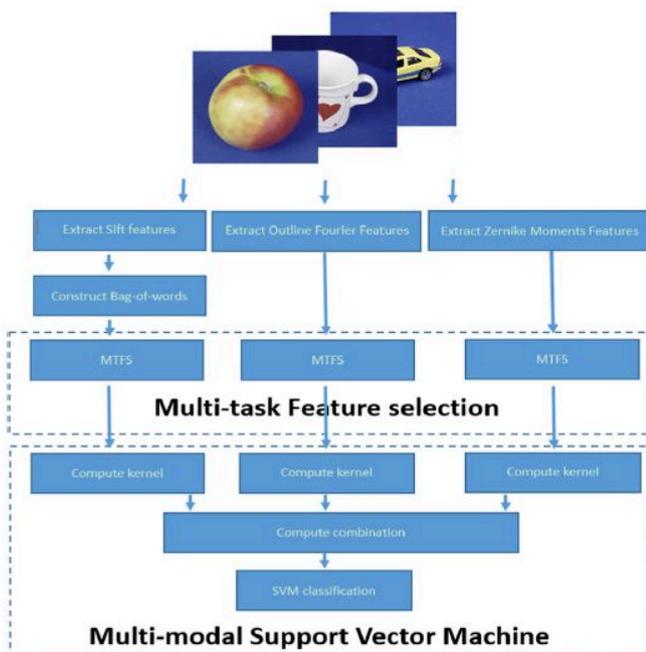


Fig. 144. Representação do método proposto por Fuhai Chen. Figura extraída de [453].

Chi Li [454] apresenta um algoritmo de segmentação semântico hierárquico, o qual particiona uma cena densa-

mente desordenada em diferentes regiões do objeto. Um registro com base no método RANSAC é subsequentemente utilizado para estimar uma pose 6-DoF do objeto dentro de cada classe de objeto. Parte deste algoritmo inclui um esquema de *pooling* generalizado utilizado para construir uma robusta e discriminativa representação do objeto de um arquitetura convolucional com múltiplos domínios de *pooling*. Uma típica arquitetura convolucional envolve dois passos principais: convolução de filtros locais sobre os sinais de entrada e *pooling* das respostas dos filtros dentro de uma vizinhança pré-definida. No trabalho em questão, explora-se dois espaços alternativos de *pooling*: SIFT [47] (gradiente) e FPFH [345] (geometria 3D) que são invariantes a variações de iluminação e rotação 3D. O *pooling* tradicional no domínio de espaço não pode ser aplicado diretamente para domínios de *pooling* de alta dimensionalidade como SIFT e FPFH devido ao crescimento exponencial do número de *pooling bins*. Deste modo, apresenta-se uma abordagem generalizada de *pooling* com base em K-means e na procura do vizinho mais próximo para domínios *pooling* arbitrários. O algoritmo de hierarchal semantic parsing é mostrado através de um *pipeline*. Primeiro é apresentado uma hierarquia de regiões propostas, a qual evita confiar nas heurísticas de fusão de região utilizadas na maioria das técnicas de segmentação de cena [455]. Esta hierarquia de regiões explora um grande conjunto de regiões parciais do objeto que vão desde os padrões locais até os globais. As características *pooled* de múltiplos domínios são eficientemente propagadas através dessa genérica hierarquia de regiões e os rótulos semânticos das regiões em todas as escalas são combinados para uma robusta segmentação semântica. Após a análise semântica, a cena original de nuvem de pontos está particionada em regiões com rótulos semânticos homogêneos. De forma similar a [456], estima-se as poses do objeto para cada classe semântica segmentada com um objRecRANSAC [232]. Para reduzir o número de falsos positivos retornados pelo objRecRANSAC, utiliza-se uma simples etapa de supressão de não máximos para filtrar estimativas de pose imprecisas. Considerando uma função que indica a pontuação de confiança da pose para um modelo, rejeita-se as hipóteses com pontuação baixa para qualquer par de hipóteses cujas intersecção 2D projetada é mais que 50 % de sua união. As hipóteses remanescentes são as estimativas de poses finais do algoritmo para a cena. O algoritmo proposto foi testado em dois conjuntos de dados, UW-RGBD Object Dataset [457] e JHUScene-50 Dataset [382], e os resultados foram apresentados em termos de valor de precisão e lembrança, comparando os resultados com outros algoritmos (SHOT+Hough Voting [167], Hypotheses Verification [229], Vanilla ObjRecRANSAC [232], Foreground Segmentation + ObjRecRANSAC, Object Segmentation + ObjRecRANSAC e Groundtruth Segmentation + ObjRecRANSAC)

Logoglu [458] propõe dois descritores locais 3D para a tarefa de reconhecimento de objetos, o Histograms of Spatial Concentric Surflet-Pairs (SPAIR) e o Colored SPAIR (Co-SPAIR). O SPAIR tem como base as relações de pares de surflets, onde um surflet é definido como um ponto orientado da superfície e as relações entre pares de surflets são definidas como relações geométricas entre surflets. Para calcular o

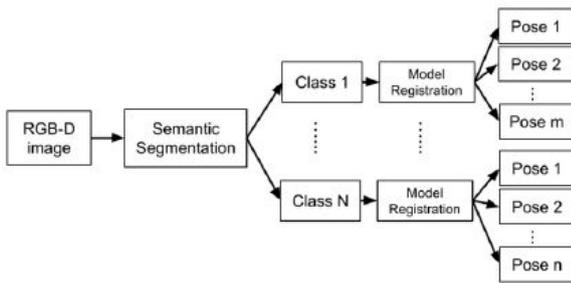


Fig. 145. Representação do método proposto por Chi Li. Figura extraída de [454].

SPAIR, inicialmente, tem-se a construção de um grid 3D de N regiões esféricas concêntricas de mesmo tamanho. Para cada concha esférica, denominada level, as relações entre pares de surflet entre os pontos dentro de um level e ponto fonte são calculadas. Quando os cálculos foram finalizados para todos os pares de surflets, os histogramas gerados são normalizados utilizando o número de pontos distintos em cada level. O descritor SPAIR é a concatenação dos três histogramas gerados em uma ordem com base na distância deles com o centro. No CoSPAIR, as informações de cor/textura e formato são codificadas para cada level do descritor SPAIR e o mesmo processo de concatenação é realizado adicionando estas informações ao descritor final. Para comparar os descritores propostos com outros descritores presentes na Point Cloud Library (Point Feature Histograms (PFH), Colored Point Feature Histogram (PFHRGB), Fast Point Feature Histograms (FPFH), Signature of Histograms of Orientations (SHOT), Color-SHOT (CSHOT)), o mesmo procedimento foi realizado para todos os descritores. Inicialmente, a base de dados (RGB-D Object Dataset [299], BigBIRD ((Big) Berkeley Instance Recognition Dataset [459] e Amazon Picking Challenge at ICRA 2015 [460]) é dividida no conjunto de consulta e no conjunto de referência dependendo do cenário de teste. Em ambos os conjuntos ocorre a extração de pontos chave (keypoints) através do Intrinsic Shape Signatures 3D (ISS3D), o cálculo dos descritores, sendo adicionado à base de descritores os descritores do conjunto de referência e os descritores das consultas entram no processo de correspondência e voto. Para a parte de correspondência e voto o OpenCV foi utilizado e para as demais partes o Point and cloud library foi utilizada. A performance dos descritores foi medida com base na acurácia média que tem como base o número de falsos positivos, o número de falsos negativos, verdadeiros positivos, verdadeiros negativos e o número total de rótulos de classe.

Naji [461] apresenta uma abordagem para o reconhecimento de objetos 3D com base em equações *heat*. Estas equações são utilizadas para calcular a distância geodésica entre qualquer pares de pontos no *Riemannian manifold*, utilizando uma transformação do *heat kernel*. O sistema de reconhecimento adotado pode ser descrito pelos seguintes passos: Cálculo das equações *heat* para os objetos 3D na base de dados; Cálculo das características base derivadas das equações *heat*; Treinamento de um classificador com base nestas características; Classificação dos objetos de teste em classes utilizando o

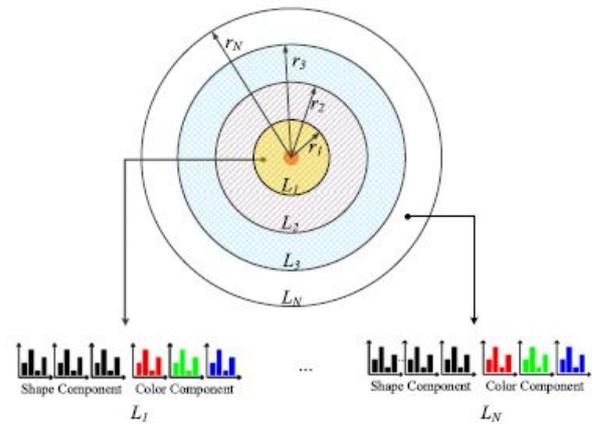


Fig. 146. Representação do método proposto por Logoglu. Figura extraída de [458].

classificador treinado. Para os experimentos foram utilizadas as bases de dados The base Princeton Shape Benchmark [39] e The Watertight datasets [462] e os resultados foram comparados com outros métodos (Light Field Descriptor [463], um método com base em visões salientes [464] e o método com base em distância geodésica [465]) com base nos resultados de reconhecimento alcançados ao se utilizar o descritor atrelado a um classificador, tal como C4.5, Random Forest (RF) e o K-Nearest Neighbour.

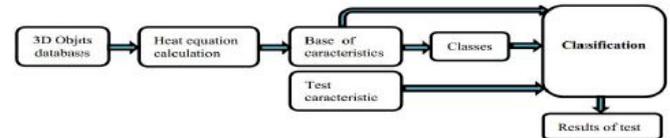


Fig. 147. Representação do método proposto por Naji. Figura extraída de [461].

O trabalho apresentado por Syed Shah [466], propõe uma representação com base em keypoints denominada Keypoints-based Surface Representation (KSR), utilizada para o reconhecimento de objetos 3D. O algoritmo completo para reconhecimento de objetos 3D passa por 2 fases, treinamento e reconhecimento. Na parte de treinamento, os keypoints são primeiramente detectados e o KSR entre os keypoints são computados para todos os modelos 3D e armazenados na base de dados de objetos. Durante a fase de reconhecimento, os KSR são calculados para uma dada cena. Estes KSR da cena são comparados com os KSR dos modelos utilizando um coeficiente de correlação linear buscando a correspondência entre KSRs. Os resultados das correspondências entre KSR são utilizados para votar em modelos candidatos e gerar as hipóteses para transformar o modelo para a cena. Os modelos candidatos são verificados em turnos através do alinhamento deles com a cena utilizando as transformações hipotéticas. Se o candidato modelo está alinhado com precisão a porção da cena, verifica-se esse alinhamento final com o algoritmo Iterative Closest Point (ICP), o candidato e a hipótese são aceitas. Como resultado os pontos da cena correspondentes

ao modelo são reconhecidos e segmentados, caso contrário, a hipótese é rejeitada e a próxima é verificada. Para verificar a performance da técnica proposta utilizou-se as base de dados UWA e Ca'Foscari em relação as taxas de precisão e lembrança. Os resultados foram também comparados contra outras técnicas de caracterização local (Rotation of Project statistics (RoPs), 3DTensor, VD-LSD, keypoint-depth local feature, Spherical Harmonics (SH) , Spin Images (SI) and Game-theoretic).

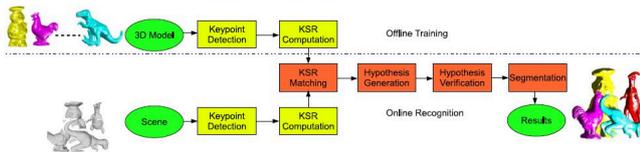


Fig. 148. Representação do método proposto por Syed Shah. Figura extraída de [466].

B. Representação com base em visões

Uma abordagem para solucionar o problema de escolha de visões ótimas (seleção pontos de visão) e fusão destas visões para um ótimo reconhecimento de objetos 3D (fusão de pontos de visão) é proposto por Deinzer [467]. Esta abordagem de seleção tem como base o reforço de aprendizagem. A seleção de pontos de visão pode ser definido como uma abordagem de duas etapas. Na primeira etapa, estima-se a função denominada *action-value function* $Q(s, a)$, durante o treinamento. Na segunda etapa, se em algum momento a fusão do ponto de visão retornar um estado s como resultado da classificação, seleciona-se o movimento de câmera o qual maximiza o esperado acumulado e o peso da recompensa. A função que maximiza o esperado acumulado e o peso da recompensa é denominado *policy*. O problema chave da abordagem é a estimativa da função $Q(s, a)$, a qual é a base do processo de decisão. Para solucionar este problema, sem a iteração do usuário, utiliza-se o reforço de aprendizagem, o qual possui muitos algoritmos diferentes para estimar a função com base em método de tentativa e erro [468]. O resultado do método de tentativa e erro é utilizado para atualizar a função $Q(s, a)$ e para aprimorar a função *policy*. A seleção do ponto de vista, ou seja, a computação de função *policy*, pode ser escrita, de acordo com [469], como um problema de otimização, o qual é solucionado pela aplicação do global Adaptive Random Search Algorithm [470] seguido por um simplex local. O objetivo primário dos experimentos foi o de mostrar que a abordagem tem a habilidade de aprender e realizar uma sequencia ótima de visões. O autor mostrou em publicações anteriores que a fusão dos pontos de vista de uma sequencia de visões selecionadas randomicamente funciona muito bem em ambientes reais e melhora os resultados de localização e classificação significativamente [470]. Por essa razão decidiu-se utilizar imagens sintéticas simples para a avaliação da abordagem de seleção de pontos de vista, avaliando a abordagem em termos de taxa de reconhecimento.

O estudo apresentado por Rui Nian [471] mostra um esquema probabilístico para reconhecimento de objetos 3D

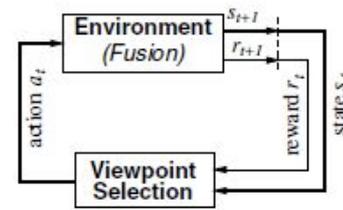


Fig. 149. Representação do método proposto por Deinzer. Figura extraída de [467].

partindo de uma sequência de visões 2D. O esquema proposto pode ser descrito em três partes: pré-processamento, aprendizagem com base em modelo de visões e reconhecimento. Na parte de pré-processamento, as imagens coletadas da base de dados, a qual consiste em vídeos 2D de tipos de objetos 3D, são processadas com a finalidade de extrair características das imagens e suprimir possíveis artefatos que dificultem o processo de reconhecimento dos objetos. A aprendizagem com base em modelos de visão é realizada através de um algoritmo de clusterização. Nesta etapa as imagens são *clusterizadas* em grupos e os modelos de visões são os responsáveis por minimizar as distâncias para todos os outros elementos no mesmo *clusters*. Modelos de visão são formados através da composição de generalizações dos membros em cada grupo, os quais correspondem a distintas visões de cada objeto. A parte de reconhecimento é realizada através de estimativas de funções de distribuição de probabilidade, isto é, para uma dada imagem de entrada é decidido se o objeto está presente ou ausente na cena com base em qual probabilidade é maior. Esta escolha é conhecida como regra de decisão Bayesiana. No trabalho em questão foi utilizada uma simplificação devido a complexidade de se representar completamente a probabilidade de um objeto pertencer a cena. Essa simplificação usa estimativas derivadas através do conjunto de treinamento, sendo a decisão final realizada através do maximum a posteriori associado a um mecanismo de acúmulo de evidência na memória de trabalho. Os testes com o método proposto foram realizados em alguns vídeos de objetos capturados por uma câmera rotacionada em 360 graus em incrementos de 5 graus e os resultados foram mostrados com base nas probabilidades estatísticas em relação aos modelos de visão.

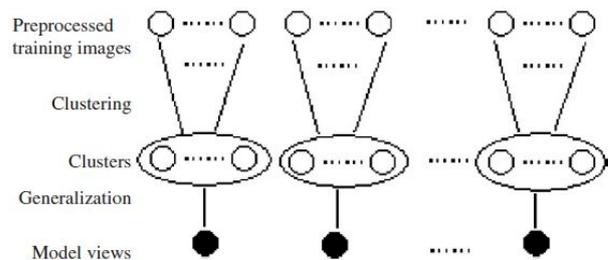


Fig. 150. Representação do método proposto por Rui Nian. Figura extraída de [471].

Zografos [472] apresenta um método com base em modelos para o reconhecimento de objetos 3D a partir de um

número pequeno de imagens 2D retiradas de pontos de visão arbitrários. O sistema proposto inicia com duas visões base armazenadas. Estas visões são imagens retangulares armazenadas contendo informações de *pixel* em tons de cinza sem informações adicionais. As imagens são obtidas de visões base escolhidas de tal forma que o ponto de visão, da qual a imagem alvo é adquirida, está na esfera de visão entre as visões base. Após a seleção de duas visões base, um número de pontos de referência correspondente, os quais estão localizados em pontos de descontinuidade de limites, bordas e outras características proeminentes, são selecionados. Quando um número apropriado de pontos de referência foi selecionado, o próximo passo é a utilização da triangulação de Delaunay, com restrições, com a finalidade de produzir malhas triangulares consistentes e correspondentes para todas as imagens. Esse processo pode ser realizado durante o estágio offline de treinamento. O sistema de reconhecimento envolve escolher os coeficientes apropriados a partir da Linear Combination of Views (LCV), sintetizando assim uma imagem, a qual é comparada a imagem alvo utilizando uma métrica de similaridade. O LCV, é uma técnica a qual pertence a teoria geral de tensores tri e multi focais, ou Algebraic Function of View (AFoV), e fornece uma forma de lidar com a variação da pose de um objeto devido a mudanças de ponto de vista. Uma vez que comparam-se duas imagens sobre todos os *pixels*, ou uma métrica de dissimilaridade tal como Sum of Squared differences (SSD) ou uma medida de similaridade tal como coeficientes de correlação cruzada podem ser utilizados. No trabalho apresentado, utilizou-se os coeficientes de correlação cruzada, pois quando aplicado a toda a imagem ele é invariante a transformações fotométricas. A seleção dos coeficientes do LCV é determinada pela maximização dos coeficientes de correlação cruzada. Essencialmente, está sendo proposto um sistema de correspondência de *templates* flexível, no qual é permitido a deformação do *template* no espaço LCV até que este corresponda a imagem alvo. Os experimentos realizados foram conduzidos em imagens sintéticas e reais com ângulos variáveis da câmera. As imagens sintéticas foram geradas através de fotografias 2D de um objeto 3D. As imagens reais foram retiradas da base de dados CMU PIE [473]. Os resultados do método apresentado foram demonstrados em termos de taxa de reconhecimento.

O trabalho proposto por Vázquez [474] [475] apresenta um método com base em visões, e em alguns aspectos biológicos da visão de crianças nos estágios iniciais da vida, para o reconhecimento de objetos 3D. Os aspectos biológicos da visão da criança que são utilizados como base, estão relacionados com a resposta a baixas frequências nos estágios iniciais de vida e em algumas conjecturas a respeito de como uma criança detecta características sutis de um objeto. Com base no aspecto de resposta em baixas frequências, inicialmente, o método proposto utiliza um filtro passa baixo para remover componentes de frequências mais altas da imagem. Depois, características sutis da imagem são detectadas através de uma seleção randômica de pontos estimulantes. Por fim, como dispositivo de aprendizagem, utiliza-se o Dynamic Associative Memory (DAM) para aprender as características e realizar o reconhecimento dos objetos. Um DAM pode ser construído em

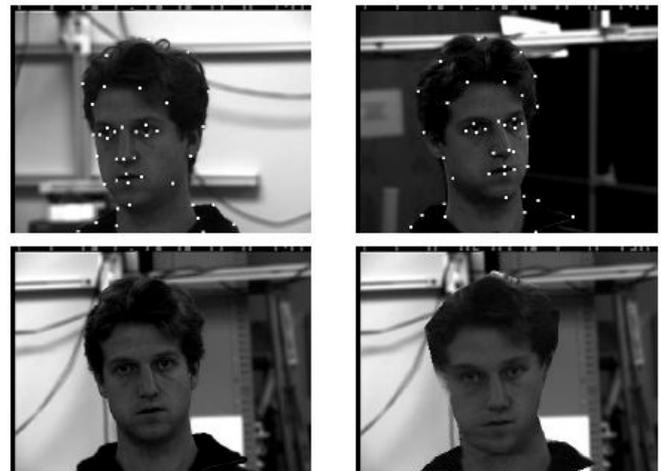


Fig. 151. Representação do método proposto por Zografos. Figura extraída de [472].

três etapas: a primeira etapa, transforma o conjunto fundamental de associações em padrões codificados e decodificados. A segunda etapa, computa uma versão simplificada dos padrões de etapa por meio de uma equação. A terceira etapa, constrói as sinapses da memória associativa em termos dos padrões codificados com outra equação. Após o treinamento utiliza-se o DAM para reconhecer as imagens de entrada com base nos padrões associados. Para testar a acurácia do método proposto utilizou-se a base de dados COIL 100 [23] com um conjunto de treinamento de 100 imagens e um conjunto de teste com 20 imagens para cada objeto em diferentes ângulos.

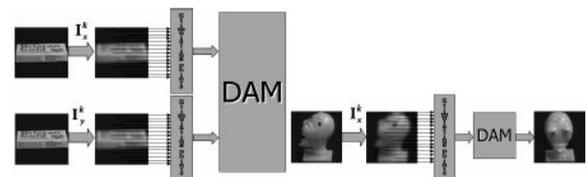


Fig. 152. Representação do método proposto por Vázquez. Figura extraída de [474].

Luciw [476] apresenta um mecanismo de Topographic Class Grouping (TCG), o qual explica como as conexões *top-down* influenciam o tipo de detector de características (neurônios) desenvolvido e sua colocação no plano neuronal. As conexões *top-down* impulsionam as variações no plano neuronal entre as direções das classes durante a etapa de treinamento. O primeiro resultado deste espaço de entrada *top-down* impulsionado é a facilitação do surgimento de detectores de características que são puros, medidos estatisticamente pela entropia média do desenvolvimento dos neurônios. Os neurônios relativamente puros são mais abstratos, isto é, caracterizam informações de entrada específicas a classe, resultando em um melhor resultado de classificação. O segundo resultado é o aumento da distância entre amostras de entrada que pertencem a diferentes classes, resultando em uma separação maior dos neurônios pertencentes a diferentes classes. Portanto, neurônios que respondem para a mesma classe ficam

relativamente próximos. Isto resulta em um TCG, medido estatisticamente por uma pequena dispersão intra-classe das respostas, quando o plano neural possui um tamanho fixo. A rede neural utilizada para desenvolver o mecanismo proposto é a Multi-layer In-place Learning Network (MILN) [477], dado que esta rede requer aprendizagem *in-place*, o que significa que cada neurônio deve aprender por conta própria, enquanto interagindo com outros neurônios, sem a necessidade de um aprendedor fora do neurônio. Para os experimentos do mecanismo proposto, utilizou-se a MSU Dataset e o normalized-centered NORB dataset [478], inicialmente mostrando os resultados de classificação com base em duas redes treinadas, uma com conexões *top-Down* excitatórias e outra sem. Depois, na base de dados NORB, comparou-se o resultado obtido pela rede MILN, com e sem TCG, com o método proposto em [478] em termos do resultado de erro.

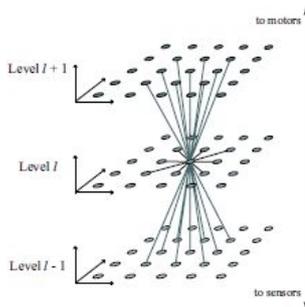


Fig. 153. Representação do método proposto por Luciw. Figura extraída de [476].

O trabalho apresentado por Polat [479] utiliza algoritmos genéticos (AG) e general regression neural network (GRNN) [480] para o reconhecimento de padrões com base em poses/visões dos objetos 3D, aplicável também para o reconhecimento de dígito escrito à mão, sem a extração de características. No estudo em questão o algoritmo genético é utilizado para otimizar o GRNN para o reconhecimento de padrões. Desta forma, o primeiro passo é a geração da população inicial de indivíduos, a qual é gerada aleatoriamente da base de dados. Depois, um valor de *fitness*, que representa uma medida de adaptação ao meio, é calculada. O próximo passo é a seleção dos indivíduos, a qual usa um esquema de seleção por roleta, que particiona a roleta conforme a probabilidade de cada indivíduo, calculada com base na soma dos valores de *fitness* da população, roda a roleta n vezes, onde n é o número de indivíduos da população, e a partição onde o ponteiro parar representa o indivíduo sendo selecionado. Por fim, mutação e *crossover* são realizados nos indivíduos para gerar a nova população e esse processo continua até a condição de parada ser alcançada. A saída do AG é então utilizada para treinar a GRNN. Para os testes 10 objetos em 72 poses para cada objeto foram utilizadas. No caso, 720 instâncias divididas em 10 classes e, para o conjunto de dados em questão (COIL-100 [23]), apenas 6 instâncias de cada objeto foram necessárias para o conjunto de treinamento. Os testes realizados compararam a abordagem proposta com o GRNN sem a otimização com AG em termos de taxa de reconhecimento.

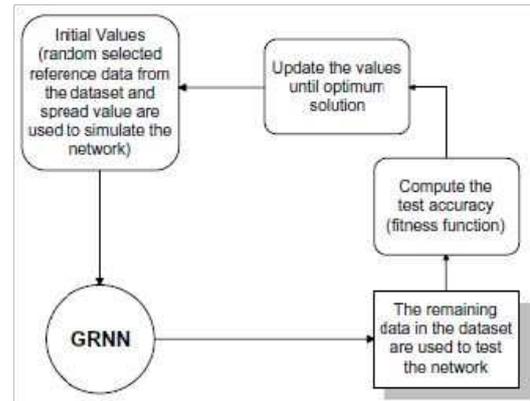


Fig. 154. Representação do método proposto por Polat. Figura extraída de [479].

Uma abordagem heurística para o reconhecimento de objetos 3D, através da consideração de múltiplas projeções 2D do objeto de interesse, é apresentada por Dimov [481]. Nesta abordagem, a identificação do objeto é interpretada como um problema convencional de Content Based Image Retrieval (CBIR). Uma imagem de entrada arbitrária de um dado objeto é tratado como uma amostra de busca dentro de uma base de dados grande com conjuntos de imagens, isto é aparências, de um número suficientemente grande de visões do objeto. Um dado objeto na frente de uma câmera é considerada um objeto 3D dinâmico representado por uma série de projeções 2D de aparências (imagens estáticas da câmera). Se um parte apropriada destas imagens, ou imagens similares a estas, estão armazenadas em uma base de dados de imagens, então pode-se procurar, nessa base de dados de imagens, a imagem mais próxima (mais similar) a imagem de entrada. Além disso, pode-se localizar um série de imagens ordenadas em uma ordem descendente a similaridade com a imagem de entrada. Para alcançar isto, a base de dados de imagem necessita ser construída como um dicionário de objetos com uma grande quantidade de imagens representativas do objeto. Para assegurar esta operação em tempo real, o tempo de comparação da imagem de entrada com todas as imagens do dicionário deve ser rápida o suficiente. Isto é possível se for utilizado um sistema CBIR rápido o suficiente para acessar uma grande base de dados de imagens. Tal método CBIR existe e está presente com o sistema EFIRS (Effective and Fast Image Retrieval System) desenvolvido em [482]. Com esta parte desenvolvida, o problema torna-se a acumulação de imagens representativas o suficiente na base de dados de imagem, isto é, montar o dicionário com imagens representativas da aparência do objeto, para encarar o problema já conhecido de aprendizagem e classificação. Toda uma heurística de aquisição de imagens, por uma câmera, de diferentes pontos de vista e tratamento para encontrar os quadros representativos é descrito detalhadamente. Os experimentos realizados com a heurística proposta tiveram como base as estrutura da base de dados de imagens do EFIRS e o método de acesso utilizado foi o PFWT [482] aplicados a dois testes para o reconhecimento de face e de poses de mãos. Os resultados dos testes foram demonstrados

em termos de erro médio e taxa de reconhecimento.

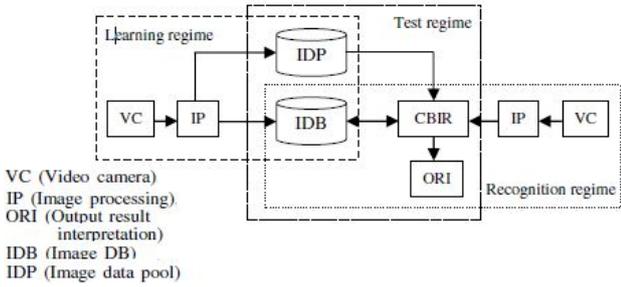


Fig. 155. Representação do método proposto por Dimov. Figura extraída de [481].

Um método para escolher a melhor sequência de visões para o reconhecimento de objetos 3D é apresentado por Urdiales [483]. O algoritmo de planejamento de visões depende de encontrar os pontos de diferença máxima entre o mapa de clusters de candidatos em potencial e, então, encontrar o melhor caminho para visitá-las em ordem. O método funciona da seguinte forma: primeiro, adquire-se uma visão inicial e encontra-se quais objetos na base de dados apresentam um vetor cujo as diferenças ao vetor de entrada está abaixo de um limiar fixado. Para aprimorar a eficiência do processo de reconhecimento, é aconselhado capturar uma segunda visão do objeto. Então, compara-se o mapa de clusters dos candidatos com o propósito de reduzir o número de candidatos existentes. Depois disso, calcula-se a diferença entre cada dois mapas e acumula-se a diferença para conferir em quais pontos eles se diferem mais. Afim de evitar problemas de aprendizagem em diferentes visões, quando a segunda visão é adquirida, realinha-se todos os candidatos, utilizando um vetor da visão 1 para a visão 2 em vez de uma única. Neste ponto, os mapas estão corretamente alinhados e o reconhecimento com base em Hidden Markov Model (HMM) pode ser aplicado as visões com as diferenças máximas. Para escolher a sequência mais eficiente, utiliza-se a solução por força bruta do problema do caixeiro viajante, porque o número de visões para capturar é sempre menor que 10. No caso as visões a serem capturadas são as cidades e as distâncias são calculadas sob a esfera de visão. Alguns resultados do método proposto são apresentados para objetos complexos como por exemplo um cubo. Todo processo de seleção de visões e como a probabilidade do HMM evolui para decidir qual o modelo correto é relatado.

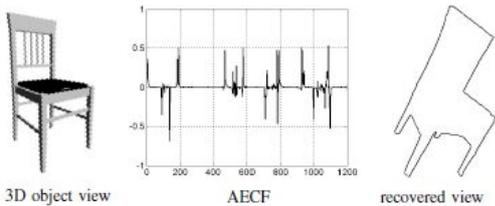


Fig. 156. Representação do método proposto por Urdiales. Figura extraída de [483].

O trabalho proposto por Bo Pang [484] mostra uma forma de descrever modelos 3D em um sistema de reconhecimento

de objetos 3D. Os modelos 3D são descritos por uma série de imagens 2D projetadas. A fusão de características com base em 3 características de formato é implementada para definir as propriedades da imagem e então a aprendizagem múltipla, é utilizada na clusterização de características para uma seleção mais eficiente de visões representativas. Um sistema de reconhecimento de objetos 3D foi construído para testar a performance da forma de representação apresentada. Esse sistema pode ser descrito por duas partes (treinamento e reconhecimento). Na parte de treinamento, um conjunto de imagens 2D projetadas são selecionadas para efetivamente representar cada modelo 3D. Primeiro, cada modelo 3D é igualmente amostrado na esfera de visão, gerando 162 imagens projetadas. Depois, extrai-se características de cada imagem projetada, utilizando Zernike descriptor, Fourier descriptor e trace feature, no conjunto de visões, e implementa-se uma seleção efetiva de visões com fusão de características e clusterização por aprendizagem múltipla, obtendo assim visões chave a partir da visões projetadas. A distância geodésica do Isomap [485] é utilizada para converter o espaço de características em um espaço múltiplo e uma clusterização hierárquica aglomerativa é aplicada no espaço múltiplo para gerar uma visão representativa. Na parte de reconhecimento, assinaturas extraídas das imagens de entrada são comparadas com os dados de cada modelo na base de dados e suas distâncias são tratadas com uma pontuação de dissimilaridade dos objetos de entrada e saída. O objeto com a menor pontuação torna-se o resultado da busca. Os experimentos foram realizados utilizando a base de dados Princeton Shape Benchmark [39] e os resultados avaliados em termos de taxa de reconhecimento comparando com outros métodos.

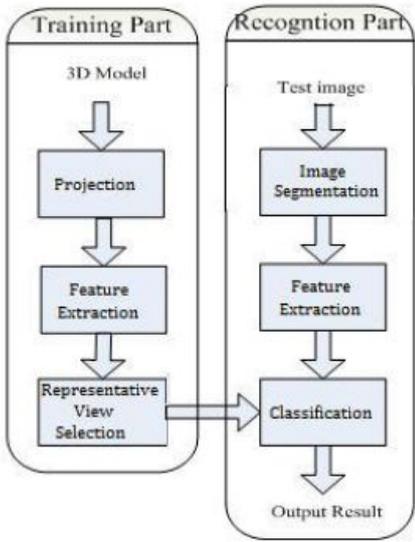


Fig. 157. Representação do método proposto por Bo Pang. Figura extraída de [484].

Ulrich [486] apresenta uma abordagem para o reconhecimento de instâncias de objetos 3D em uma única imagem de câmera e para determinar suas poses 3D. Inicialmente, um modelo hierárquico é gerado com base apenas nas informações geométricas de um modelo CAD 3D do objeto. Durante

a geração do modelo hierárquico, apenas as informações geométricas do objeto, importantes para o processo de reconhecimento, são incluídas no modelo hierárquico. A principal tarefa da geração do modelo é derivar uma hierarquia 2D de visões do objeto que possa ser usado para encontrar o objeto eficientemente em uma imagem. As diferentes visões do objeto são criadas automaticamente através do posicionamento de câmeras virtuais ao redor do objeto 3D e pela projeção do objeto no plano de imagem de cada câmera virtual. A derivação dos grafos de aspecto com base na similaridade, isto é, a amostragem das visões dentro do alcance da pose, é determinado automaticamente durante a geração do modelo hierárquico. Para aumentar a velocidade de reconhecimento, os modelos 2D são gerados em múltiplos níveis de uma pirâmide de imagens. Devido ao fato que os níveis mais altos da pirâmide permitem uma amostragem grosseira das visões [487], a computação das visões é realizado separadamente para cada nível da pirâmide. Durante a fase de reconhecimento, o modelo hierárquico gerado é utilizado para reconhecer o objeto 3D em uma única imagem de câmera e determinar a pose do objeto em relação ao sistema de coordenadas da câmera. Inicialmente, uma pirâmide de imagens é construída com a imagem de entrada. O reconhecimento inicia nos nível mais alto da pirâmide, onde ao menos um modelo 2D esta disponível. Todos os modelos 2D neste nível da pirâmide são analisados através do cálculo da medida de similaridade, proposta em [488], entre os modelos 2D das visões e o nível atual da pirâmide de imagem. Os modelos 2D são rotacionados e escalonados em passos discretos dentro do alcance necessário. A medida de similaridade é computada em cada posição dos modelos 2D rotacionados e escalonados na imagem. As poses 2D com correspondências que ultrapassam o limiar de similaridade são armazenados em uma lista de candidatos. Nos próximos níveis inferiores da pirâmide, todos os modelos 2D que não possuem um nodo pai são analisados da mesma forma que as visões nos níveis mais altos da pirâmide. Adicionalmente, os candidatos que foram encontrados nos níveis anteriores da pirâmide são refinados. Este refinamento é executado através da seleção de todas as visões filhas na árvore e computação da similaridade entre os modelos 2D das visões filhas e o nível atual da pirâmide de imagens. Este processo é repetido até todos os candidatos terem sido rastreados até o nível mais baixo da pirâmide. Os demais passos se resumem na conversão da pose de 2D para 3D, com base nos candidatos encontrados e no refinamento das poses encontradas. Os experimentos realizados analisaram a velocidade e a acurácia do método proposto com base no erro absoluto resultante.

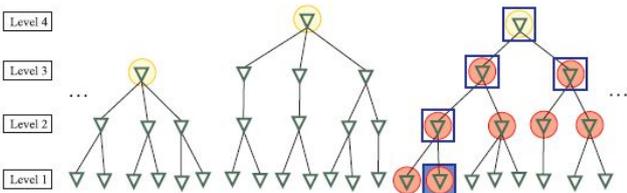


Fig. 158. Representação do método proposto por Ulrich. Figura extraída de [486].

O trabalho apresentado por Elons [489] propõe uma técnica para lidar com as variações de pose no processo de reconhecimento de objetos 3D. A técnica proposta utiliza pulse-coupled neural network (PCNN) para gerar assinaturas únicas através de imagens capturadas em diferentes ângulos. A técnica proposta é composta de duas partes: construção do modelo e reconhecimento. Na parte de construção do modelo, no caso aplicado para o reconhecimento da linguagem de sinais, cada sinal de mão a ser reconhecido deve ser colocado em uma posição estável no centro de uma mesa circular giratória e com duas câmeras, uma posicionada no 0 graus e outra posicionada no 90 graus, são capturadas as imagens do objeto. Cada câmera é movida 5 graus e captura-se uma nova imagem para as duas câmeras. Esse processo se repete até a câmera localizada no 0 graus chegar na posição 90 graus. Desta forma cada objeto terá 19 imagens de cada câmera. Estas imagens são divididas em dois grupos 10 imagens de cada câmera para o treinamento e 9 para testes. O conjunto de treinamento é utilizado para construir o modelo do objeto 3D. As imagens capturadas são utilizadas para produzir uma assinatura através do PCNN. As duas assinaturas são então ponderadas e combinadas linearmente para produzir uma assinatura 3D da imagem. O próximo passo é a avaliação da qualidade da assinatura através de uma função de aptidão com a finalidade de construir características 3D com qualidade otimizada. Após a montagem da base de modelos utiliza-se uma rede neural, Multilayer perceptron, para aprender e posteriormente classificar as imagens de entrada, as quais passam pelo mesmo processo de geração de uma assinatura 3D. Os testes realizados para o método proposto foram com base na língua de sinais dos árabes e os resultados obtidos foram comparados em termos de acurácia, descrevendo experimentalmente a taxa de acurácia obtida para cada número de épocas e número de nodos utilizados no treinamento e a função de aptidão utilizada. Os resultados obtidos também foram comparados com outros sistemas de reconhecimento de sinais.

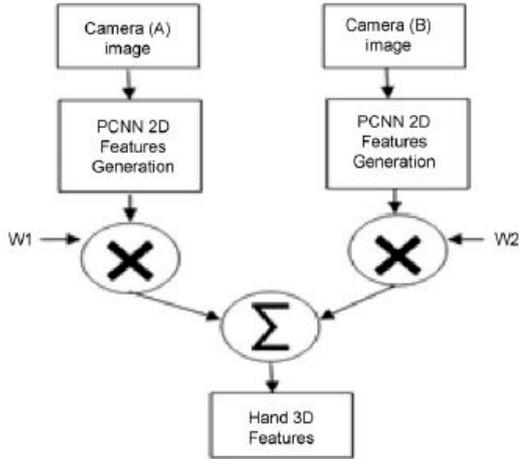


Fig. 159. Representação do método proposto por Elons. Figura extraída de [489].

Domingo Mery [490] propõe o uso de um método automático com base em múltiplas imagens de raio-x de difer-

entes visões para o reconhecimento de objetos regulares com formatos e tamanhos definidos. O método consiste de duas etapas: análise monocular, para obter possíveis detecções em cada visão de uma sequência, e análise de múltiplas visões, para reconhecer os objetos de interesse utilizando correspondências em todas as visões. O método proposto também possui dois estágios: o estágio offline serve para a aprendizagem do modelo utilizado no reconhecimento e para a estimativa do modelo geométrico de múltiplas visões utilizado para associação de dados. O estágio online é realizado com a finalidade de reconhecer o objeto de interesse em uma sequência de imagens de teste. As imagens são adquiridas com rotações do objeto sendo testado, por um ângulo informado e estas analisadas através das análises monocular e de múltiplas visões. A análise monocular é realizada sobre cada imagem da sequência de imagens de teste. Inicialmente, o objeto como um todo é segmentado do plano de fundo utilizando limiarização e os operadores morfológicos. Depois, pontos chave SIFT, ou outros descritores, são extraídas da porção segmentada. O descritor de cada ponto chave é classificado utilizando um classificador treinado no estágio offline. Todos os pontos chave classificados com a classe c , onde c é a classe de interesse, são selecionados. Devido a problemas de classificação errada dos pontos chave, pontos chaves vizinhos são clusterizados no espaço 2D utilizando o algoritmo de mean shift. Apenas os clusters que possuem um número grande suficiente de pontos chave são selecionados. Esse pontos são denominados *detected monocular keypoints*. A análise de múltiplas visões tem como entrada os *detected monocular keypoints* e realiza o reconhecimento do objeto de interesse em três passos. O primeiro passo é a associação dos dados, no qual encontram-se correspondências para todos os *detected monocular keypoints* em todas as imagens consecutivas. O passo subsequente é a análise 3D, na qual para cada par de pontos chave correspondidos na imagem atual e nas imagens consecutivas, estabelecidos na etapa anterior, um ponto 3D é reconstruído utilizando as matrizes de projeção dos modelos geométricos previamente calculados. De forma similar a detecção monocular, pontos 3D vizinhos são clusterizados em um espaço 3D utilizando o Mean shift e apenas aqueles que possuem um número grande suficiente de pontos 3D são selecionados. Por fim o último passo é a análise final, na qual para cada cluster 3D selecionado, todos os pontos 3D reconstruídos pertencentes ao cluster são reprojatados em todas as imagens da sequência utilizando a matriz de projeção do modelo geométrico. Os descritores extraídos dos pontos chave localizados perto destes pontos reprojatados são classificados individualmente utilizando um classificador. O cluster será classificado como classe c' se existe um grande número de pontos chave individualmente classificados como c' e esse número representa a maioria dentro do cluster. Esta estratégia de voto majoritário pode superar o problema falsas detecções monoculares quando a classificação da minoria falha. Um cluster pode ser classificado erroneamente se a parte que esta sendo reconhecida estiver oclusa por uma parte de outra classe. Neste caso, existiram pontos chave no cluster atribuídos para ambas as classes. Contudo, é esperado que para a grande maioria dos pontos chave seja atribuída a classe verdadeira

se existir um número pequeno de pontos chave classificados erroneamente. Os experimentos realizados visaram classificar os objetos teste em três classes e a performance do método foi demonstrado em termos de taxa de precisão e lembrança.

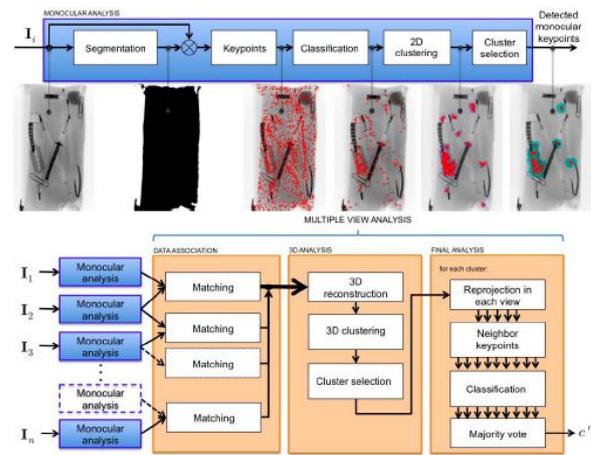


Fig. 160. Representação do método proposto por Domingo Mery. Figura extraída de [490].

Um modelo de rede neural do fluxo visual ventral para o reconhecimento e classificação de objetos 3D é proposto por Efremova [491]. Um estudo similar foi realizado pelo autor em [492]. O modelo representa uma hierarquia de módulos, a qual se assemelha as áreas V1-V4 e o córtex temporal inferior. A arquitetura do modelo tem como base o conceito da rede neural denominada Self-Organized Map (SOM). Este tipo de rede neural é treinado utilizando procedimentos de aprendizagem não supervisionada para produzir um mapa bidimensional do espaço de entrada das amostras de treinamento. Mais especificamente, foi utilizado o conceito do SOM associada a rede Radial Basis Function (RBF). O algoritmo de aprendizagem é uma modificação do algoritmo do SOM e consiste de quatro processos: processo de cálculo, processo competitivo, processo cooperativo e processo adaptativo. No primeiro passo os pesos são definidos randomicamente em um intervalo de 0 a 0.5. Na para de avaliação, calcula-se todas as saídas para todas as entradas em uma única unidade RBF de acordo com uma regra apresentada. No processo competitivo, o módulo que minimiza o erro é considerado o módulo vencedor. No processo cooperativo, os pesos são calculados através de uma função de vizinhança. No processo adaptativo, todos os módulos são atualizados através do algoritmo de aprendizagem por back-propagation. Essa etapa de aprendizagem é repetida até que a rede alcance um estado estável. Após o treinamento da arquitetura apresentada, utiliza-se a mesma para o reconhecimento de objetos 3D. Nos experimentos realizados, foram utilizados objetos gerados computacionalmente (com a aparência de arames, para o primeiro experimento, e com vários formatos geométricos para o segundo experimento). Um total de 10 objetos foram utilizados, no primeiro experimento, cada um consistindo em seis pontos sequencialmente conectados com traços. Três tipos de objetos de arame com diferentes números de segmentos retos foram utilizados. O número de segmentos retos teve a intenção de adicionar similaridade para os objetos.

Dez estímulos 3D foram apresentados a rede em projeções a cada grau de rotação. Para cada projeção, as coordenadas 2D das projeções dos objetos 3D foram as entradas da primeira camada da rede descrita. A saída da rede formou uma mapa de similaridade, em que três regiões podem ser claramente distinguidas. Portanto, a distribuição da ativação no mapa resultante representa a classificação dos objetos pelo número de segmentos retos. Logo, a arquitetura do SOM de módulos RBF consiste de basicamente dois níveis (denominados RBF e SOM níveis). No nível SOM o resultado do experimento é a habilidade da rede de gerar a similaridade dos objetos de acordo com certos princípios (no caso número de segmentos retos). No nível RBF o resultado é o padrão de ativação de uma única célula durante a apresentação do objeto. Para testar a arquitetura proposta com objetos mais complexos, adicionou-se uma etapa de pré-processamento composta de módulos que se assemelham ao fluxo visual ventral do córtex visual de primatas. Com esta extensão, mostrou-se que o modelo tem capacidade de realizar o reconhecimento e classificação de objetos 3D.

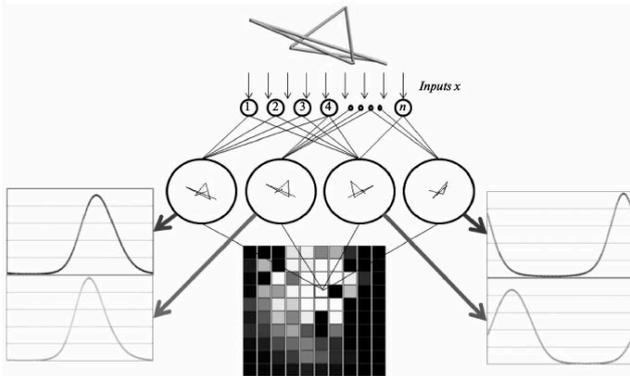


Fig. 161. Representação do método proposto por Efremova. Figura extraída de [491].

A abordagem descrita por Guan pang [493] mostra uma forma de reconhecimento de objetos 3D em nuvem de pontos que projeta a nuvem de pontos 3D em várias imagens de profundidade de diversos pontos de vista, transformando o problema de reconhecimento 3D em uma série de problemas de detecção 2D. Para assegurar que a informação 3D original não foi perdida, a projeção 3D para 2D é realizada em múltiplos ângulos de visões, igualmente escolhidos em uma esfera. A informação de profundidade é utilizada quando projeta-se as imagens 2D para cada visão, e mantidas para depois reprojeter de volta em 3D para a fusão dos resultados 2D. Após a nuvem de pontos de entrada ser projetada em imagens 2D de múltiplas visões, cada visão é utilizada para localizar o objeto alvo. Por fim, todos os resultados de detecções 2D são reprojetaadas de volta para o espaço 3D para uma estimativa combinada da localização do objeto 3D. A detecção do objeto 3D acontece apenas se várias detecções 2D reprojetaadas acontecerem em uma região 3D próxima, filtrando assim as detecções das múltiplas visões. A confiança final é computada como o peso médio através de todas as visões, filtradas por um limiar empírico. O método proposto

foi testado em uma combinação de dados industriais e de rua (3D Keypoint detection benchmark [81], UWA 3D Object Dataset [138], CMU Oakland 3D Point Cloud Dataset [494] e Washington Urban Scenes 3D point Cloud Dataset [495]) contendo vários tipos de objetos e condições de cena. Para avaliar a performance o algoritmo foi comparado com outros descritores de pontos de nuvens 3D incluindo FPFH, Spin images e SHOT e com um método desenvolvido anteriormente pelo autor[331] em termos de tempo de execução e curvas de precisão e lembrança.

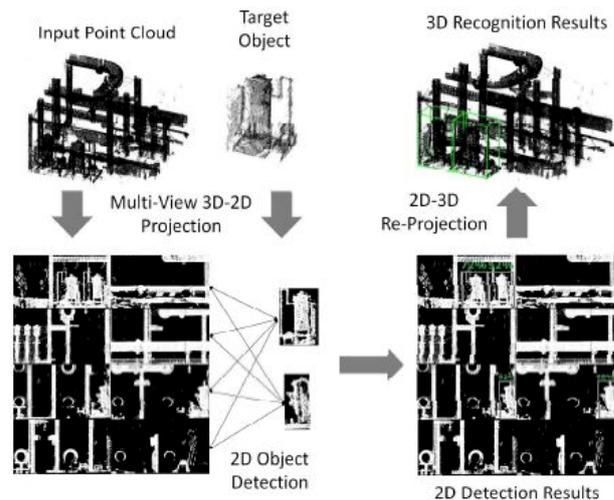


Fig. 162. Representação do método proposto por Guan Pang. Figura extraída de [493].

Um método online com base em múltiplas-visões, o qual integra informações do ambiente capturado, fundindo saídas individuais do reconhecimento com uma única visão, é apresentado por Faulhammer [496]. O método proposto utiliza dados RGB-D e transfere, continuamente as hipóteses construídas em vários pontos, em um *framework* para juntar a máxima quantidade de informações para todos os objetos na cena. Contrário aos métodos de *batch* existentes [497], [498], a abordagem proposta permite ao sistema melhorar o reconhecimento online, ou seja melhora o reconhecimento a cada nova observação, utilizando uma representação dinâmica em grafo do ambiente observado. O reconhecedor por visão única gera para cada cena de nuvem de pontos um conjunto de objetos candidato (hipótese) potencialmente presentes na cena. Essas hipóteses são obtidas utilizando o sistema de reconhecimento por visão única proposto em [444]. Depois, de um refinamento da posição com o ICP, um estágio final de verificação retorna um subconjunto de hipóteses que melhor representam a cena em relação a um critério global [444]. Para explorar a informação proveniente de múltiplas visões de um ambiente, cria-se um grafo com os vértices representando informações de uma única visão e arestas conectando as visões. Para cada foto da cena, um vértice é conectado a outro vértice que compartilham uma hipótese de objeto em comum pelas arestas. Um grafo totalmente conectado é obtido se existem suficiente hipóteses de objetos em comum para cada vértice. Para evitar vértices isolados ou, em caso de uma esti-

mativa fraca da pose do objeto, arestas adicionais são criadas por meio de características visuais da cena. Em particular, cada visão é correspondida utilizando uma estratégia de primeiro vizinho mais próximo em relação a suas respectivas características SIFT, fornecendo um conjunto de correspondência entre ambos os frames dois quais uma transformação rígida foi estimada. Utilizando a representação de grafo, as visões individuais são fundidas em uma representação global da cena através da estimativa das poses das câmeras. Dado que estimativas erradas de pose diminuiriam a qualidade do ambiente reconstruído, isola-se a visão se o peso da aresta for maior que um limiar pré-definido, ignorando efetivamente informações de múltiplas visões neste ponto e construindo subgrafos para observações futuras. Escolhendo o sub-grafo contendo as observações mais recentes, constrói-se uma 3D+RGB reconstrução da cena assim como um conjunto de modelo de múltiplas visões candidato pelo percorrimento deste subgrafo e composição da hipótese com todos os conjuntos de múltiplas visões verificadas geradas nas observações anteriores. Por fim, verifica-se a hipótese fundida contra o ambiente reconstruído por uma etapa de verificação 3D [497]. A saída final é um conjunto de hipóteses verificadas, as quais substituem as hipótese por visão única armazenadas no vértice. O método proposto foi testado em duas bases de dados disponíveis, TUV[499] e Challenge[497], e comparado com um sistema de reconhecimento com única visão assim como aos resultados reportados em [500], [501].

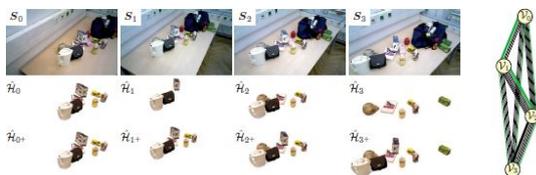


Fig. 163. Representação do método proposto por Faulhammer. Figura extraída de [496].

Uma método para determinar as visões salientes de objetos 3D é proposto por Yi-Chen Chen [502]. O método proposto possui dois estágios. Inicialmente, visões são extraídas, cortadas e redimensionadas de uma sequência de vídeos. Depois, no primeiro estágio, as pontuações de fronteiras são computadas utilizando uma métrica de espalhamento baseada em dispersão para estimar as Boundary Representative Views (BRVs), um dos dois tipos de categorias de visões salientes definidas, e determinar as side view classes. Uma side view class é definida como um conjunto de todas as visões de um correspondente lado do formato, enquanto boundary views class refere-se a visões onde dois ou mais lados podem ser vistos simultaneamente. No segundo estágio, para cada lado, um conjunto de side representative views (SRVs), que melhor representam um correspondente lado, é selecionado através da minimização do erro de representação. Resumidamente, as visões salientes consistem em BRVs e SRVs as quais são computadas através da maximização da pontuação de fronteira e da minimização do erro de representação, respectivamente. Com base nestas visões salientes, são montados dicionários, denominados view-dependent dictionaries (VDDs), os quais

codificam informações da geometria através de visões e representações do objeto. Estes VDDs são utilizados na parte de reconhecimento e recuperação de objetos 3D. O processo para descrição, representação, reconhecimento e recuperação dos objetos pode ser descrito em duas etapas. A primeira etapa é o treinamento, na qual os objetos na base de dados são descritos em termos de visões salientes e classes de side views e estas utilizadas para montar os VDDs. A segunda etapa é a parte de reconhecimento, a qual utiliza os VDDs e um critério de vizinho mais próximo para classificar uma visão alvo na classe mais apropriada. Os experimentos realizados com a abordagem proposta utilizaram vídeos, nos quais cada *frame* representa uma visão do objeto 3D, e os resultados foram avaliados em termos de taxa de reconhecimento e comparados com outras abordagens .

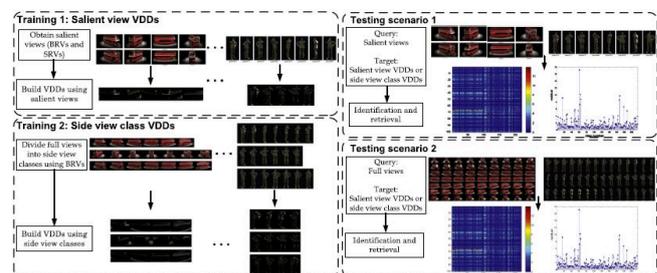
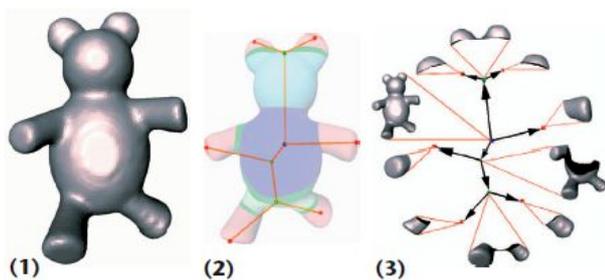


Fig. 164. Representação do método proposto por Yi-Chen Chen. Figura extraída de [502].

C. Grafos

Marini [503] propõe um método para construir protótipos criativos de classes de objetos 3D descritos por assinaturas estruturais codificadas em um grafo de atributo, o qual sumariza os aspectos do formato topológico e geométrico. A metodologia utilizada propõe atribuir para cada classe de modelos 3D um protótipo de formato que sumariza as características mais relevantes ocorrendo nos membros daquela classe. No *framework* proposto, os objetos 3D são descritos por descritores estruturais, os quais explicitamente codificam as características relevantes, como arestas e vértices de um grafo. Esta representação explícita permite aos desenvolvedores utilizar operações de edição entre os elementos dos descritores estruturais (tal como vértices, arestas e seus atributos) e portanto entre as subpartes 3D relevantes de um objeto. A metodologia proposta compreende dois passos: computação do subgrafo comum entre membros de uma classe e computação de um conjunto de operações de edição adequadas. O grau de liberdade destes dois passos permite ao desenvolvedor dirigir a construção dos protótipos de acordo com a semântica que o usuário quer explorar. As operações de edição que devem ser consideradas são a adição e deleção de arestas e vértices do grafo e modificação de seus atributos, como formalizado pela noção de Error Tolerant Subgraph Isomorphism (ETSI)[504]. O modelo consultado é então classificado na classe representada pelo protótipo que mais se assemelha a ele. Formalmente dado um conjunto M de modelos 3D, cujos os membros estão agrupados em m classes, o modelo consultado é classificado

na classe que minimiza a distância, equação definida pelos autores. Como cada modelo é representado por um descritor estrutural, a distância entre o modelo consultado e a classe pode ser medida pela distância entre descritores estruturais. Desta forma, compara-se o descritor estrutural, codificado como grafo, da classe com o descritor estrutural do objeto consultado. Para validar a metodologia apresentada utilizou-se um conjunto de dados com 400 modelos 3D agrupados em 20 classes, cada uma consistindo em 20 elementos. Esse conjunto de dados foi formado de um conjunto contendo mais de 1000 modelos coletados de vários repositórios de modelos 3D (AIM@SHAPE repository, National Design Repository at Drexel University, the Caesar Data Samples e McGill 3D Shape Benchmark). Os resultados da validação foram demonstrados em termos de taxa de classificação.



A (1) A 3D model displayed with its structural descriptor. (2) The geometric attributes associated to the structural descriptor encoded as (3) directed graph.

Fig. 165. Representação do método proposto por Marini. Figura extraída de [503].

Shengping Xia [505] apresenta duas contribuições para os problemas relacionados aos métodos com base em características locais invariantes, como por exemplo o SIFT, pouco uso ou uso complexo das restrições geométricas e são confundidos quando detectam características que são superabundante. A primeira contribuição é o ranqueamento dos pontos SIFT (R-SIFT) utilizando saliência visual. O método para ranqueamento utiliza um procedimento de síntese da imagem para localizar pontos de características que são robustos a perturbações randômicas na imagem. Novas imagens são sintetizadas utilizando o método de Monte Carlo, o qual redimensiona, rotaciona e adiciona ruído a imagem original. As imagens sintetizadas são processadas para a extração de características e estas são correspondidas com as características extraídas da imagem original. A fração de pontos correspondidos corretamente para cada ponto SIFT na imagem original e o gradiente de magnitude são utilizados para cálculo da medida usada no ranqueamento dos pontos de característica. A segunda contribuição, utiliza o conjunto reduzido de características R-SIFT para construir um Class Specific Hyper Graph (CSHG), o qual utiliza de forma compreensiva as restrições do SIFT e da geometria global. Além disso, o CSHG captura de forma eficiente instâncias de aparência de múltiplos objetos. De forma resumida a iteração utilizando o CSHG pode ser descrita em três passos: primeiro, constrói-se o o grafo

G1 com base no conjunto de características R-SIFT. Depois, procura-se o melhor conjunto de grafos correspondentes do CSHG. Por fim, teste-se se o G1 é redundante ou não. Se não for, G1 será incrementado a estrutura do CSHG. Para múltiplos objetos, CSHGs são treinados separadamente e podem ser agregados juntos como um CSHG de múltiplos objetos. Para os experimentos realizados com as duas contribuições propostas, testou-se primeiro a efetividade do R-SIFT e então foram realizados experimentos com a modelagem e reconhecimento utilizando o CSHG, avaliando os resultados em termos de taxa de reconhecimento.

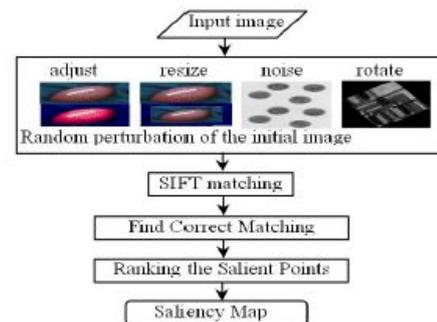


Fig. 166. Representação do método proposto por Shengping Xia. Figura extraída de [505].

Um método para particionamento de formatos 3D complexos em partes simples, voltado para a correspondência e reconhecimento de objetos 3D, é proposto por Aouada [506]. O particionamento dos objetos conta com a extração da topologia com os grafos Reeb dos objetos. Tirando vantagem das propriedades da teoria de Morse [507], detectam-se os pontos críticos da Global Geodesic Function (GGF). O uso da GGF está relacionado com a sua invariância a transformações isométricas. Para preservar a geometria dos objetos, escolheu-se utilizar curvas de nível ao invés de intervalos. Para realizar a correspondência dos objetos, propõe-se uma técnica com base em *kernel* para registrar grafos Reeb. Este posicionamento ótimo de grafos Reeb prepara para uma comparação em pares da geometria dos formatos primitivos. A comparação é realizada através de uma medida de distância que possui a mesma diretiva que a edit-distance para correspondência de formatos como definido em [508]. Logo, computa-se a pontuação de similaridade entre dois grafos Reeb com uma formulação, descrita pelos autores, associada a um *kernel* Gaussiano. Uma ilustração do funcionamento da técnica de particionamento é apresentada como um dos resultados dos prováveis teste realizados.

Noma [509] apresenta uma representação de formatos esparsa, a qual é testada no reconhecimento de objetos 3D e de dígitos manuscritos. Esta representação esparsa é alcançada através da representação dos padrões dos objetos como grafos e, para cada grafo, ocorre uma amostragem uniformemente espaçada de pontos do contorno do objeto, representando cada ponto por um vértice. De forma similar a [510], contornos poligonais são utilizados para aproximar aos contornos reais do objeto. Com a representação esparsa do formato computada, o próximo passo é definir uma métrica para realizar

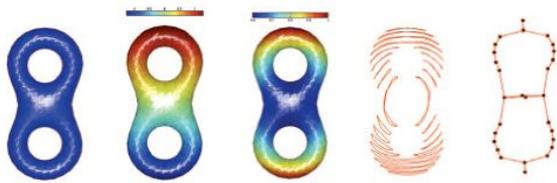


Fig. 167. Representação do método proposto por Aouada. Figura extraída de [506].

a correspondência do formato. A métrica proposta tem como base a soma das crenças computadas pelo método de belief propagation (BP), considerando apenas os vértices de entrada rotulados, e acrescenta uma penalidade em relação aos vértices sem correspondência. Nos experimentos realizados testou-se a representação de formato proposta acoplada ao classificador K-NN, demonstrando os resultados em termos de taxa de erro. Duas bases de dados foram utilizadas nos experimentos, uma para o reconhecimento de objetos 3D, COIL [23] e a outra para o reconhecimento de dígitos manuscritos, MNIST [511].

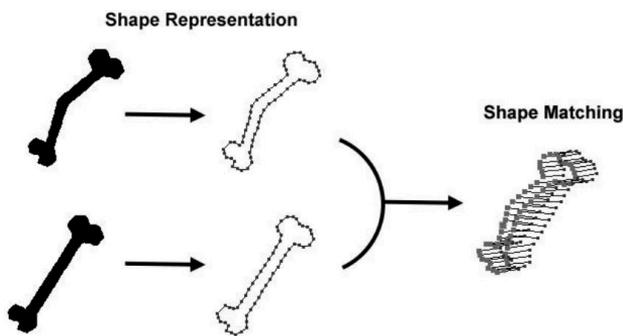


Fig. 168. Representação do método proposto por Noma. Figura extraída de [509].

Um método de multi-resolução para reconhecimento de objetos 3D, com base no modelo visual humano é apresentado por Huimin Ma [512]. O Sistema de reconhecimento proposto possui duas partes principais (a montagem da biblioteca de grafos de aspectos e o reconhecimento do objeto 3D pesquisado). A montagem da biblioteca de grafos de aspectos inicia com uma simplificação do modelo 3D original de acordo com o método de cálculo do Visual Resolution Limit (VRL). Depois, aplica-se um método de Viewpoint Space Partitioning aos modelos simplificados com diferentes níveis de resolução. Por fim, um processo adicional de clusterização pode ser realizado nos resultados particionados para obter representações por grafos de aspectos de múltiplas resoluções do modelo 3D. Para a parte de reconhecimento do objeto alvo, o mesmo é inicialmente segmentado da cena 2D. Depois, o reconhecedor utiliza o objeto alvo para o processo de correspondência entre o objeto alvo e as representações na base de dados, após uma série de fases de pré-processamento que incluem o detector de bordas (Canny) e transformações de translação, rotação e espelhamento. Para a comparação entre o objeto da cena e a representação na base, pode-se utilizar a distância de Hausdorff. Para os experimentos do

método proposto foram utilizados modelos 3D publicados pela universidade de Princeton [39], da qual 7 objetos foram escolhidos para os testes. Os testes executados analisaram o cálculo do VRL, a montagem da biblioteca de grafos de aspectos, reconhecimento por múltipla resolução e a comparação entre o método proposto e outras métodos de reconhecimento por múltiplas resoluções.

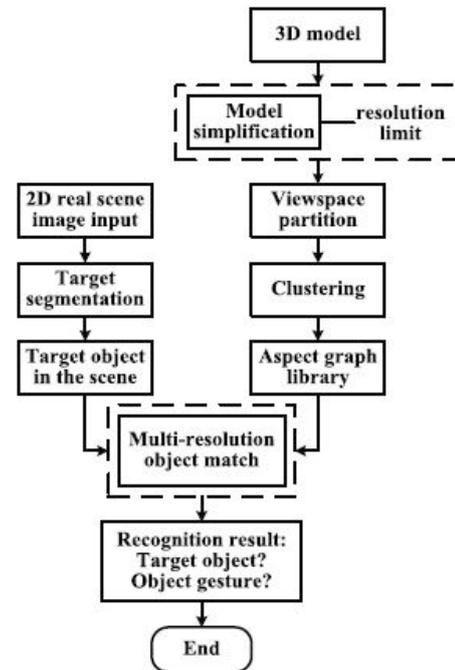


Fig. 169. Representação do método proposto por Huimin Ma. Figura extraída de [512].

Bonev [513] [514] avalia medidas estruturais de grafos para a classificação de objetos 3D. Inicialmente, cada objeto 3D é representado por uma malha de triângulos. Para cada objeto, três tipos de *Reeb graph* [515] são extraídos, um com base na distância geodésica e os outros dois, um com base na distância com o centro de massa e o outro com base na distância do centro da esfera que circunscreve a malha de triângulos. Apenas a informação estrutural dos grafo é utilizada para classificação. Para cada grafo, 9 diferentes medidas são calculadas (*Complexity Flow, Friedler, Adjacency Spectrum, Degrees, Perron-Frobenius, N.Laplacian Spectrum, Node Centrality, Commute Times 2 e Commute Times 1*). Estas medidas são transformadas em histogramas, após a normalização destas pelo volume do grafo. A *Commute time* é normalizada duas vezes (linearmente e quadraticamente). Dado que não há uma forma ótima para selecionar o número de *bins*, foram realizados experimentos com diferentes números de *bins* para cada medida. Todos os histogramas formam um *bag of features* com 540 características. O processo de seleção de características decide qual *bin* de cada medida e de qual grafo será descartado. O processo de seleção de característica estima a Mutual Information (MI) entre o conjunto de características e a classe, de forma a selecionar a característica que melhor representa a classe. Os experimentos foram realizados

em uma base de dados de formatos 3D [516], a qual consiste em 15 classes com 20 objetos por classe. Cada uma das 300 amostras é caracterizada por 540 características e tem uma classe rótulo. Os resultados foram avaliados com base nas medidas de erro e cada uma das 9 medidas foram avaliadas visando determinar qual é a mais importante, isto é, de maior influência e com qual combinação.

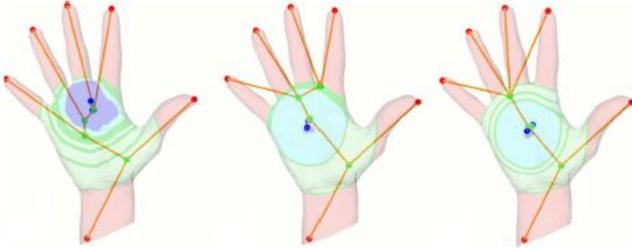


Fig. 170. Representação do método proposto por Bonev. Figura extraída de [513].

Kuk-Jin [517] apresenta um *framework* para o reconhecimento geral de objetos 3D com base na invariáveis das características locais e suas informações 3D obtidas com câmera estéreo. O primeiro passo do *framework* proposto, após a aquisição das imagens pelo par estéreo, é a extração de características locais. As características locais extraídas, por G-RIF[518] ou SIFT[47] da imagem da câmera da esquerda e da câmera da direita, são então correspondidas entre as imagens. Após encontrar as correspondências, computa-se a coordenadas 3D das características locais correspondidas em relação a câmera. A forma utilizada para representar o objeto 3D com informações 3D foi uma representação por grafo, mais especificamente um attributed relational graph (ARG) direcionado. Neste ARG, cada vértice corresponde a uma característica local correspondida e o atributo do vértice é o vetor descritor das características locais. Na parte de aprendizagem, os grafos dos objetos a serem reconhecidos são construídos e armazenados na base de dados. Então, as hipóteses de reconhecimento são geradas pela comparação dos grafos armazenados na base de dados com o grafo da imagem de entrada sendo reconhecida, através do cálculo de uma matriz de correspondência entre os grafos [519] e da escolha dos objetos, os quais suas matrizes de correspondência possuem um custo de correspondência menor que um limiar. Os candidatos selecionados são então verificados na última etapa, através da utilização das informações 3D, calculando uma transformação rígida via RANSAC. Se a verificação com os pontos 3D falhar, tenta-se verificar os candidatos selecionados utilizando homografia 2D, tal como nos métodos convencionais. Para verificar a performance do método proposto, construí-se uma base de dados de objetos 3D contendo 11000 imagens estéreo de 11 objetos. Para cada objeto, 200 imagens foram capturas com variações de iluminação, escala e oclusão. Os resultados foram demonstrados com uma variação do descritor sendo utilizado e do uso ou não do par de imagens em termos de correspondências corretas, falsos positivos e falsos negativos.

Mengjie Hu [520] propõe um método de reconhecimento

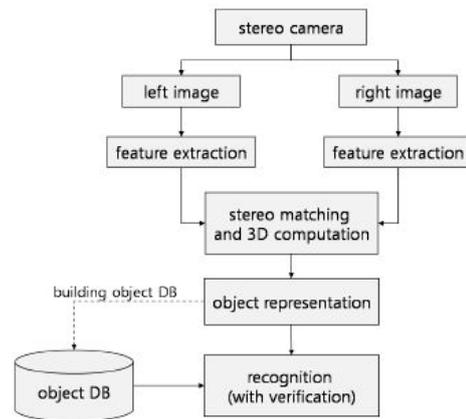


Fig. 171. Representação do método proposto por Kuk-Jin Yoon. Figura extraída de [517].

de objetos com base na representação de objetos 3D via *aspect graph aware*. Também é proposto um filtro de dois estágios para falsas correspondências 2D para 3D com base na representação proposta para alcançar pares de correspondências mais consistentes. O método de reconhecimento proposto é composto de dois estágios, offline e online. No estágio offline, a estrutura do motion method Bundler [521] é aplicada no conjunto de dados de objetos e os pontos do plano de fundo são removidos manualmente para obter o modelo de nuvem de pontos do objeto. Então, a representação por *aspect graph aware* é obtida pela análise da estrutura do modelo. Durante o estágio online, correspondências 2D-3D grosseiras são geradas por similaridade de características dos descritores (SIFT) entre os pontos extraídos da imagem e o modelo 3D. Então a visão mais provável é selecionada por votação de aspecto. Depois disto, um filtro de dois estágios para falsas correspondências grosseiras para remover falsas correspondências. Por fim, a estimativa da pose é determinada pelo RANSAC em associação com o EPnP [131] e a localização do objeto é obtida pela re-projeção do modelo 3D. A performance do método proposto foi avaliada em três aspectos: acurácia da localização 2D, acurácia da estimativa da pose e consumo de tempo. A base de dados utilizada foi a Weizmann Cars ViewPoint dataset [522] e os resultados foram comparados com outros métodos (Nearest Neighbour, Nearest Neighbour+RANSAC e deformable part model [295]) em termos dos três aspectos previamente citados.

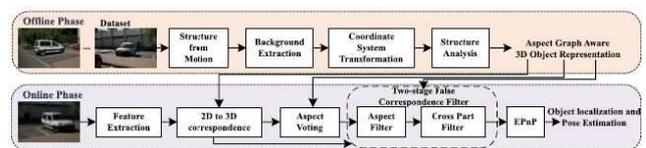


Fig. 172. Representação do método proposto por Mengjie Hu. Figura extraída de [520].

Madi [523] propõe um algoritmo para medir a distância entre objetos 3D representados por grafos (*triangular tessellations*). Esta distância tem como base a decomposição da *triangular tessellations* em estrelas triangulares. Uma estrela

triangular é uma componente conexa formada pela união de um triângulo com sua vizinhança. A decomposição em estrelas triangulares tem como foco reduzir o número de componentes e cobrir um espaço maior de vizinhos. Em adição, a decomposição proposta permite obter uma representação a qual é invariante ou ao menos inconsciente sob as deformações mais comuns. Antes da decomposição, uma ordem total estrita nos triângulos deve ser estabelecida. Esta ordem visa reduzir o número de estrelas triangulares geradas e garantir a unicidade do resultado da decomposição. Por fim aplica-se a distância, medida de dissimilaridade, entre as estrelas triangulares das duas *triangular tessellations* e utiliza-se esta medida para determinar a melhor correspondência entre os pares de *triangular tessellations*. Para avaliar o método proposto utilizou-se a base de dados TOSCA [524] e comparou-se os resultados obtidos com outros algoritmos de correspondência de formato (CAM[525], GeodesicD2[73], DSR [526] e RSH [527]) em termos de precisão e lembrança.

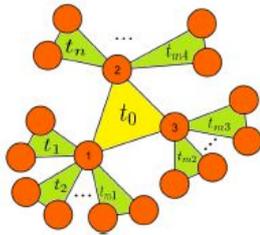


Fig. 173. Representação do método proposto por Madi. Figura extraída de [523].

D. Mapas

Sen Wang [528] propõe um *framework* para o cálculo de correspondência entre superfícies. Para realizar a correspondência de formatos 3D com precisão, tem-se o desenvolvimento de uma representação 2D, Least Squares Conformal Shape Images (LSCSIs) utilizando o Least Squares Conformal Maps (LSCMs). A utilização desta representação simplifica o problema de correspondência de formatos 3D para um problema de correspondência de imagens 2D. Para utilizar o LSCMs, primeiro é necessário estabelecer restrições de características interiores entre os formatos 3D. Para este propósito, primeiro tem-se a seleção pontos candidatos com uma curvatura maior que um limiar, e então tem-se a comparação de suas *spin-images* para detectar correspondências entre características. Depois, geram-se as LSCSIs através do LSCMs, o qual pode mapear cada superfície 3D com uma topologia de disco para o domínio 2D canônico. Os LSCSIs são gerados através da associação de atributos de formato com cada vértice. Além disso, LSCMs podem também ajudar a gerar representações de formato adicional através da associação de outros atributos, como por exemplo textura, o que conduz a uma solução natural de combinação de múltiplas pistas (textura e formato) para a correspondência e reconhecimento de superfícies 3D. Para a parte de correspondência de superfícies utiliza-se um cálculo de correspondências de LSCSIs,

ou seja, gera-se as LSCSIs das duas superfícies e realiza-se um cálculo de similaridade entre elas. O autor utiliza este mesmo *framework* exemplificando como utilizá-lo na parte de correspondências de formatos 3D, no reconhecimento e na parte de costura de superfícies 3D parciais [529]. Para avaliar a performance do *framework* proposto, foram realizados experimentos analisando a robustez da representação proposta, no quesito de correspondência de superfícies 3D, na presença de oclusão, ruído e variação de resolução, e experimentos reconhecendo formatos 3D em dados reais, comparando os resultados de reconhecimento com outra técnicas, em termos de taxa de reconhecimento. Para os experimentos em dados reais, foi utilizado um conjunto de dados de faces 3D, a qual contém 100 escaneamentos, capturados por um sistema de mudança de fase de luz estruturada, de faces de 10 pessoas.

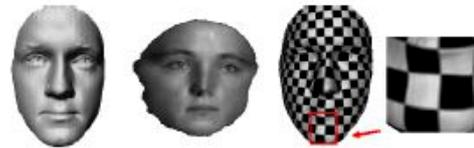


Fig. 174. Representação do método proposto por Sen Wang. Figura extraída de [528].

O trabalho apresentado por Kordelas [208] apresenta uma abordagem que aborda o reconhecimento de objetos em cenas com oclusões e desordem. A abordagem proposta pode ser dividida em duas partes: a primeira parte é a extração e armazenamento dos mapas de distância para cada objeto 3D que pode estar presente na cena. A segunda parte é a extração dos mapas de distância na cena e o reconhecimento dos objetos da cena com base nos mapas de distância da cena e dos objetos com a aplicação de uma medida de similaridade entre os descritores extraídos das imagens. Para a extração e armazenamento dos mapas de distância o primeiro passo é colocar o objeto 3D em uma posição adequada e formar um sistema de coordenadas base utilizado para extrair os mapas de distância para o objeto. Esse processo é realizado com base na aplicação de um principal component analysis e da derivação das três componentes principais ortogonais. O objeto é então rotacionado em torno do centro de massa, de forma que a primeira componente principal fique paralela ao eixo z e a segunda componente principal fique paralela ao eixo y. O objeto é então transladado de forma a deixar o objeto em uma posição que o eixo z passe centralmente através do volume do objeto dado que a coordenada base utilizada para extrair os mapas de distância iniciais é ao redor do eixo z. Após o processo de posicionamento, ocorre o processo de extração dos mapas de distância para o objeto e seu armazenamento em uma biblioteca de modelos. Outros passos realizados nesta primeira etapa foram: a extração e o armazenamento dos mapas de distância final, das imagens correspondentes destes mapas em grayscale e de seus SIFT frames. Para o processo de reconhecimento, inicialmente, são extraídos os mapas de distância da cena e, subsequentemente, as imagens correspondentes em grayscale e os respectivos SIFT frames. Depois, as imagens grayscale do modelo e da cena são correspondidas

com base em dois critérios de correspondência. O primeiro tem como base apenas na comparação entre as imagens em grayscale. O segundo critério combina informações dos mapas de distância, das imagens em grayscale e dos SIFT frames para estipular se a cena e o objeto modelo são correspondentes. Os experimentos realizados para o método proposto incluíram dados reais e sintéticos. Para avaliar a performance do sistema de reconhecimento uma biblioteca com 20 modelos sintéticos foi utilizada. O sucesso no reconhecimento foi verificado em termos das taxas de verdadeiros positivos, verdadeiros negativos e falso positivos. A parte de oclusão e desordem também foi medida com base na segmentação manual dos objetos da biblioteca na cena e comparadas com base nas taxas de reconhecimento.

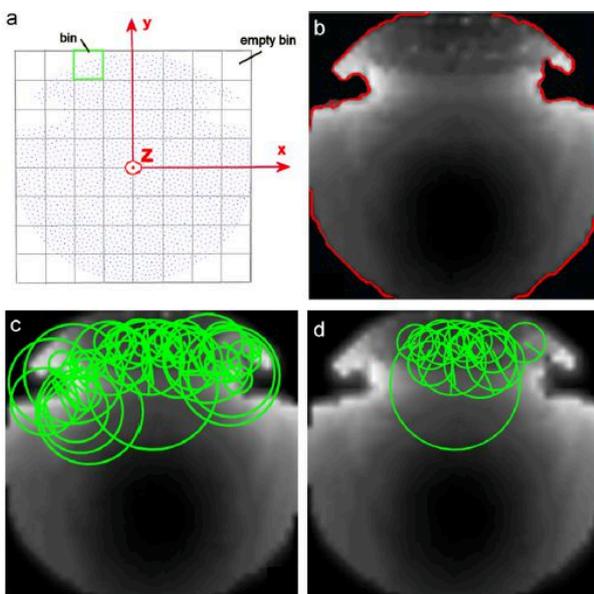


Fig. 175. Representação do método proposto por Kordelas. Figura extraída de [208].

Pintilie [530] apresenta uma nova forma de calcular o conformal map e como utilizar este para a classificação de objetos 3D. Um conformal map é uma representação de baixa dimensão, onde os ângulos formados por três pontos vizinhos na base de dados com alta dimensão são iguais aos ângulos entre esses mesmos três pontos nesta representação. A nova forma de cálculo do conformal map passa por um problema de otimização e esta detalhadamente descrita e avaliada por experimentos apresentados pelo autor. Partindo para a parte de classificação dos objetos, a ideia básica é a de classificar as transformações as quais mapeiam o objeto 3D para um conformal map 2D ao invés de classificar os dados bruto diretamente. De forma mais precisa, propõe-se modelar os objetos 3D pela caracterização das transformações que preservam as invariantes que elas codificam. Deste modo, a transformação que produz o conformal map será utilizada para classificar os objetos. Infelizmente, não existe uma transformação única que possa ser aplicada aos pontos 3D iniciais para produzir a representação bidimensional. O algoritmo de conformal map é um algoritmo de dois passos iniciando com o Locally Linear

Embedding (LLE), o qual produz uma matriz de transformação e terminando com o pós-processamento utilizando o algoritmo de conformal map, o qual produz outra matriz de transformação. Logo, a representação final é composta das duas matrizes de transformação e seus valores são utilizados na caracterização da transformação. Quando comparam-se duas superfícies, os valores das transformações são combinados para formar um vetor e para julgar a similaridade das transformações utiliza-se a distância euclidiana entre todos os pares de pontos. Para os experimentos, foram utilizadas 5 superfícies 3D e os resultados dos testes foram demonstrados em termos da taxa de erro de classificação.

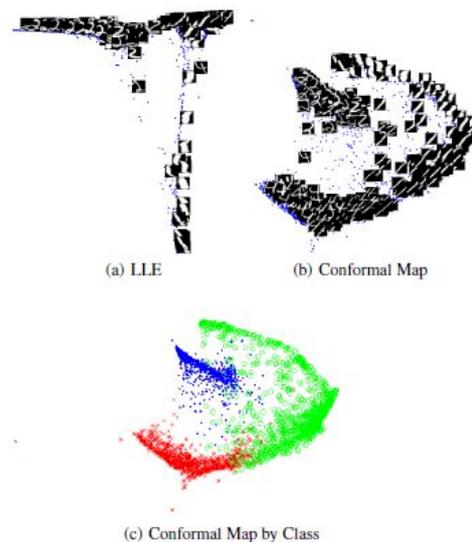


Fig. 176. Representação do método proposto por Pintilie. Figura extraída de [530].

Rodrigues [531] apresenta um modelo cortical para o reconhecimento de faces 3D a partir de suas projeções 2D, explorando sua relação de aspecto sem o uso de disparidade estéreo. Todos os graus de rotação comum, tais como pan e tilt, são considerados. No trabalho em questão, estudou-se o número de *templates* de características 2D requeridos para a representação de todas as visões. No processo de reconhecimento, primeiro detecta-se a visão e então corresponde-se a face de entrada com os *templates* com base em visão armazenados em memória. O modelo cortical utilizado tem como base trabalhos anteriores [532] [533], os quais utilizam bordas e linhas em múltiplas escalas e representações por ponto chave, inspirado na área V1 do córtex visual. Os seguintes passos constituem o modelo cortical como um todo: Segregação da face da cena, detecção de pontos chave e de linhas/bordas em múltiplas escalas, determinação da visão da face de entrada, construção das representações por mapas simbólicos e de saliência e por fim o processo de reconhecimento, onde assume-se que os *templates* das faces (visões) estão armazenadas em memória e que estes foram construídos através de experiência. Cada *template* de face em memória é representado por 400 mapas (5 visões x 4 tipos de eventos x 20 escalas). O esquema de reconhecimento compara as representações das imagens de entrada com os *templates* em memória. Dependendo da

visão da face de entrada selecionada, as duas visões mais similares dos *templates* são selecionadas e pesadas. Se a face de entrada é classificada como frontal, os dois *templates* selecionados são o frontal e lateral-direita. O peso de cada *template* é determinado como uma função da relação de aspecto e os resultados do reconhecimento tem como base esses pesos. Para a parte de reconhecimento de objetos, utiliza-se o mesmo *framework* empregado no reconhecimento de faces, com algumas pequenas modificações. Os testes foram realizados com base em 8 modelos 3D de animais, para a parte de reconhecimento de objetos, e com base em 100 imagens, 10 visões por pessoa, para a parte de reconhecimento de faces. Os resultados dos testes foram demonstrados em termos de taxa de reconhecimento.

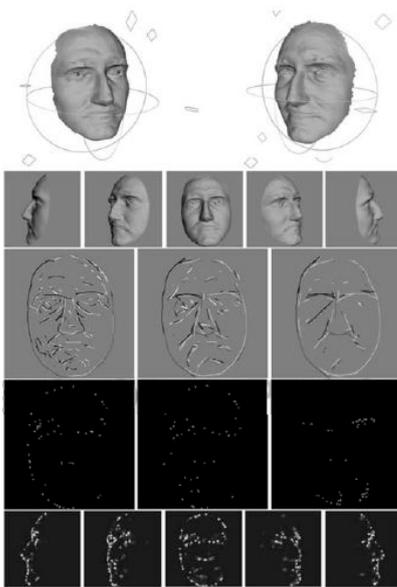


Fig. 177. Representação do método proposto por Rodrigues. Figura extraída de [531].

E. Modelos

1) *Modelo geométrico*: Um método com base em características geométricas quase invariantes em imagens 2D é proposto por Aouat [534] para o reconhecimento de objetos 3D. A abordagem proposta pode ser dividida em duas etapas: a primeira é a construção da base de modelos geométricos e a segunda é o processo de recuperação, o qual é dividido em duas etapas (análise em vários pontos de vista do objeto fazendo o cálculo dos índices e correspondência entre os índices calculados e os presentes na base de modelos geométricos). Para a construção da base de modelos geométricos, são considerados imagens de diferentes pontos de vista para cada objeto e para cada imagem são extraídos pontos e segmentos que fornecem a configuração de segmentos adjacentes necessários para calcular os índices geométricos e adicioná-los a estrutura da base de modelos geométricos. Para o processo de recuperação, ocorre, inicialmente, um processo similar a construção da base de modelos. Primeiramente o objeto em questão é decomposto em várias imagens de pontos de vista diferentes e ocorre o

cálculos dos índices geométricos. Depois, com base nestes índices, ocorre um processo de comparação da similaridade dos índices através de uma busca recursiva na estrutura utilizada na base de modelos geométricos. No final do processo de procura, obtém-se uma lista de possíveis candidatos, na qual realiza-se um processo de votação para apontar o melhor candidato. A votação é realizada com base na análise do descritor Vector Approximation file (VA-file), o qual é um *array* de aproximações geométricas compactas para os dados. Para a avaliação do método proposto foram considerados 28 objetos poliédricos e várias imagens de cada objeto retirados de diferentes pontos de vista e sobre esta estrutura foram realizados testes de tempo de resposta e análises sobre a similaridade de índices, os parâmetros dos VA-file e os resultados obtidos para consultas com objetos presentes na base e objetos desconhecidos.

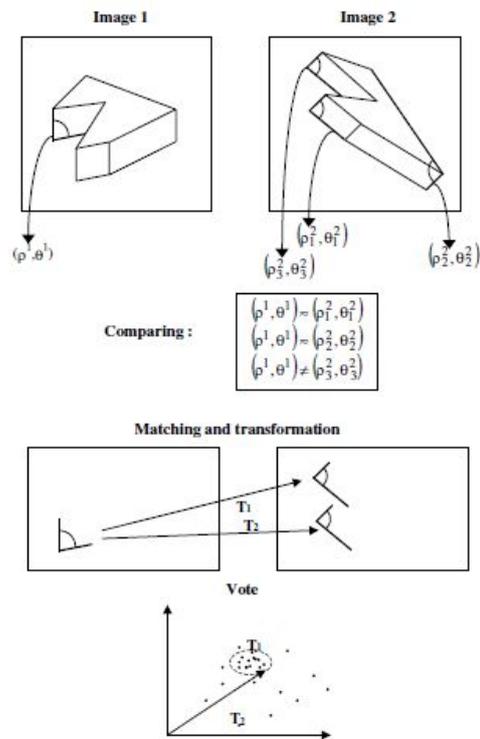


Fig. 178. Representação do método proposto por Aouat. Figura extraída de [534].

Truong [535] apresenta um método de reconhecimento de objetos 3D com base em modelos, utilizando o cruzamento de linhas. Os autores se concentraram em utilizar pares de linhas perpendiculares para testar o reconhecimento de um modelo paralelepípedo e representar a face visível do objeto. Partindo das imagens 2D e da nuvem de pontos, primeiro, os segmentos de linha 3D são extraídos, e então linhas intersectantes são selecionadas. Três passos são utilizados para selecionar um par de segmentos de linhas perpendiculares: Primeiro, eliminam-se casos onde o ponto de interseção é interior ao segmento de linha. Depois, considera-se o ângulo entre duas linhas intersectantes, que satisfazem o requerimento do primeiro passo. Como terceiro passo, mantém-se os

pares de linhas perpendiculares, os quais o tamanho de cada linha é menor que o máximo das dimensões altura, largura e comprimento do modelo e que são 20% maiores que a dimensão mínima do modelo. Após esses três passos obtém-se um conjunto de linhas perpendiculares 3D que possuem alta probabilidade de correspondência com a face retangular do modelo. Através da estimativa da taxa de cobertura, acha-se a correspondência mais precisa entre os pares de linhas perpendiculares detectados e o modelo da base de dados. Por fim, a posição e a pose do objeto são determinados com base nas melhores correspondências encontradas. Os experimentos com o método proposto foram realizados utilizando uma cena com uma caixa e outros objetos, demonstrando o passo a passo, do método proposto, com imagens resultantes de cada etapa.

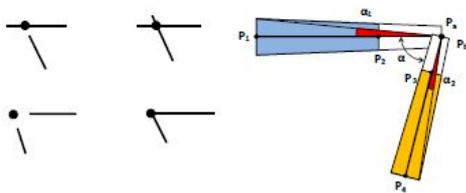


Fig. 179. Representação do método proposto por Truong. Figura extraída de [535].

Khatun [536] [537] propõe uma forma de representação da função de formato 3D utilizando elipsoides e, com base nesta forma de representação, apresenta características *wavelet* com base em elipses as quais formam o vetor de características do formato 3D. O principal objetivo relatado pelo autor é o de representar um formato 3D de forma adequada e computar coeficientes *wavelet* elipsoidais como um vetor de característica definido por uma função do formato 3D na elipsoide, medindo a similaridade do formato. O método proposto consiste de três estágios principais: primeiro, obtém-se uma aproximação elipsoidal parametrizada pelo mapeamento do modelo de formato 3D triangular radialmente do formato para os vértices do centro da unidade de elipsoide. Depois, mapeia-se a função aproximada do formato para uma imagem quadrada, levando em conta que cada ponto na superfície mapeia para um único ponto no quadrado de textura. Por fim, estima-se o vetor de característica do coeficiente *wavelet* aplicando a transformação *wavelet*, mais especificamente Haar *wavelet*, na imagem produzida. Por torna-se custoso trabalhar com todos os coeficientes, utilizaram-se os primeiros 256 dados dos coeficientes para extrair o descritor. Para propósitos de normalização utilizou-se o PCA como pré-processamento. O erro de aproximação foi medido pela distância entre a superfície e a elipsoide. Para avaliar a exatidão do método proposto experimentos foram realizados utilizando a base de dados Princeton Shape Benchmark [39]. Os resultados foram demonstrados com base nos erros de aproximação em relação ao modelo.

Sukhan Lee [538] propõe, tendo como base outro trabalho desenvolvido pelo autor em parceria com Zhaojin Lu [227], uma abordagem probabilística para o reconhecimento de objetos 3D utilizando evidências positivas e negativas em

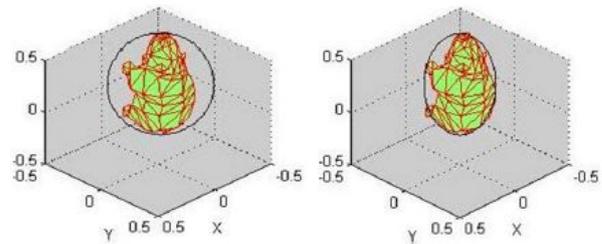


Fig. 180. Representação do método proposto por Khatun. Figura extraída de [536].

ambientes desordenados. O processo como um todo pode ser descrito nos seguintes passos: primeiro, as imagens são adquiridas por uma câmera estéreo, seguido por uma extração de linhas. Então, as linhas 3D extraídas são agrupadas em dois tipos de conjuntos de características com base nas restrições do modelo (pares de linhas paralelas e pares de linhas perpendiculares), as quais usualmente aparecem em objetos feitos pelo homem. Os modelos são armazenados utilizando um modelo poliédrico descrito em termos de arestas, vértices e superfícies. Cada par de características da imagem atribuídas ao conjunto de características do modelo contribui para uma hipótese de pose como sendo uma interpretação consistindo em uma transformação 6-DOF (graus de liberdade). Tipicamente, cada conjunto de características da imagem pode corresponder a múltiplos conjuntos de características do modelo, o que resulta em múltiplas interpretações. Se as interpretações geradas satisfazem o teste de visibilidade, então evidências positivas e negativas serão detectadas e identificadas para a computação de probabilidade. Por fim, um conjunto de interpretações com melhor classificação é produzido após a verificação da pose e dos refinamentos. Devido a ambiguidade, algumas correspondências podem estar incorretas, devido a este fato a abordagem proposta realiza a parte de classificação de interpretações de uma maneira probabilística utilizando regras Bayesianas. Com a finalidade de fazer a probabilidade estimada mais discriminativa, tanto as evidências positivas quanto negativas são utilizadas e o método proposto estima as probabilidades de cada interpretação não apenas através das medidas de probabilidades entre os conjuntos de características do modelo e da imagem, mas também através das medidas de improbabilidade, analisando as informações de contexto em volta da pose estimada. Afim de levar em conta as medidas de incerteza nos valores medidos na imagem durante o processo de geração de interpretações, representa-se cada interpretação como uma região no espaço de pose em vez de um ponto no espaço. Para a validação da abordagem, um número de objetos utilizados diariamente foram capturados com uma câmera estéreo e os resultados obtidos foram avaliados com base nas probabilidades obtidas.

Dahua Lin[539] apresenta uma extensão do *framework* CPMC [540] para 3D com a finalidade de solucionar, de forma conjunta, problemas de compreensão de cena e de reconhecimento de objetos 3D. No intuito de representar os objetos do mundo em termos de cuboides e modelar as iterações físicas e estatísticas entre os objetos e a cena, assim como as iterações

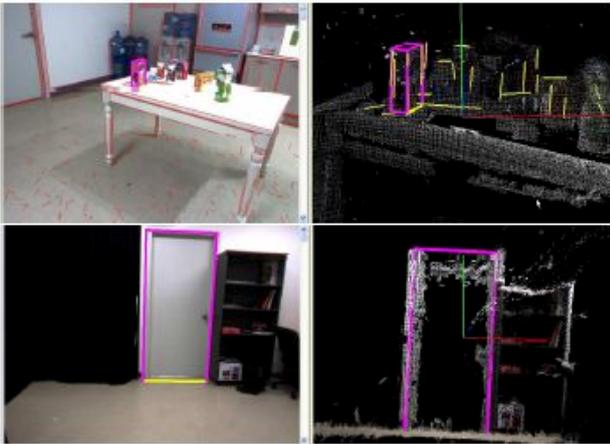


Fig. 181. Representação do método proposto por Sukhan Lee. Figura extraída de [538].

entre os objetos, propõe-se a extensão do CPMC de forma a gerar hipóteses de cubóides em nuvens de pontos através da colocação dos mesmos em volta das regiões 3D candidatas. Estas regiões candidatas são ranqueadas de acordo com um valor de objetividade [540] em aparência e são encorajadas a respeitar os limites de oclusões em 3D. A extensão realizada no CPMC é a inclusão das informações de profundidade as já utilizadas informações de aparência. Por fim, o problema de detecção conjunta é formulado em um Conditional Random Field para modelar a relação contextual entre objetos 3D, com base na integração de informações de aparência, geometria e de contexto. Em particular, tenta-se manter os rótulos dos objetos em concordância com o tipo de cena e fazer os objetos respeitarem as relações geométricas estatísticas em relação ao layout da cena, tais como proximidade com uma parede em um quarto e o tamanho absoluto em 3D, assim como as iterações entre objetos, as quais evidenciam certos tipos de relações de suporte a coocorrência espacial. A abordagem foi avaliada na base de dados NYUv2 RGBD [435] e comparada com diferentes combinações de detectores de cubóides e configurações de CRF e com o DPM [295] em termos de *F1-score*, o qual é a média harmônica entre os valores de lembrança e precisão.

Hejrati [541] apresenta uma abordagem para o reconhecimento e reconstrução de objetos 3D em imagens. Esta abordagem tem como base a estratégia de análise pela síntese. Um modelo de *forward synthesis* constrói possíveis interpretações geométricas do mundo, e então seleciona a interpretação que melhor concorda com a evidência visual medida. O modelo forward sintetiza *templates* visuais definidos em características invariantes, características HOG. Estes *templates* visuais são discriminativamente treinados para serem precisos para estimativa inversa. O modelo apresentado requer dois conjuntos de parâmetros, aqueles associados com a síntese do formato e aqueles associados com os *templates* dos pontos chave locais. Ambos são aprendidos utilizando imagens com localizações dos pontos chave 2D anotadas. Assim como realizado em [542], emprega-se uma *structure from motion* (SFM) [543] não rígida para aprender a base 3D e estimar a base do

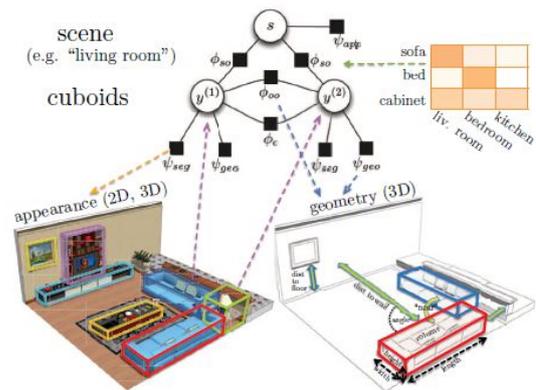


Fig. 182. Representação do método proposto por Dahua Lin. Figura extraída de [539].

formato e o vetor de características 3D para cada exemplo de treinamento. Os clusters de aparência são aprendidos através da clusterização de formatos 3D dependentes da visão do conjunto de características 3D. Os *templates* aprendidos são aqueles que minimizam uma função de treinamento especificada. Para a parte de inferência, apresenta-se uma abordagem eficiente de força bruta que procura através de um grande número de reconstruções candidatas, retornando a reconstrução ótima. Os seguintes passos definem o algoritmo de inferência: Offline, enumera-se o conjunto invariante a translação e armazena-se o conjunto associado de pontos chave renderizado; Online, dado uma imagem, computa-se o mapa de resposta para todas as partes e misturas de *templates* dos pontos chave para cada localização; Avalia-se a pontuação associada com parâmetro em particular do objeto para cada conjunto de parâmetros; Por fim, retorna-se os parâmetros acima de um limiar de detecção, junto com a reconstrução associada. O modelo foi avaliado utilizando dois conjuntos de dados, o SUN primitive dataset [544] e o UCI-Car dataset [542] e os resultados demonstrados em termos de precisão média.

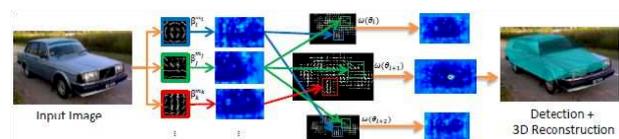


Fig. 183. Representação do método proposto por Hejrati. Figura extraída de [541].

Mabel M. Zhang [545] apresenta um descritor para classificação de objetos 3D por contato. Este descritor, histograma 3d de triângulos, é construído pelo conjunto de amostragens de pontos de contato na superfície do objeto. Cada conjunto de três pontos são conectados para formar um triângulo. Após amostrar vários triângulos, os parâmetros dos mesmos são agrupados em um histograma 3D, um parâmetro por eixo do histograma. O histograma normalizado é o descritor final do objeto. Uma estimativa grosseira da posição do objeto é assumida para assegurar que o contato pode ser feito. Então o histograma é construído puramente através do contato com alguns passos independentes, para minimizar a

propagação de erros de rastreamento no reconhecimento. O resultado é uma representação robusta ao ruído e que tem poucas variáveis dependentes. Dois tipos de resultados simulados foram apresentados em adição ao resultado experimental utilizando um braço robótico. Primeiro, foram amostradas as superfícies de objetos sintéticos para construir o descritor base e realizar o reconhecimento do objeto utilizando uma simulação por nuvem de pontos. Então repetiu-se o processo em uma simulação física e abordou-se o viés de amostragem introduzido por limitações físicas da garra do robô. Para os testes com a nuvem de pontos utilizou-se 185 objetos divididos em 10 categorias, extraindo a nuvem de pontos dos modelos de malha de objetos 3D e calculando o histograma 3D de triângulos. Então essas informações foram utilizadas para treinar um SVM linear para a classificação do conjunto de teste. Os resultados foram demonstrados em termos de acurácia na classificação.

2) *Modelo com base em appearance manifold e matriz de covariância*: O trabalho apresentado por Lina [546] propõe a construção de um *appearance manifold* embutido com uma matriz de covariância dependente da visão para o reconhecimento de objetos 3D em imagens com ruídos. O *appearance manifold* é utilizado para capturar a variabilidade da pose, enquanto a matriz de covariância é utilizada para aprender a distribuição das amostras afim de ganhar invariância ao ruído. O autor apresenta dois modelos de construção do *appearance manifold* denominados View-Dependent Covariance matrix by training-Point Interpolation (VCPI) e View-Dependente Covariance matrix by Eigenvector Interpolation (VCEI). O VCPI é obtido através do interpolação de todos os pontos de treinamento de uma pose aos outros pontos de treinamento nas poses consecutivas. Enquanto o VCEI é obtido através do interpolação apenas dos autovalores e autovetores sem considerar a correspondência de cada imagem de treinamento. Para a parte de reconhecimento com base no VCPI e VCEI utiliza-se a distância de mahalanobis, a qual fornece uma forma suficiente para classificar imagens considerando as características relacionadas e a probabilidade em cada pose da classe. A métrica de mahalanobis é uma forma útil de determinar a similaridade entre uma amostra desconhecida e conjuntos conhecidos. Para avaliar os métodos propostos vários experimentos foram propostos em 2 conjuntos de dados, o primeiro com 7 objetos com cada objeto gerando 360 imagens e o segundo com 10 objetos com cada objeto gerando 360 imagens. O sistema como um todo pode ser descrito nos seguintes passos: captura das imagens por uma câmera, redimensionamento das imagens, extração das características utilizando Principal component analysis e projeção em *eigenspace*, construção dos *appearance manifold* e classificação das imagens de entrada utilizando a distância de mahalanobis. Os dois métodos de construção de *appearance manifold* foram comparados com outros 3 métodos de construção de *appearance manifold* (Simple Manifold with Non-Degraded center, Simple Manifold with Mean center e Constant Covariance matrix), com diferentes níveis de ruído, em termos de acurácia de reconhecimento.

Lina [547] propõe uma técnica para construção do *appearance manifold* em um *eigenspace* para reconhecimento robusto de objetos 3D. Quatro métodos são apresentados para

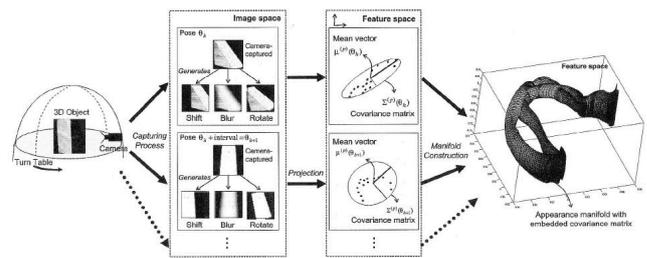


Fig. 184. Representação do método proposto por Lina. Figura extraída de [546].

a construção do *appearance manifold*: método de Parametric Eigenspace (PE), Constant Covariance matrix (CC), método Point Interpolation e o método Covariance Interpolation (CI). Apesar de cada método utilizar um tipo diferente para a construção *appearance manifold*, no geral utilizam os mesmos passos básicos. Dado que neste trabalho foi utilizado apenas o parâmetro horizontal da pose, a superfície do *appearance manifold* terá a aparência de um tubo. Primeiro, após transformar as imagens de treinamento para o *eigenspace*, calcula-se o vetor médio e a matriz de covariância para cada objeto para o ponto de vista horizontal. Então, cria-se um *manifold* contínuo com base nessas duas informações. O processo de criação do *manifold* pode ser diferente de um método para o outro. O método PE utiliza um *manifold* simples obtido da interpolação do vetor médio dos *eigenpoints* em duas poses consecutivas e aplica os valores da matriz identidade para suas matrizes de covariância. O método CC, após o cálculo da matriz de covariância e o vetor médio para cada pose de aprendizagem, aplica-se um método de interpolação para o vetor médio de duas poses de aprendizagem consecutivas para obter o *manifold* do vetor médio. Por outro lado, o *manifold* da matriz de covariância contém os mesmos valores para cada ponto de vista através da aplicação de uma matriz de covariância média. O método PI, obtém o *appearance manifold* pela interpolação de cada *eigenpoint* em cada pose de classe com outros *eigenpoints* nas poses de classes consecutivas que possuem as mesmas características, tal como o mesmo efeito de degradação. Após criar esses *manifolds* para cada *eigenpoint*, geram-se novos *eigenpoints* para todas as poses de classes intermediárias, e então calcula-se seus vetores médios e matriz de covariância para todas as poses da classe. O método CI utiliza o *appearance manifold* embutido com uma matriz de covariância dependente de visão que muda juntamente com a função dos pontos de vista. O vetor médio pode ser obtido pela aplicação de um método de interpolação entre dois vetores médios consecutivos. Então, a matriz de covariância pode ser obtida pela interpolação de duas matrizes de covariância consecutivas. Para reconhecer uma imagem de entrada utiliza-se a medida da distância de Mahalanobis, a qual fornece uma forma suficiente para classificar imagens com base em suas características relacionadas e na probabilidade em cada classe de pose. Para avaliar a performance dos métodos propostos, desenvolveu uma aplicação no reconhecimento de objetos 3D. O sistema desenvolvido foi utilizado para reconhecer 7 objetos com várias posições de poses horizontais e influenciados

por efeitos de degradação geométrica e de qualidade. Na fase de treinamento foi utilizado 6552 imagens para criar os *appearance manifolds* com base nos métodos anteriormente explicados e testou-se o sistema com imagens diferentes das utilizadas para treinamento. Os resultados foram demonstrados em termos de acurácia no reconhecimento.

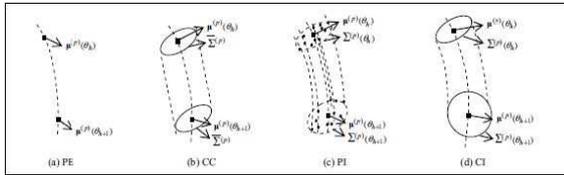


Fig. 185. Representação do método proposto por Lina. Figura extraída de [547].

3) *Modelo de superfície*: Kushal [548] propõe um *framework* para o reconhecimento visual de objetos, onde as classes de objetos são representados por montagens de partial surface models (PSMs) obedecendo restrições geométricas locais. Os PSMs são formados por montagens localmente densas e rígidas das características da imagem. Estes PSMs são aprendidos pela correspondência de padrões de características repetidas através das imagens de treinamento de cada classe. A abordagem de hipótese e validação utilizada para aprendizagem é similar a utiliza em [549], no entanto uma diferença crítica é que as partes relativamente esparsas e rígidas de [549] são substituídas por PSMs localmente rígidos e densos. Pares de PSMs, os quais regularmente ocorrem próximos um do outros em posições relativas consistentes, são ligados por arestas cujos rótulos refletem a relação geométrica local entre estas características. Estas conexões locais são utilizadas para construir um modelo probabilístico de grafo para a geometria e a aparência dos PSMs formando um objeto. Por sua vez, o grafo PSM correspondente é a base para um efetivo algoritmo de detecção e localização de objetos, o qual utiliza a correspondência de todos os PSMs no modelo do objeto para a imagem de teste, de forma independente, para obter um conjunto de correspondências de PSM e sobre este conjunto detectar o objeto correspondente e a sua localização. Os experimentos foram realizados no conjunto de dados PASCAL VOC Challenge 2005 Cars Test 1 [550], comparando os resultados obtidos pelo algoritmo proposto contra outros algoritmos (Dalal & Triggs [195], Voting+MS, Fritz et al. [551], Garcia & Duffner [552]) em termos de precisão média.

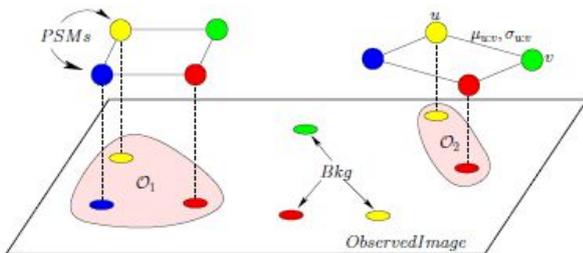


Fig. 186. Representação do método proposto por Kushal. Figura extraída de [548].

Ibrayev [553] apresenta um método para reconhecimento de objetos 3D com superfícies curvas a partir de dados táteis lineares. Para cada modelo de superfície em uma dada base de dados, uma *lookup table* é construída para armazenar curvaturas principais pré-computadas em pontos de discretização na superfície. Para reconhecer um objeto, uma braço robótico com um sensor de toque obtém dados de pontos em sua superfície ao longo de três curvas concorrentes. As duas curvaturas principais estimadas no ponto de intersecção das curvas são utilizadas para procurar na tabela associada com cada modelo para localizar os pontos de discretização da superfície que possuem geometria local similar. Então, buscas locais são realizadas começando com esses pontos para registrar os dados táteis no modelo. O modelo com melhor resultado de registro é reconhecido como resultado de reconhecimento. Três passos principais são executados pelo método para o reconhecimento dos objetos. O primeiro passo é a configuração dos dados, no qual, para facilitar o processo de correspondência, os dados de pontos amostrados estão ao longo de três curvas concorrentes na superfície. A principal razão é para simplificar o controle, o pino move na superfície enquanto restringido sequencialmente em três planos normais distintos através de um ponto q de superfície. As curvas são, portanto, as respectivas intersecções da superfície com este plano. A tangente, a normal e a curvatura de cada curva no ponto q podem ser estimadas via um ajuste de parábola aos dados de pontos correspondentes. A partir destas informações pode-se computar duas curvaturas principais junto com as direções principais [554]. Arbitrariamente escolhem-se dois vetores tangentes ortogonais para formar um *frame* local com a normal da superfície. Todos os dados de pontos ao longo das curvas são então convertidos em coordenadas locais. O próximo passo é a superposição dos dados das curvas. Seja M , o modelo a ser correspondido contra as três curvas, montam-se os dados das curvas em M coincidindo o ponto de intersecção q com algum ponto p na superfície do modelo, e alinhando a normal da superfície dos dados com a normal da superfície no ponto p . O terceiro e ultimo passo é o reconhecimento com base no modelo. Na superfície de cada modelo M , se pré-computa a curvatura principal em uma coleção de pontos, e armazena-se em uma *lookup table* T . Procura-se em T , com o par de curvaturas principais estimadas dos dados de curva, por aqueles pontos com geometrias locais similares na superfície do modelo. O ponto é selecionado como ponto candidato se a distância das curvaturas principais do ponto e do modelo for menor que um valor de tolerância. Com esses pontos candidatos, sobrepõe-se os dados das curvas de cada ponto e computa-se o erro para cada ponto. Então, move-se o ponto q na vizinhança do ponto p em M até que o erro alcance um mínimo local. O erro de registro é o mínimo de erros de sobreposição sobre todos os pontos no modelo os quais q pode ser colocado. O modelo na base de dados que fornecer o menor erro de registro será reconhecido como o formato do objeto. Nos experimentos com o método proposto utilizou-se um sensor de captura dos formatos pelo toque com base em 10 modelos de objetos. Os resultados foram demonstrados em termos de tempo de reconhecimento e erros de registro entre o modelo na base e o objeto medido.

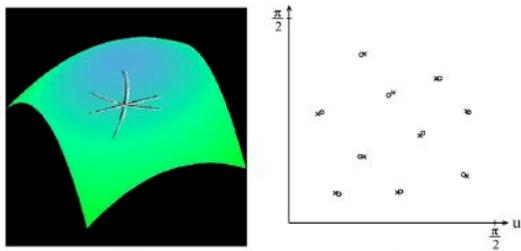


Fig. 187. Representação do método proposto por Ibrayev. Figura extraída de [553].

4) *Modelo com base em distribuições*: Wentao fan [555] apresenta um *framework* estatístico para o reconhecimento e modelagem de objetos 3D. O *framework* proposto tem como base a descrição dos objetos 3D utilizando descritores locais, dos quais um vocabulário visual é construído, e o processo hierarchical Pitman-Yor (HPY)[556] de mistura de distribuições de Beta-Liouville [557]. Uma abordagem online com base em Bayes variacional é desenvolvida para a aprendizagem do *framework* proposto. O modelo gerado é denominado OnHPY-BL e a metodologia para avaliar o modelo no reconhecimento de objetos 3D é descrita pelos seguintes passos: primeiro, extraí-se o 31-dimensional heat kernel signatures (HKS) [558] como descritor local para cada modelo de objeto 3D e então realiza-se a normalização. Então, as características HKS extraídas são modeladas utilizando o modelo proposto OnHPY-BL. Mais especificamente, cada modelo 3D é considerado um grupo e está assim associado a um modelo Pitman-Yor process mixture. Portanto, cada vetor de características HKS de um modelo 3D está supostamente distribuído de acordo com o modelo de Pitman-Yor, onde os componentes de mistura do modelo podem ser considerados palavras visuais. Então, um vocabulário global é construído e compartilhado entre todos os grupos através de um modelo Pitman-Yor process mixture global do modelo hierárquico apresentado. Dado que o objetivo é determinar qual categoria de modelos 3D que um modelo 3D de teste pertence, é requerido a introdução de uma variável indicadora associada com cada modelo 3D, ou grupo, no *framework* proposto. Esta variável indicadora significa que o modelo 3D é de uma certa categoria e é extraído de outro modelo Pitman-Yor process mixture, truncado em um determinado nível. Isto significa que necessita adicionar um novo nível na hierarquia ao modelo hierárquico de mistura infinita com um vocabulário compartilhado entre todas as categorias de modelo. Para os testes utilizou-se um subconjunto do Princeton Shape Benchmark [39], o qual foi dividido em dois grupos, um para treinamento e outro para teste. A performance do modelo proposto foi avaliada comparando os resultados obtidos contra outras três abordagens (HPY-BL, OnHPY-Gau e uma abordagem proposta em [559]) em termos da precisão média de reconhecimento.

5) *Modelos híbridos*: Anand [560] apresenta um sistema com base em visão ativa para o reconhecimento de objetos 3D e estimativa de pose, o qual emprega um time autônomo de robôs, para obter múltiplas visões do objeto, fusão de dados de múltiplos sensores e utiliza um mecanismo de auto-

organização para completar a tarefa. Métodos teóricos de grafos permitem ao sistema localizar múltiplas instâncias de um objeto em um território desconhecido. O objetivo geral é uma equipe de robôs explorar um local desconhecido de dimensões conhecidas e realizar a localização, reconhecimento e estimativa de pose de múltiplas instâncias de objetos neste local. Os robôs utilizados são simples, pequenos e contém dispositivos de locomoção e odometria. As acurácias de odometria necessárias são baixas e são compensadas pela natureza probabilística do modelo do objeto. Cada robô contém os modelos dos objetos, classificadores de partes do objeto e outras estruturas de dados dependentes da implementação. O modelo do objeto é a base de conhecimento fornecida para cada robô e fornece um *framework* probabilístico aproximado mas razoavelmente preciso sobre o objeto. A formulação do modelo tem como base: grafo de função de distribuição de probabilidade (FDP) radial, conjunto de FDP direcionais e árvore de nível de Confiança. O primeiro estágio no processo de detecção é para o time explorar o local e localizar os objetos desejados. Durante esse processo, cada robô continuamente atualiza o time com relação a sua própria posição e da presença de obstáculos no local. Quando uma parte de um objeto classificável é detectada, existe a possibilidade da existência de um objeto naquela localização e o time passa para o estágio de auto-organização. O objetivo do estágio de auto-organização é o de identificar partes suficientes do objeto na localização a fim de atravessar um determinado limiar na confiança de detecção e confirmar a detecção via estimativa de pose. A detecção da primeira parte permite prever círculos esperados, nos quais as outras partes ficarão, utilizando o grafo FDP radial. A detecção de uma segunda parte permite utilizar as tabelas de FDP direcionais e prever uma área de pesquisa mais focada para as partes restantes. Sabendo as posições dos robôs e a posições esperadas das partes no local, pode-se alocar partes para os robôs de tal forma que a distância acumulada percorrida pelo time é minimizada. Os processos colocar na lista negra e de atualização do nível de confiança, de acordo com o modelo do objeto, continuam ocorrendo a cada estágio do processo. Uma vez que o nível de confiança de detecção cruza um limiar definido, o time realiza uma estimativa da pose do objeto. Quando uma instância do objeto é detectada no local, o time volta para o estágio de exploração, marcando a área deste objeto como completada. A simulação do sistema projetado foi realizada utilizando o Gazebo [561], um simulador de robôs 3D de código aberto, associado ao Player [561], servidor de controle de robôs de código aberto.

Um *framework* para reconhecimento de objetos 3D invariante a visões é proposto por Raytchev [562]. O *framework* proposto tem como base o que o autor denomina como Visibility Map (VM), o qual codifica um modelo compacto de um objeto 3D para o qual um conjunto de imagens tiradas de diferentes visões está disponível. Um VM define um subespaço de visibilidade para um objeto 3D através de bases representadas pelos vetores de características. O modelo codifica diferentes visões do objeto através de um vetor binário, onde cada índice do vetor corresponde a um vetor de características representativas. Para comparar novas imagens com este modelo, utiliza-se uma métrica adequada para o

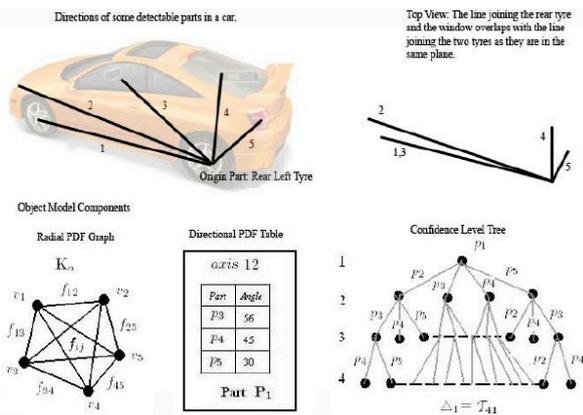


Fig. 188. Representação do método proposto por Anand. Figura extraída de [560].

subespaço de visibilidade, a qual compara duas visões codificadas levando em conta o número de componentes conexas e os pesos de co-ocorrência, os quais implementam restrições geométricas codificadas pela matriz de co-ocorrência. Para testar a performance do VM utilizou-se a base de dados COIL-100 [23], separando em imagens para treinamento e teste. Os resultados para os testes executados foram demonstrados em termos número de características, tempo de cálculo e taxa de reconhecimento e comparados com uma implementação convencional da estratégia de correspondência (*view-to-view* (V2V)).

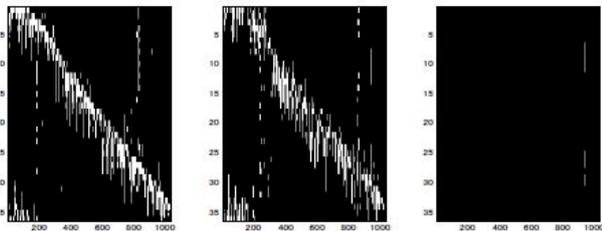


Fig. 189. Representação do método proposto por Raytchev. Figura extraída de [562].

Kent [563] apresenta um sistema para construção de modelos de objetos para o reconhecimento de objetos 3D e manipulação possibilitada por avanços na web robótica. Duas contribuições são realizadas por esse trabalho. A primeira é o sistema com base em web para coletar dados de nuvem de pontos para um conjunto de objetos desejado. A segunda e foco do trabalho é a construção de um sistema de reconhecimento de objetos utilizando apenas nuvens de pontos. O sistema constrói modelos 3D dos objetos, dado um conjunto de representações por nuvens de pontos de um objeto capturado de diferentes pontos de vista, através da determinação dos conjuntos de disparidade sobre nuvens de pontos sobrepostas e da fusão de cada conjunto de nuvens de pontos com características em comum. Para a construção do modelo partindo de visões individuais, primeiro utiliza-se Point Cloud Library (PCL) [404] para implementar um *pipeline* de registro de nuvem de pontos. Este processo permite a fusão de qualquer duas

nuvens de pontos em uma única nuvem de acordo com características SIFT correlacionadas. O *pipeline* é composto pelos seguintes passos: Para cada par de nuvem de pontos, primeiro calcula-se um conjunto de características (diferença de tamanho, diferença do desvio de cor, diferença de cor média, erro de distância, erro de cor e sobreposição) que caracterizam a fusão em potencial. Utilizando estas métricas, e um conjunto de treinamento de rotulamentos manuais de sucesso e falha do registros de nuvens de pontos, treina-se um árvore de decisão, por meio do algoritmo C4.5 [564], para prever fusões bem sucedidas. Por fim, utiliza-se a árvore de decisão para construir uma representação por grafo de fusões de pares de candidatos que são, então, realizadas iterativamente para gerar o modelo de objeto. O resultado é uma representação do objeto por um conjunto de múltiplos modelos do objeto. Para testar o modelo gerado, implementou-se um processo similar ao de geração do modelo. Especificamente, utilizando o mesmo *pipeline* de registro de nuvem de pontos utilizado na construção do modelo, o sistema de reconhecimento tenta registrar uma nuvem de pontos de um objeto desconhecido para cada modelo de objeto na base de dados. Pelo cálculo das métricas listadas anteriormente, o algoritmo atribui uma medida de erro para cada tentativa de registro aos pares. O algoritmo então reconhece o objeto desconhecido como o objeto da base de dados com o menor resultado de erro. Para demonstrar a generalização do sistema de construção e reconhecimento, avaliou-se os dados utilizando uma validação por repetida subamostragem randômica. Para cada iteração de teste, as nuvens de pontos para cada objeto foram separadas aleatoriamente em conjunto de treinamento (construção do modelo) e conjunto de teste (utilizado para reconhecimento do objeto). Os resultados dos testes foram demonstrados em termos de taxa de classificação geral dos objetos.

6) *Modelos de máquina Boltzman*: Um novo tipo de modelo de alto nível para Deep Belief Nets (DBNs), avaliado na tarefa de reconhecimento de objetos 3D, é introduzido por Nair [565]. Este modelo de nível superior é uma máquina Boltzman de terceira ordem, treinada utilizando um algoritmo híbrido que combina gradientes generativos e discriminativos. As duas principais decisões que devem ser feitas quando realiza-se o treinamento das DBNs são: o número de camadas escondidas e o número de unidades em cada camada. Realizado o treinamento com os casos de treinamento rotulados e selecionado o número de camadas e número de unidades por camada, o próximo passo é o teste do modelo treinado. A performance do modelo foi avaliada na base de dados NORB [478], a qual contém pares de imagens estéreo de objetos em diferentes condições de iluminação e de visão. Os resultados obtidos foram apresentados em 3 partes comparando-os em termos das taxas de erro. A primeira parte compara modelos *deep* com modelos *shallow*, todos treinados utilizando Contrastive Divergence (CD). A parte 2 compara o CD com o algoritmo híbrido de aprendizagem para o treinamento do modelo de nível superior de uma DBN. A parte 3 compara DBNs treinadas com e sem dados rotulados, utilizando ou o CD ou o algoritmo híbrido no nível superior.

7) *Modelos estatísticos*: Grzegorzek [566] apresenta uma abordagem probabilística para classificação e localização de

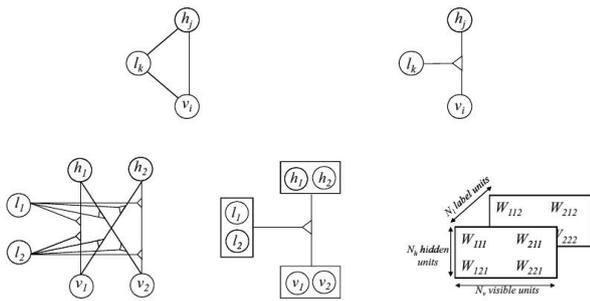


Fig. 190. Representação do método proposto por Nair. Figura extraída de [565].

objetos 3D em imagens 2D com múltiplos objetos. Esta abordagem pode ser dividida em duas fases, treinamento e reconhecimento. Na fase de treinamento, modelos estatísticos dos objetos e do contexto são aprendidos separadamente. Para a modelagem do objeto, o sistema de reconhecimento extrai vetores de características locais para o treinamento, utilizando a transformação *wavelet* e modela estatisticamente estes vetores através de funções de densidade. Uma vez que em ambientes contextuais a probabilidade a-priori para a ocorrência de diferentes objetos não pode ser assumida como a mesma, a modelagem estatística do contexto é introduzida. As probabilidades de ocorrência a-priori são aprendidas na fase de treinamento, para cada contexto separadamente utilizando um conjunto de dados de treinamento grande, e armazenadas em modelos de contexto. Uma vez que a modelagem do objeto e a modelagem do contexto terminaram, o sistema é capaz de classificar e localizar objetos nas imagens retiradas do ambiente contextual. Primeiro, uma imagem de teste é adquirida, pre-processada e os vetores de características são computados da mesma forma que foi realizada na fase de treinamento. Então, um dos algoritmos de reconhecimento integrados no sistema, algoritmo de reconhecimento de um único objeto na cena, algoritmo de reconhecimento para múltiplos objetos na cena sem a modelagem de contexto ou o algoritmo de reconhecimento para múltiplos objetos na cena com a modelagem de contexto, é iniciado. Os experimentos foram realizados em imagens da base de dados 3D-REAL-ENV [567] e os resultados foram demonstrados em termos de taxas de determinação do número de objetos, localização e classificação, com e sem a modelagem contextual para fins de comparação.

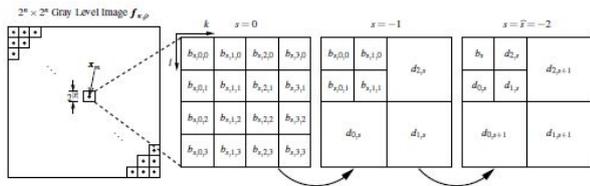


Fig. 191. Representação do método proposto por Grzegorzec. Figura extraída de [566].

8) *Modelos CAD*: Muhammad zia [568] apresenta uma abordagem para o problema de interpretação da cena. Nesta

abordagem, ao invés de construir o modelo da cena com um detector de bounding box 2D, utiliza-se um modelo de classes de objetos 3D, estendendo-o para juntamente representar cenas com múltiplos objetos. A interpretação da cena resultante engloba o detalhado formato 3D de todos os objetos na cena, estabelecendo uma conexão explícita entre a evidência 2D na imagem e a geometria 3D através de um modelo de *wireframe*. A abordagem proposta pode ser descrita nos seguintes passos: primeiro, descreve-se o modelo de cena 3D consistindo em uma área plana comum, um conjunto de objetos deformáveis 3D e uma explícita máscara de oclusão para cada objeto. Então, expressa-se a probabilidade de uma hipótese de cena particular naquele espaço como sendo a combinação das probabilidades por objetos, computados através de modelos existentes, em CAD, para cada objeto. Por fim, utiliza-se um procedimento de inferência com base em amostras, o qual levanta uma hipótese 2D inicial sobre a localização do objeto para um modelo 3D grosseiro e refina este modelo grosseiro em uma coleção final de formatos 3D consistentes. Para a avaliação dos resultados, foi analisada a habilidade da abordagem na recuperação da localização e da pose do objeto 3D em cenas de ruas de uma base de dados pública [569]. Os resultados foram demonstrados com base na acurácia para estimar a localização e o ponto de visão.

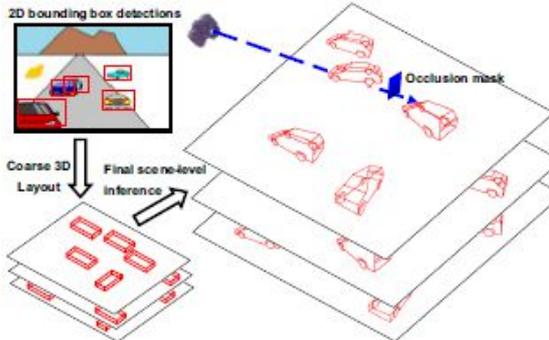


Fig. 192. Representação do método proposto por Muhammad Zia. Figura extraída de [568].

Um método de reconhecimento de objetos 3D em imagens monoculares é proposto por Pengfei Han [570]. O método de reconhecimento com base em modelos CAD busca enfrentar o desafio de reconhecimento de objetos 3D de uma imagem monocular em realidade aumentada móvel em uma pequena quantidade de tempo e minimizando a interação com o usuário. Em vez de estimar a posição da câmera diretamente, o método proposto tenta estimar a pose do objeto 3D no sistema de coordenadas da câmera. A correspondência geométrica entre o centro de massa do objeto e a sua projeção 2D na câmera é utilizada para estimar a posição do objeto. Então fora da tela, o modelo CAD é renderizado em diferentes azimutes sobre um restrição de inclinação assumida fornecida pelos sensores inerciais dos dispositivos móveis para hipotetizar possíveis azimutes do objeto e comparar com a imagem de entrada através de uma correspondência de contorno. O necessário na parte offline do método proposto é apenas o modelo CAD do objeto 3D. O método proposto foi testado em experimentos

utilizando imagens reais e dados dos sensores avaliando os resultados com base em acurácia e consumo de tempo do método.

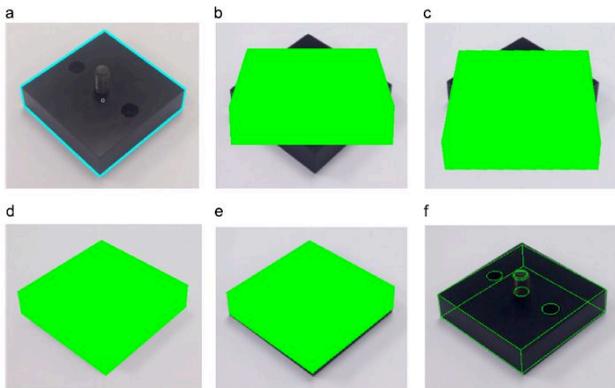


Fig. 193. Representação do método proposto por Pengfei Han. Figura extraída de [570].

9) *Modelo de segmentação com base em octree*: Com o objetivo de reconhecer e segmentar os objetos 3D em uma imagem RGB-D capturada pelo Kinect com a informação de imagens rotuladas em um base de dados, Yong-Hui Xu [571] propõe um método supervisionado com base na tecnologia de transferência de rótulos. A relação entre o objeto observado e os objetos rotulados é a parte central do método proposto, o qual pode ser dividido em três partes: recuperação de imagem, transferência de rótulos e segmentação do objeto 3D. Na parte de recuperação das imagens, utiliza-se características invariantes a escala, características SIFT extraídas do espaço de cor, para procurar um conjunto de vizinhos mais próximos da imagem de entrada na base de dados de imagens rotuladas. A parte de transferência de rótulos é realizada com base nos pares de características SIFT e na utilização dos métodos mínimos quadrados e RANSAC para calcular a matriz de projeção ótima entre a imagem de entrada e as imagens rotuladas. Com a utilização destes métodos, transfere-se os rótulos das imagens RGB rotuladas para a imagem RGB-D de entrada. Como ultima parte tem-se a segmentação dos objetos 3D. Com o resultado da transferência dos rótulos, cria-se um modelo de segmentação e se estabelece um classificador com base neste modelo. O modelo de segmentação elimina o ruído herdado pela transferência de rótulo e recupera as partes perdidas do objeto observado. Então, clusteriza-se o modelo de sub-cubo criado e formam-se as regiões do objeto espacialmente e semanticamente consistentes. Os experimentos com o método proposto foram conduzidos com base em 3 objetos selecionados para identificação e segmentação.

F. Subespaços

Um Kernel Orthogonal Mutual Subspace Method (KOMSM) para o reconhecimento de objetos 3D é apresentado por Fukui [572]. O KOMSM é um método com base na aparência para classificar um conjunto de padrões tais como quadros de vídeos ou imagens de múltiplas visões. O KOMSM, assim como Kernel Constrained Mutual Subspace Method (KCMSM) apresentado anteriormente

pelo autor [573], é derivado diretamente do Kernel Mutual Subspace Method (KMSM) que por sua vez é a extensão não-linear do Mutual Subspace Method (MSM). O MSM classifica um conjunto de padrões com base no ângulo canônico entre classes lineares de subespaços, o que representa a distribuição do conjunto de treinamento de cada classe respectivamente. Devido a queda de performance do MSM quando a distribuição dos padrões possui estruturas não-lineares, foi proposto o KMSM. Neste método um padrão de entrada é mapeado em um espaço de características com grande dimensão através de um mapa não-linear. Consequentemente, o KMSM traz o MSM no subespaço linear gerado do mapeamento dos parâmetros utilizando a expansão de Karhunen-Loève (KL), também conhecida como Principal Component Analysis (PCA). O KMSM funciona bem, mas possui problemas de performance em algumas aplicações devido ao fato que as classes de subespaços não-lineares são gerados independentemente uns dos outros. Desta forma apresenta-se o KOMSM, o qual alcança uma perfeita ortogonalização das classes de subespaços não-lineares e sob estes espaços ortogonais aplica o KMSM. Logo o KOMSM pode ser descrito como a aplicação do linear MSM aos subespaços lineares gerados pelo *whitening transformed vectors* no espaço de característica. No estágio de treinamento aplica-se a transformação pelo *kernel whitening matrix* para todos os padrões pertencentes a uma classe k e obtém-se os vetores base do subespaço. Na parte de reconhecimento, o subespaço linear ortogonal de entrada é gerado do conjunto de padrões de entrada, os ângulos canônicos entre o subespaço de entrada e todos os subespaços de referência são calculados como a similaridade e finalmente a classe do objeto é determinado com base no maior valor de similaridade em relação aos subespaços de referência, dado que o valor de similaridade é maior que um valor de limiar. A performance do KOMSM foi comparada com outros métodos (MSM, CMSM, OMSM, KMSM, KCMSM) utilizando a base de dados pública ETH-80 [105] e um conjunto de dados de faces coletados pelos autores em termos de taxa de reconhecimento e *Equal Error Rate*.

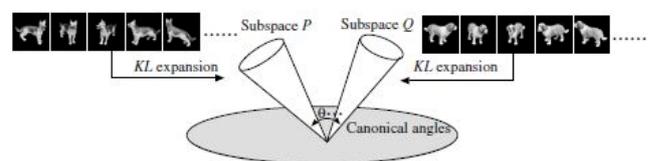


Fig. 194. Representação do método proposto por Fukui. Figura extraída de [572].

Um método com base em visões para classificação de objetos 3D é proposto por Jianing Wu [574]. A ideia chave do método é a de aproximar a distribuição dos vetores de característica com múltiplos subespaços locais. Esta operação enfraquece a não linearidade em cada subespaço local. Entretanto, é difícil otimizar o número de subespaços locais e a dimensão de cada subespaço local, logo, geram-se múltiplos conjuntos de subespaços locais através da mudança desses parâmetros. Então aplica-se a aprendizagem de conjunto com

os subespaços locais gerados. Para realizar a ideia central apresentada, aplica-se o método k-means sob a distribuição dos vetores de característica para obter k subconjuntos. Depois, aplica-se o PCA sob cada subconjunto para obter os vetores base do subespaço. Os vetores base de cada subespaço são calculados como os autovetores correspondendo aos maiores autovalores da matriz de autocorrelação. Para calcular a similaridade entre subespaços utiliza-se o ângulo canônico entre os subespaços como definido em [575]. O fluxo do processo de classificação consiste de duas fases, aprendizagem e classificação. Na fase de aprendizagem, utiliza-se um conjunto de subespaços locais para aproximar a distribuição de cada classe. Múltiplos conjuntos destes subespaços locais são gerados com a mudança dos números de *clusters* e da dimensão dos subespaços locais. O peso de cada subespaço local é ajustado após a geração dos subespaços locais. Na parte de classificação, a similaridade de um subespaço de entrada com cada subespaço local é calculada utilizando duas formulações definidas pelos autores para primeiramente, calcular a similaridade entre um subespaço local e o de entrada e depois combinar essas similaridades em um único valor. A classe com o maior valor de similaridade é o resultado de classificação. Para os experimentos com o método proposto utilizou-se a base de dados The ETH-80 Image Set [105]. Os resultados dos experimentos foram comparados com outros dois métodos (Mutual Subspace Method (MSM) e Kernel Mutual Subspace Method (KMSM)) em termos de taxa de acurácia, EER (Equal Error Rate) e separabilidade.

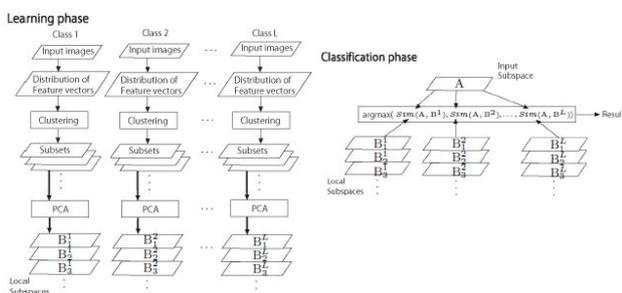


Fig. 195. Representação do método proposto por Jianing Wu. Figura extraída de [574].

Um método para classificação de objetos, independente de pose e com um pequeno número de amostras para treinamento é apresentado por Hotta [576]. O método proposto tem como base os kernels das partes locais. Para utilizar efetivamente os kernels das partes locais, características locais de aparência são utilizadas. Para tanto, as características de Gabor foram utilizadas. O método proposto pode ser descrito nos seguintes passos: o primeiro passo envolve a extração das características de Gabor, através dos filtros de Gabor, seguido pelo extração das partes locais das características de Gabor. Depois, ocorre a aplicação do Kernel principal component analysis às partes locais extraídas. Os subespaço obtido é utilizado na parte de classificação. Cada classe possui um subespaço. O próximo passo é a classificação, onde, para computar a similaridade entre o subespaço construído e as partes locais extraídas da amostra de teste, foi utilizado o

Class-featuring information compression (CLAFIC), o qual fornece um valor de similaridade através da soma ponderada das similaridades com as partes locais de treinamento. Logo, o processo de classificação se resume a extrair as características de Gabor da amostra de teste e, subsequentemente, extrair as partes locais das características de Gabor. Depois, mede-se a similaridade entre o subespaço da classe k com todas as partes locais extraídas da amostra. Como várias partes locais são extraídas da amostra, gerando vários valores de similaridade, são utilizadas regras de votação (voto para a classe com o maior valor de similaridade e seleção da classe com maior número de votos), somatório (soma dos valores de similaridade de todas as partes locais com cada classe e seleção da classe com maior valor somado) e mediana (cálculo da similaridade das partes locais com todas as classes e posterior cálculo da mediana para cada classe, selecionando a classe com maior valor de mediana) para combinar os resultados de similaridade e selecionar a classe que a amostra de teste deve ser classificada. A avaliação do método proposto foi realizada com base nos resultados da aplicação do método ao problema de reconhecimento de faces. Foram utilizadas imagens de 300 pessoas em diferentes ângulos utilizadas para construção do subespaço e testes. Os resultados obtidos foram comparados, em termos de taxa de reconhecimento, com outros métodos em diferentes configurações de representações de objetos (SVM com características globais, Linear PCA, Nearest Neighbour no espaço KPCA, entre outras)

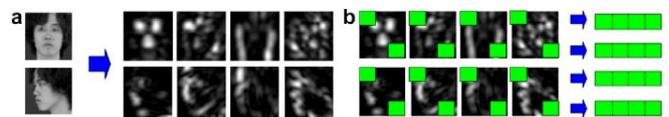


Fig. 196. Representação do método proposto por Hotta. Figura extraída de [576].

Akihiro [577] propõe o Compound Mutual Subspace Method (CPMSM) como uma extensão teórica do Mutual Subspace Method (MSM), o qual pode lidar com múltiplos conjuntos de padrões através da representação dos mesmos como subespaços. O autor mostra a aplicação do método para faces, entretanto ele pode ser usado para diferenciar qualquer tipo de classe de objetos 3D. O CPMSM pode ser dividido em duas etapas aprendizagem e teste. Na parte de aprendizagem aplica-se a expansão Karhunen-Loève (KL) nas classes s e t do conjunto de imagens de treinamento para obter os subespaços de referência P_t e P_s . Depois, obtém-se o subespaço da diferença, D_{st} , através do conjunto de vetores de diferença entre os vetores canônicos. Para o teste, aplica-se novamente a expansão KL no conjunto de imagens de entrada para obter o subespaço de entrada I . Então, calcula-se a similaridade entre o subespaço de entrada e os subespaços de referência ($S(I, P_t)$ e $S(I, P_s)$) e, igualmente, calcula-se os subespaços das diferenças (D_{Is} e D_{It}). Por fim, calcula-se as similaridades entre os subespaços das diferenças $S(D_{Is}, D_{st})$ e $S(D_{It}, D_{st})$ e combina-se a similaridades ($S(I, P_t)$ e $S(D_{It}, D_{st})$) e ($S(I, P_s)$ e $S(D_{Is}, D_{st})$) para obter as similaridades $SCPMSM(I, P_t)$ e $SCPMSM(I, P_s)$ respectivamente. Para a identificação compara-se a similaridade

dades $SCPMSM(I, P_t)$ e $SCPMSM(I, P_s)$ para classificar o conjunto de entrada na classe com maior valor de similaridade. Para os experimentos utilizando o CPMSM, utilizouse a base de dados de faces VidTIMIT [578], e os resultados foram comparados com os resultados obtidos da aplicação dos métodos MSM e o Constrained MSM (CMSM) em termos de taxa de reconhecimento e taxa de erro.

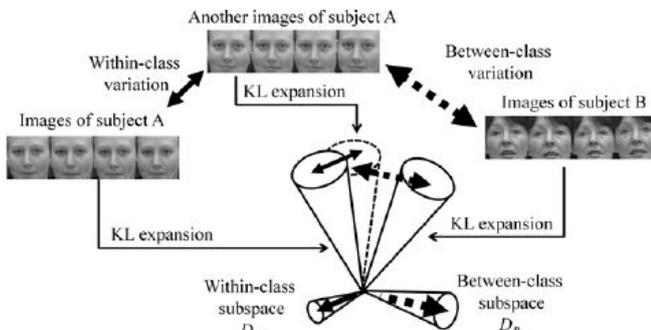


Fig. 197. Representação do método proposto por Akihiro. Figura extraída de [577].

Um método para medir a similaridade de formatos para objetos 3D utilizando subespaços de formato 3D produzidos por um método de fatoração é apresentado por Igarashi [579]. Inicialmente, os pontos de características são rastreados através de uma sequência de imagens, para cada objeto, por um rastreador de características tal como o Kanade-Lucas-Tomasi (KLT). Então, os subespaços de formato dos dois objetos são derivados dos conjuntos de pontos de característica rastreados por um método de fatoração [580], o qual pode, de forma robusta, recuperar o formato e o movimento de um objeto de uma sequência de imagens sem assumir um modelo de movimentação, tal como translação constante ou rotação. Por fim, os ângulos canônicos entre os subespaços de formato são encontrados e utilizados para construir a medida de similaridade de formatos. Para obter uma medida de similaridade robusta entre os subespaços de formato utiliza-se o conceito de matriz de projeção ortogonal, a qual é unicamente determinada a partir dos vetores de base ortogonal do subespaço de formato. O centro da ideia é minimizar a diferença entre duas matrizes de projeção ortogonal, as quais são geradas dos pontos de características dos dois objetos, através do rearranjo de linhas e colunas de uma das matrizes. Os pontos de característica são considerados como tendo sido correspondidos entre os dois objetos quando a diferença entre as duas matrizes é a menor. Com base nos ângulos canônicos define-se uma medida de similaridade, a qual se dois subespaços coincidirem completamente um com o outro o valor de similaridade é 1, dado que todos os ângulos canônicos são zero. A similaridade fica menor a medida que os dois subespaços se separam. Por fim, a similaridade é zero quando os dois subespaços são ortogonais entre eles. Para os experimentos com o método proposto, inicialmente, foram utilizados dados sintéticos para avaliar a acurácia do algoritmo proposto para correspondência de pontos de característica, e então foram utilizadas imagens reais de faces para demonstrar a eficácia do método proposto para o reconhecimento de

objetos, comparando com outro método para reconhecimento de faces em termos das taxas de reconhecimento e *Equal Error Rate* (EER).

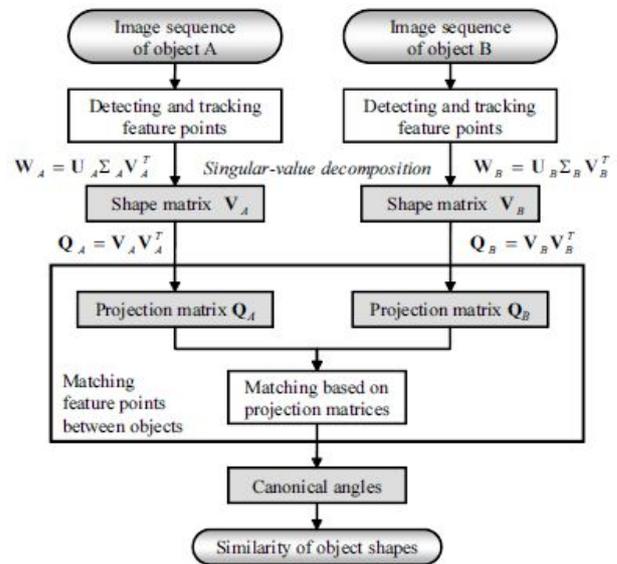


Fig. 198. Representação do método proposto por Igarashi. Figura extraída de [579].

Kise [581] apresenta três métodos de reconhecimento de objetos 3D, nos quais a principal característica é o uso de um grande número de subespaços gerados de um número grande de características locais. As características locais são extraídas via uma cadeia de características SIFT presentes em pares de *frames*. Para a construção dos subespaços com base nas características utilizou-se o PCA. Com o intuito de corresponder as características locais com uma grande quantidade de subespaços, uma simples aproximação por procura pelo vizinho mais próximo é utilizado. Com base nesta aproximação, os três métodos propostos são os seguintes: um método com um subespaço comum, o qual corresponde cada característica local inquerida para um subespaço através da projeção da característica no subespaço, para cálculo de similaridade; um método de correspondência em duas etapas, o qual no primeiro passo os subespaços são utilizados apenas para selecionar os subespaços candidatos e no segundo passo os subespaços candidatos são examinados em uma larga dimensão para selecionar o melhor. Isto é, um vetor de consultas é projetado em todos os subespaços candidatos e o subespaço com maior similaridade é empregado para votação; um método com subespaços mútuos (tanto consulta quanto modelos são representados como subespaços e a similaridade é definida com base nos ângulos canônicos). O resultado do reconhecimento do objeto é obtido pela votação dos objetos com base na correspondência. Os métodos propostos foram avaliados em uma base de dados gerada, a qual consiste em 1002 objetos e cada objeto teve suas imagens adquiridas de diferentes pontos, e os resultados foram analisados em termos de taxa de reconhecimento e tempo de processamento.

Kobayashi [582] propõe o generalized Mutual Subspace Method (gMSM) para minimizar o problema relacionado

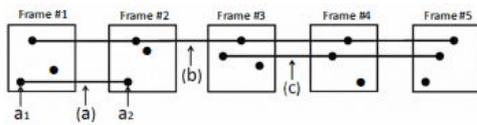


Fig. 199. Representação do método proposto por Kise. Figura extraída de [581].

com a dimensionalidade do subespaço. O método proposto introduz uma ponderação suave nos vetores base compondo o subespaço, sem definitivamente selecionar um pequeno número de bases principais, ou seja, adiciona-se um pesos nos cálculos com a finalidade de relaxar o peso binário (1 ou 0) no momento de seleção dos vetores base do espaço. Este método é uma generalização do Mutual Subspace Method (MSM), o qual classifica um conjunto de vetores (subespaços) através do cálculo de similaridade entre eles com base em seus ângulos canônicos. Logo, o gMSM é uma generalização do MSM por meio do uso de uma ponderação suave para efetivamente combinar a bases para a computação dos ângulos dos subespaços. Adicionalmente, também reformulou-se o Constrained Mutual Subspace Method (CMSM), uma variação do MSM que busca construir o espaço discriminativo para fazer os ângulos medidos entre os subespaços mais discriminativos. Esta reformulação do CMSM, construindo o espaço discriminativo de uma forma diferente da originalmente mostrada em [583], permite a generalização do método através de uma ponderação suave apresentado o generalized Constrained Mutual Subspace Method (gCMSM). Para a classificação do conjunto de vetores de características, inicialmente treina-se o espaço discriminativo, utilizando o gCMSM, no qual o gMSM é subsequentemente executado para computar os ângulos do subespaço. Depois, classifica-se os conjuntos aplicando o gMSM no espaço discriminativo montado no treinamento e, subsequentemente, aplica-se o k-NN, com k igual a 1. Para os experimentos dos métodos propostos na parte de classificação de objetos 3D utilizou-se duas bases de dados (ETH-80 [105] e RGB-D [257]) e os resultados foram comparados com outros métodos (MSMS, OSM, DCC, CMSM-mono, CMSM) em termos de taxa de acurácia.

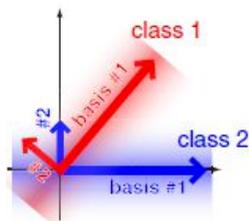


Fig. 200. Representação do método proposto por Kobayashi. Figura extraída de [582].

Fukui [584] apresenta um *framework* para extrair as diferenças dos formatos locais entre dois objetos distintos e então realizar a classificação utilizando estas diferenças. Para tanto, primeiro se define geometricamente o conceito de difference subspace (DS), o qual representa as componentes da diferença entre dois subespaços, estas componentes

tem como base os ângulos canônicos, e descreve-se sua eficácia em analisar diferenças de formato. Então, generaliza-se o DS formando o generalized difference subspace (GDS) para subespaços de múltiplas classes. Para esta generalização primeiro mostra-se que o DS entre dois subespaços pode ser definido analiticamente pelo uso das duas matrizes de projeção dos subespaços. Desta definição analítica, pode-se então gerar sistematicamente o GDS e mostra-se o benefício de aplicar isto aos métodos de Subspace Method (SM) [585] e Mutual Subspace Method (MSM) [586] em termos de capacidade de reconhecimento. Denominam-se as versões aprimoradas do SM e do MSM como constrained SM e MSM (CSM e CMSM), respectivamente. O termo *constrained* reflete o fato que utiliza-se o SM e o MSM em um subespaço *constrained* que satisfaz algumas *constraints*. Além disso, estende-se estes métodos para o kernel DS (KDS) e kernel GDS (KGDS), através de um mapeamento kernel não linear, kernel principal component analysis (KPCA) [587], para lidar com casos envolvendo grandes mudanças na direção da visão. Para avaliação do *framework* proposto utilizou-se múltiplas imagens de formatos de mãos para classificação. Os experimentos compararam os resultados do *framework* proposto contra outros métodos (SM [585], DCC [588], SRC [589], GDA [590], MSM [586], FDA+MSM, Orth+MSM [591], KMSM [592], KOrth+KMSM [572]) em termos de *error rate* e *equal error rate*

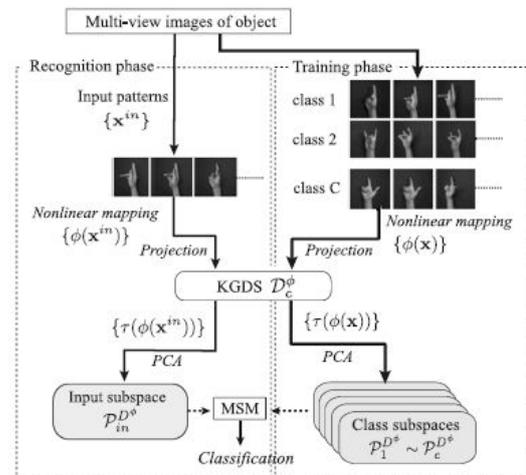


Fig. 201. Representação do método proposto por Fukui. Figura extraída de [584].

G. Tensores

Mian [81] apresenta um algoritmo com base em modelo, o qual realiza o reconhecimento de objetos de formato livre e suas segmentações, independente do ponto de vista, na presença de oclusão e desordem. O modelo 3D do objeto é automaticamente construído, offline, a partir de de suas múltiplas imagens de profundidade (visões) desordenadas. Essas visões são convertidas em representações por tabelas multidimensionais, referenciadas como tensores. Correspondências são automaticamente estabelecidas entre estas visões através da correspondência simultânea dos tensores de uma visão com aqueles das visões restantes utilizando um esquema de

votação com base em tabelas hash. Isto resulta em um grafo de transformações relativas usado para registrar as visões antes de elas serem integradas em um modelo 3D. Estes modelos e suas representações por tensores constituem a biblioteca de modelos. Durante a fase online de reconhecimento, um tensor de uma cena é simultaneamente correspondido com aqueles na biblioteca através de atribuição de votos. Medidas de similaridade são calculadas para os tensores do modelo que recebeu o maior número de votos. O modelo com maior similaridade é transformado para a cena e, se ele alinhar perfeitamente com um objeto na cena, aquele objeto é declarado como reconhecido e segmentado da cena. Este processo é repetido até que a cena esteja completamente segmentada. Os experimentos foram realizados em dados sintéticos e reais. Os resultados dos experimentos que usaram dados sintéticos são demonstrados em termos de taxa de reconhecimento por porcentagem de oclusão e desordem. Os experimentos que utilizaram dados reais, foram comparados com o algoritmo spin image [16] em termos de taxa de reconhecimento por porcentagem de oclusão.

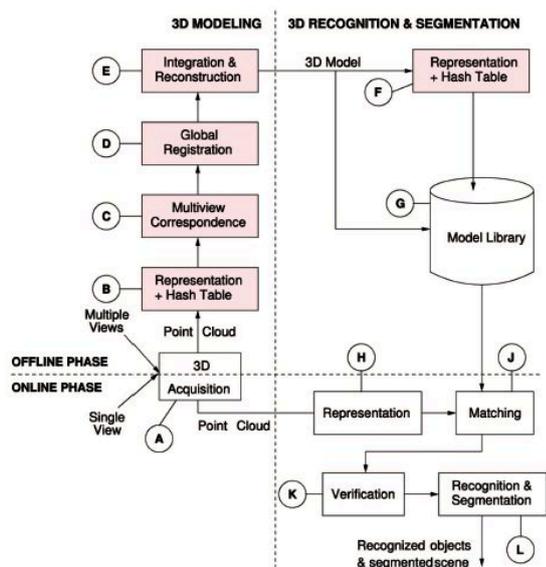


Fig. 202. Representação do método proposto por Mian. Figura extraída de [81].

Ben-Yaacov [593] propõe um método de reconhecimento de objetos 3D com base em Implicit polynomials (IPs). O método apresentado utiliza uma representação de tensores de IPs afim de derivar um conjunto de rotações invariantes. Deste modo, desenvolveu-se um novo conjunto de rotações invariantes 3D IP, assim como expressões de forma fechada destas invariantes. Seguindo o método de reconhecimento 2D IP, Multi-Order and Fitting Error Technique (MOFET), propõe-se um classificador com base nessas invariantes de rotação 3D IP. O classificador proposto usa vários graus de IP afim de utilizar a estabilidade dos IPs de baixo grau e o poder descritivo dos IPs de alto grau. Deste modo a abordagem de reconhecimento proposta é uma abordagem com base em modelo, dado que as características do classificador são combinações dos coeficientes do IP e o IP é uma representação

paramétrica da superfície do objeto. O classificador escolhido tem como base as funções de densidade de probabilidade (FDP) dos vetores de características. Assume-se que o vetor de característica possui uma distribuição gaussiana. Cada função FDP é estimada do vetor de característica pertencente a uma ou mais diferentes visões do objeto do dicionário. Portanto, cada objeto no dicionário é representado por uma ou mais FDP. Os experimentos realizados com o método proposto iniciaram com simulações com dados sintéticos com a finalidade de analisar a sensibilidade do processo reconhecimento. Depois, utilizou-se um conjunto de dados adquiridos, utilizando um equipamento com base na técnica de luz estruturada, com 40 objetos em 9 posições diferentes e comparou-se o método proposto com outros métodos (Shape Spectrum Descriptor (SSD) e outros métodos de estimativa de pose com base em ajuste de IP) em termos de estimativa de pose.

Um método para o reconhecimento de objetos invariantes a deformação inelástica é proposto por Smeets [594]. Este método representa os objetos 3D por Diffusion Distance Tensors (DDT), isto é, tensores de terceira ordem contendo a distância média de difusão para diferentes tempos de difusão entre cada par de ponto na superfície do objeto. Em adição ao DDT, utiliza-se a Geodesic Distance Matrices (GDM) para representar os objetos independentes do quadro de referência. Ao transformar esses tensores de distância em representações modais, fornece-se um descritor de formato invariante a ordem da amostragem. Deste modo, o reconhecimento do objeto se reduz a comparação direta das representações modais sem a necessidade de estabelecer correspondência explícita de pontos. O método proposto de representação foi validado utilizando as bases de dados TOSCA [153] e SHREC 2010 [595], comparando os resultados, de diferentes configurações dos dois descritores, em termos das curvas Receiving Operating Characteristic (ROC) e da Cumulative Matching Curve (CMC).

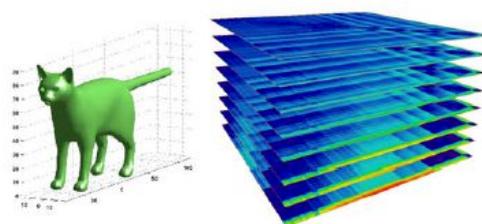


Fig. 203. Representação do método proposto por Smeets. Figura extraída de [594].

Uma técnica para o reconhecimento de objetos 3D para construir vetores de característica invariantes a rotação para high angular resolution diffusion imaging (HARDI) é proposto por Yaniv Gur [596]. Este método utiliza expansão de spherical harmonics (SH) e tem como base a geração de tensores contravariantes utilizando os coeficientes do SH, contraindo eles com os tensores covariantes para obter as invariantes. A técnica proposta permite a construção sistemática de invariantes para expansões SH de qualquer ordem utilizando simples operações matemáticas (soma e multiplicação de coeficientes). O primeiro conjunto de invariantes são construídos

pela contração de coeficientes SH. Sendo estes coeficientes de expansão, eles são componentes contravariantes de um vetor, logo, através da contração deles com suas contrapartes covariantes, constrói-se um tensor de ordem 0 (um escalar), o qual é invariante a rotação. Apesar de sua usabilidade [597], [598], [599], esses descritores são suscetíveis a ruído e podem gerar classificações erradas. No entanto, essas invariantes são um caso particular de um conjunto de invariantes, as quais são construídas pela mistura de tensores de diferentes subespaços e a contração deles. Os elementos centrais são os coeficientes Clebsch-Gordan (CG), os quais são utilizados para acoplar subespaços. Utilizando esses coeficientes pode-se construir novos tensores contravariantes, e estes tensores podem ser contraídos com os coeficientes SH para criar novos conjuntos de invariantes. Adicionalmente, eles podem ser contraídos uns com os outros para construir outro conjunto de invariantes. O processo de construção de tensores e contração deles pode ser continuado como desejado para produzir mais invariantes. Esta técnica foi utilizada, no trabalho apresentado, para construir vetores de características para fiber orientation distributions (FODs) de oitava ordem reconstruídas utilizando constrained spherical deconvolution (CSD). Os experimentos foram realizados em dados sintéticos e de cérebro. Nos dados simulados mostra-se que estas invariantes são robustas ao ruído e podem classificar voxels com base no número de compartimentos das fibras e suas difusividades. Nos dados reais de cérebro, mostra-se que as invariantes capturam informações anatômicas significantes e que podem ser utilizadas para medir a integridade da matéria branca.

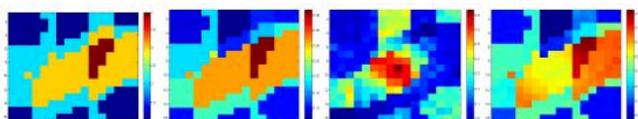


Fig. 204. Representação do método proposto por Yaniv Gur. Figura extraída de [596].

Orts-Escalano [600] propõe uma implementação, com base em hardware, da extração de características 3D e do reconhecimento de objetos 3D para acelerar estes métodos e, conseqüentemente, o pipeline inteiro de sistemas de visão computacional, onde tais características são utilizadas. O descritor de características proposto para implementação em GPU tem como base o descritor apresentado em [601]. Este descritor extrai um modelo semi-local da cena, computando características semi-locais que auxiliam o reconhecimento do objeto até mesmo em condições de oclusão. Esta característica extraída é referenciada como tensor. Um tensor é definido pela superfície do modelo que é atravessada por cada *voxel* em um *grid* centralizado. Este conjunto de valores definem um tensor de terceira ordem. Para computar este descritor, os seguintes passos de pré-processamento, os quais também foram implementados em GPU, se fazem necessários: transformação dos mapas de profundidade e cor para uma nuvem de pontos colorida, remoção de ruído (filtro bilateral), estimativa de normais da superfície e reconstrução da superfície através da triangulação dos pontos da nuvem de

pontos, o que fornece uma malha aproximada da nuvem de pontos (mesh). Com as informações das normais da nuvem de pontos e com a triangulação da superfície obtidas realiza-se o cálculo dos tensores 3D. Todos os tensores extraídos de diferentes visões parciais do objeto são armazenados com suas coordenadas base, permitindo o uso desta informação for agrupar todos os tensores com ângulos similares entre suas normais. Desta forma uma correspondência eficiente é possibilitada. Esta coleção de tensores é armazenada durante a fase de treinamento, criando uma tabela *hash* para uma recuperação rápida durante a fase de teste. A tabela *hash* é montada de forma eficiente a partir dos tensores de múltiplas visões e ela é preenchida utilizando o ângulo dos tensores como chave. Esta técnica de *hash* aparenta ser uma forma eficiente e apropriada de realizar a correspondência entre tensores devido ao baixo custo computacional comparado com as técnicas de correspondência lineares. Esta tabela é armazenada na GPU para fins de implementação do método de correspondência também na GPU. Na fase de correspondência, após o tensor ser extraído da cena, lança-se um número de *threads* equivalente ao número de tensores armazenados no *bin* correspondente. Isso permite checar em paralelo todos os tensores armazenados no *bin* e portanto acelera-se o processo de correspondência. Os experimentos realizados procuraram mostrar a velocidade alcançada com a implementação em GPU e a acurácia alcançada, demonstrada por uma matriz de confusão, pelo mesmo em termos de reconhecimento de objetos sozinhos na cena e em ambientes com vários objetos.

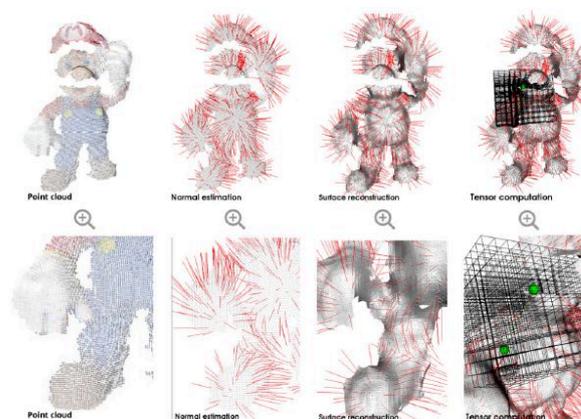


Fig. 205. Representação do método proposto por Orts-Escalano. Figura extraída de [600].

H. Outras formas de representação

1) *Conjuntos de filtros de correlação*: Chye-Hwa Loo [602] propõe um algoritmo de classificação de objetos 3D com base em uma técnica de fringe-adjusted joint transform correlator (FJTC) [603] para reconhecimento automático de alvos em tempo real. A técnica de classificação emprega um reconhecimento com base em fragmentos [604] e um novo tipo de synthetic discriminant function (SDF), denominado WSDF, na geração dos conjuntos de filtros de correlação invariantes a distorção. Então emprega-se o FJTC para fornecer a correlação do conjunto de filtros com a entrada sob um arranjo adequado.

O sistema em tempo real pode ser descrito como a junção da arquitetura descrita em [605] adequada ao sistema de classificação proposto em [606], o qual utiliza o conjunto de imagens para o treinamento e posterior classificação da imagem de entrada por meio dos filtros e da correlação destes filtros das imagens de referência com a imagem de entrada por meio do FJTC. Para avaliar a performance do algoritmo proposto um programa de simulação detalhado foi desenvolvido, mostrando os resultados das simulações em termos de peak to sidelobe ratio (PSR) para os filtros WSDF e SDF e o valor de correlação máxima em relação as classes de objetos presentes na imagem.

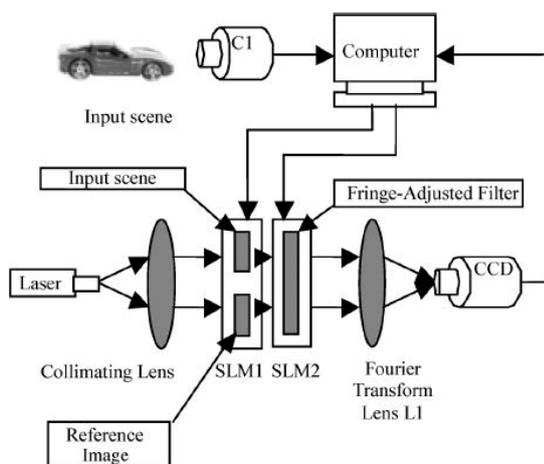


Fig. 206. Representação do método proposto por Chye-Hwa Loo. Figura extraída de [602].

2) *Content-Adaptive Pyramid*: Uma forma de representação para a classificação de imagens 3D é introduzido por Kounalakis [607]. A representação proposta tem como base uma pirâmide que se adapta ao conteúdo da imagem e utiliza as regiões da imagem ao invés de escalas retangulares da pirâmide. As características da imagem, tais como profundidade e cor, são utilizadas para definir regiões dentro da imagem. Regiões de múltiplas escalas são formadas afim de construir a proposta representação da imagem por pirâmides. Com a finalidade de construir a pirâmide adaptativa ao conteúdo, cada célula do grid retangular é rotulada e estas utilizadas para formar regiões através do agrupamento em cada nível da pirâmide. A correspondência de cada bloco para cada escala individual é indicada através dos rótulos das regiões, nomeadas region indexes. Para o cálculo de correspondência da representação proposta utiliza-se uma reformulação da formula geral proposta para Spatial Pyramid Matching kernel (SPM), para levar em conta as escalas que estão em um formato de regiões arbitrariamente moldadas que se adaptam com o conteúdo da imagem, resultando na representação proposta, denominada content-adaptive pyramid, e no método de correspondência denominado content-adaptive pyramid matching (CAPM). Para os testes com a representação proposta utilizou-se a base de dados RGB-D [299] e os resultados, do SPM e do CAPM, foram comparados com diversas configurações de características associadas ao treinamento e classificação de

dois classificadores, SVM e o SVSVM, aplicados no contexto de classificação de objetos.

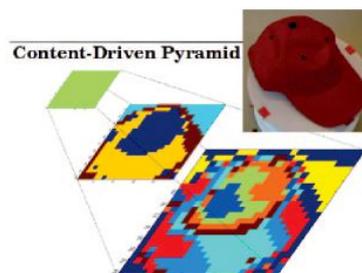


Fig. 207. Representação do método proposto por Kounalakis. Figura extraída de [607].

3) *Curvas de nível*: Uma abordagem para recuperação de objetos 3D parciais com base na correspondência de curvas é apresentado por Mahiddine [608]. Inicialmente é gerado, na etapa offline, um conjunto de dados utilizando uma esfera de visão para extrair curvas de nível em diferentes pontos da visão. As curvas de nível são um conjunto de contornos 2D planares, os quais são a projeção dos pontos em vários planos perpendiculares. Depois, as curvas de nível de cada objeto parcial inquerido são extraídas e comparadas com um conjunto de curvas de nível que definem um objeto 3D da base de dados. O algoritmo de correspondência utilizado para comparação entre as curvas tem como base o alinhamento de curvas planares 2D através do uso de propriedades intrínsecas das curvas (curvatura e tamanho do arco) e foi inspirado no trabalho apresentado em [609]. O número de curvas correspondidas entre o objeto parcial e o objeto completo representa o peso daquela classe. A classe com maior peso é identificada como a classe do objeto inquerido. Experimentos de recuperação de objetos 3D parciais, utilizando a base de dados [610], foram realizados e os resultados demonstrados na forma de matriz de confusão, *Mean square error* e similaridade entre o objeto parcial 3D inquerido e os modelos na base de dados.

4) *Depth Aspect Image*: Kitaaki [611] propõe um método para reconhecimento de objetos 3D com alta velocidade de processamento com base na correspondência do DAI (Depth Aspect Image) [612] e no Hierarchical Modified Iterative Closest Point (HM-ICP) [613]. O método de reconhecimento de objetos proposto consiste basicamente de duas etapas, estimativa de pose e posicionamento, ou seja, correspondência DAI para o passo mais grosseiro e o HM-ICP para o passo mais fino. No contexto de alta performance utiliza-se uma implementação em GPU dos métodos, para acelerar o processamento. O fluxo do método proposto inicia com a construção da base de dados de modelos de visões 3D denominada DAI, e que é gerada de um ponto de vista virtual. Após a construção da base de dados, procura-se a posição de melhor correspondência e a pose do alvo com base na correlação entre a visão alvo e a visão da base de dados. Depois, para um registro mais preciso utiliza-se o HM-IPC, o qual tem como base uma versão modificado do algoritmo ICP, o M-ICP [614] melhorado por duas outras técnicas. A primeira técnica é um processo hierárquico de registro utilizando múltiplas resoluções dos dados e a segunda utiliza regiões parciais, as

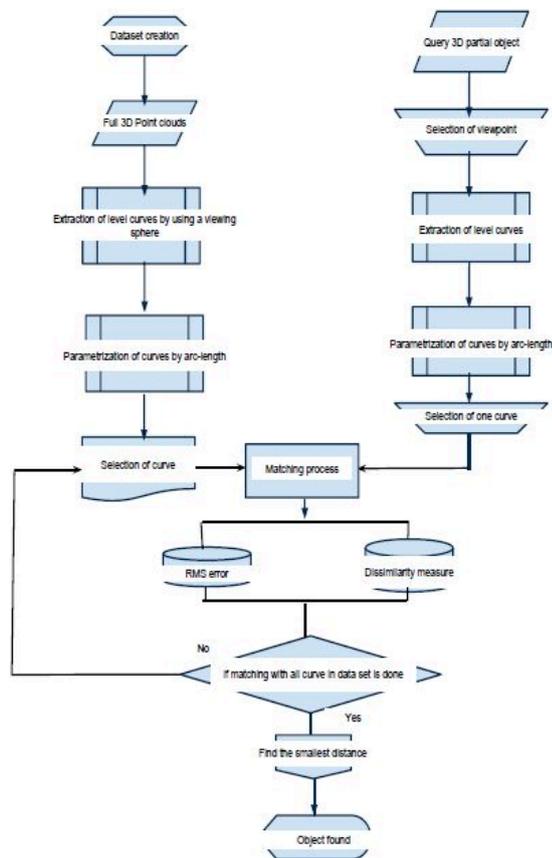


Fig. 208. Representação do método proposto por Mahiddine. Figura extraída de [608].

quais são selecionadas com base no valores de característica para registro. Os experimentos realizados pelo autor tem como foco a comparação dos métodos para CPU em relação aos métodos implementados para GPU.

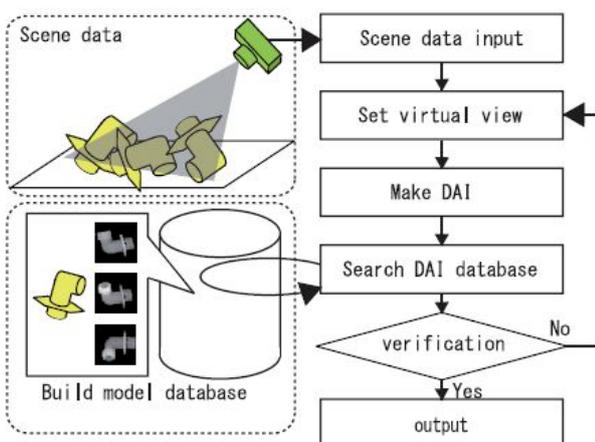


Fig. 209. Representação do método proposto por Kitaaki. Figura extraída de [611].

5) *Descritor volumétrico*: O trabalho apresentado por Gafar [615] mostra um descritor volumétrico para representar objetos 3D. Este descritor representa os objetos através da divisão

do mesmo em conchas, adquirindo a distribuição da área do objeto através destas conchas esféricas. As conchas esféricas são calculadas com base no centro de massa do objeto da seguinte forma. Primeiramente, acha-se o centro de massa do objeto e com base nesse centro de massa gera-se uma esfera de limites. Depois, a esfera é dividida em conchas e dentro de cada concha são calculadas áreas triangulares com a finalidade de computar a área de participação em cada concha. Por fim, a área das conchas é normalizada através da divisão da área de cada concha pela área total do objeto, fornecendo uma independência em relação a escalonamento. Com base neste descritor de área ocorre o processo de descrição e reconhecimento do objeto. Este processo pode ser dividido em duas fases: aprendizagem e reconhecimento. Na fase de aprendizagem calcula-se o vetor de características para cada objeto, proveniente do descritor de área previamente descrito, e armazena-se na base de objetos conhecidos. Na fase de reconhecimento, o vetor de características é também calculado para o objeto sendo pesquisado e então são comparados os vetores de características do objeto sendo comparado com os vetores de características presentes na base. Esta comparação é realizada com base em uma função de distância. O objeto da base de objetos conhecidos que obtiver a menor distância com o objeto sendo pesquisado é considerado a melhor correspondência com objeto sendo pesquisado. Nos experimentos realizados as funções de distância utilizadas foram: Euclidiana, Chi-Square, Intersection e TaniMoto. Os experimentos com o descritor proposto foram realizados em duas bases de dados com modelos de dinossauros (INRIA Gamma research dataset e Princeton Shape Benchmark). Nestes experimentos estudou-se a estabilidade do descritor 3D proposto com relação a ruído, variação na densidade da amostragem, rotação e escala. Também foi analisado a efetividade dos métodos de computação de distância para medir a similaridade entre os descritores dos objetos. Em todos os experimentos calculou-se os dados obtidos em relação ao objeto original. O tempo de computação também foi analisado em relação ao número de faces do objeto.

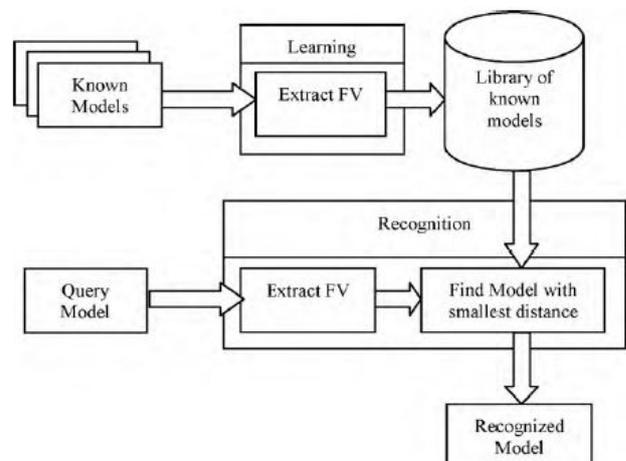


Fig. 210. Representação do método proposto por Gafar. Figura extraída de [615].

Xiaofan Xu [616] apresenta uma eficiente representação volumétrica do objeto 3D, denominada Volumetric Accelerator (VOLA), a qual requer muito menos memória que uma representação volumétrica normal. Adequadamente, VOLA consegue reduzir a complexidade computacional das Convolutional Neural Networks (CNNs) drasticamente quando aplicadas no reconhecimento de objetos 3D. Os CNNs 3D existentes utilizam nuvens de pontos 3D como dado de treinamento [617] ou canais RGB junto com o canal de profundidade para construir 3D CNNs [254]. Entretanto, a representação VOLA é utilizada pelos autores para representar os objetos 3D para as CNNs. Mais especificamente, VOLA lida com grid volumétricos irregulares, conhecidos como voxels, e armazena apenas um único bit de informação para representar cada voxel. VOLA utiliza uma octree como estrutura de dados para armazenar os voxels. Dois conjuntos de dados de dígitos foram introduzidos e utilizados nos testes com a representação proposta (3D-MNIST datasets e 3D-Fonts dataset). Por fim, introduziu-se uma combinação de múltiplos modelos de CNN com base no modelo LeNet[416]. Os quais foram treinados com o conjunto de dados de dígitos utilizando VOLA e comparados em termos de acurácia e tempo de execução.



Fig. 211. Representação do método proposto por Xiaofan Xu. Figura extraída de [616].

6) *Eigen-images*: Xu Zhang [618] propõe um algoritmo de estimativa de pose denominado Isomap-Eigenanalysis-Regression (Isomap-E-R), o qual estima uma pose arbitrária do objeto 3D no espaço de visão. Para o conjunto de treinamento, o *embedding* de baixa dimensão do conjunto de padrões de entrada é computada pelo *Isomap*[485], e as *eigen-images* do *embedding* são deduzidas com base num *eigenspace*. Uma direção de projeção diferente no *embedding* de baixa dimensão é utilizada para melhorar a acurácia da estimativa de pose. As métricas em cada direção, derivadas por regressão linear, são então utilizadas para deduzir a projeção do conjunto de treinamento. Para uma dada amostra de entrada, a sua projeção em *eigen-images* é primeiramente computada, e as imagens de treinamento próximas a elas deduzidas para a amostra de entrada pelo algoritmo, fornecendo assim a estimativa de pose. A performance do algoritmo Isomap-E-R foi analisada utilizando a base de dados the Princeton Shape Benchmark (PSB) [39], da qual 4 classes de objetos foram selecionadas e uma série de poses diversas (imagens de projeção) do modelo foram geradas para representar o objeto 3D eficientemente. Os resultados de correspondência obtidos foram comparados com os resultados de uma versão estendida do Locally Linear Embedding [619].

7) *Holografia*: O trabalho apresentado por Nelleri [620] demonstra uma técnica holográfica para classificação e reconhecimento de objetos 3D. O reconhecimento do objeto 3D é realizado utilizando apenas um holograma digital de Fresnel. Um holograma de cada objeto 3D a ser classificado é gravado eletronicamente utilizando um setup de holografia

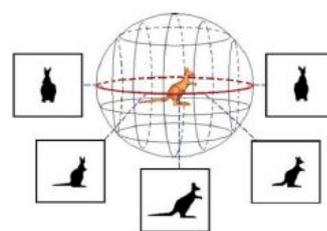


Fig. 212. Representação do método proposto por Xu Zhang. Figura extraída de [618].

digital. Os hologramas eletrônicos são processados com a finalidade de recuperar os valores complexos da imagem 2D correspondendo aos pontos do objeto 3D no plano do objeto. O processo complexo não-linear de recuperação dos pontos intrinsecamente elimina os distúrbios zero order e twin image na imagem reconstruída, permitindo a correlação de apenas uma informação do objeto na etapa de reconhecimento. Depois com a aplicação do Mexican hat digital wavelet matched filtering na imagem reconstruída obtém-se maiores picos de correlação, o que melhora a discriminação quando comparado com outros filtros classicamente utilizados no reconhecimento de objetos 3D. O método proposto foi utilizado para reconhecer e classificar 4 objetos em uma cena em duas classes e os resultados obtidos demonstraram que o método é capaz de discriminar e reconhecer um dos objetos independentemente de sua localização na cena em relação aos outros três objetos.

Kumar [621], propõe um método de reconhecimento de objetos 3D utilizando correlacionadores do método de joint fractional Fourier transform (JFRT), no qual a técnica de holografia digital de Fresnel é utilizada. Técnicas de correlação com base em holografia digital podem automaticamente focar no volume 3D em exame e assim, combinando os benefícios de um correlacionador óptico com holografia digital pode-se aplicar este esquema para o reconhecimento de objetos 3D, como por exemplo objetos 3D microbiológicos. Para desenvolver o correlacionador proposto, um esquema com base em fractional Fourier transform foi utilizado. Inicialmente a holografia digital de uma perspectiva de dois objetos 3D diferentes são simuladas utilizando holografia digital de Fresnel e as imagens 3D correspondentes são reconstruídas numericamente utilizando integral de Fresnel-Kirchhoff para propagação livre no espaço. As imagens reconstruídas são então comparadas com o objeto 3D alvo com a auxílio das duas estratégias de correlacionamento propostas (JFRTC e do não linear JFRTC (NJFRTC)). Para estudar a força das técnicas de correlacionamento para reconhecimento de objetos 3D, uma simulação numérica foi realizada utilizando MATLAB. Algumas medidas de performance (Taxa de discriminação, peak-to-correlation energy e peak-to-sidelobe ratio) foram aplicadas na saída da correlação do JFRTC e do NJFRTC como forma de avaliar os resultados obtidos.

8) *Imagens photon-counting*: Cuong Manh Do [622] apresenta um método para *photon-counting sensing* com imagem integral 3D para o reconhecimento de objetos utilizando independent component analysis (ICA). Uma matriz de lentes é utilizada para capturar múltiplas imagens de perspectiva

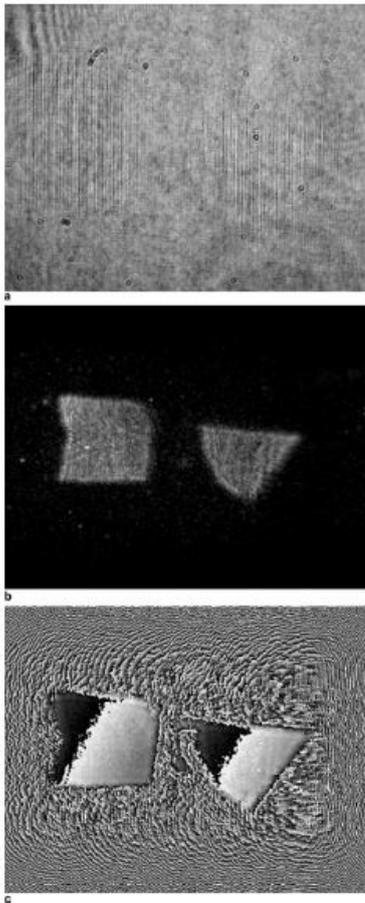


Fig. 213. Representação do método proposto por Nelleri. Figura extraída de [620].

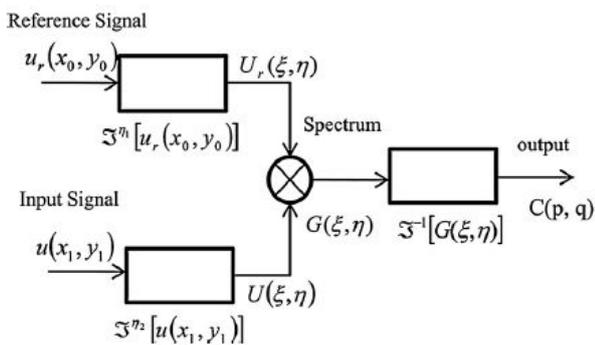
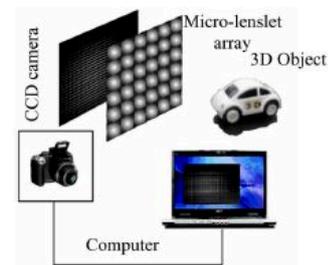


Fig. 214. Representação do método proposto por Kumar. Figura extraída de [621].

da cena 3D projetada em um sensor de imagem. Imagens *photon-counting* da imagem integral capturada são geradas utilizando uma distribuição de Poisson. Um algoritmo com base em kurtosis-maximization é utilizado como um método de maximização não Gaussiano para extrair características independentes do conjunto de dados *photon-counting* de treinamento. As imagens *photon-counting* são pré-processadas utilizando Principal Component Analysis (PCA) para reduzir a dimensão das características extraídas, aumentar a veloci-

dade do passo com o ICA e aprimorar a performance de classificação. Uma imagem *photon-counting* de uma cena de entrada desconhecida é classificada utilizando as métricas K-Nearest Neighbor (KNN) e ângulo cosseno. No método de classificação pelo ângulo cosseno, o ângulo cosseno entre o vetores de teste projetados e os vetores de treinamento projetados, provenientes da transformação pelo ICA aplicado aos vetores PCA projetados, representando os vetores no domínio do ICA, é calculado e a similaridade entre o objeto de teste e o objeto pertencente a classe é obtida com o valor máximo de cosseno. Nos experimentos com o método proposto foram utilizados três carros de brinquedo para o treinamento e teste, comparando os resultados de classificação correta para o método de classificação por ângulo cosseno e para o K-NN com os seguintes números do parâmetro k (3, 5, 10 e 20).



Optical system for 3-D sensing and imaging using II.

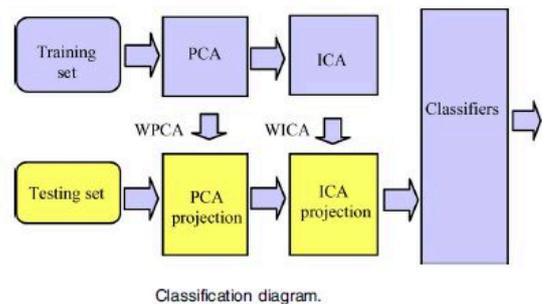


Fig. 215. Representação do método proposto por Cuong Manh Do. Figura extraída de [622].

9) *Multiplexação complexa de amplitude*: Yoshikawa [623] propõe um método de reconhecimento de objetos 3D utilizando multiplexação complexa de amplitude com uma função de suporte. O método de reconhecimento de objetos 3D utilizando descrição complexa de amplitude codifica diretamente a informação de altura do objeto em fator de fase da amplitude complexa com base na Fourier Transform Profilometry (FTP). O reconhecimento do objeto é realizado por correlação 2D utilizando a amplitude complexa, denominando esse método de reconhecimento 3D como Fourier transform Profilometric Correlator (FPC) e a amplitude complexa como Height transformed Complex Amplitude (HCA). O método de multiplexação de mudança espacial habilita obter-se resultados de correlação em muitos objetos 3D através de uma única operação de correlação. Logo, o FPC é aplicável para o reconhecimento de múltiplos objetos. O HCA obtido pelo FTP é utilizado diretamente no FPC convencional, o que

gera um problema dado que o fator de contraste é incluído no HCA. Devido ao fato que o fator de contraste inclui a propriedade de reflectância e as características de superfície do objeto em adição a informação do formato do objeto, informações desnecessárias foram utilizadas simultaneamente no reconhecimento do objeto. Neste caso, vem a proposta de utilizar a função de suporte com base no fator de contraste do HCA, extraindo apenas a informação de formato, desenvolvendo o FPC utilizando apenas informação 3D. O conjunto de dados formado por HCA de um objeto rotacionado tem baixa correlatividade com respeito ao ângulo de rotação. Logo, se a combinação linear deste conjunto de dados é assumido como os dados de referência, tolerância a rotação do objeto pode ser adicionado no FPC. No FPC, considerando que o HCA é a informação 2D, multiplexação acompanhada por deslocamento espacial do ponto de saída é proposto para evitar sobreposição das saídas de correlação. Quando o FPC é executado utilizando dados de referência multiplexados, o reconhecimento de objetos 3D associado com um número de objetos de referência com vários ângulos de rotação pode ser realizado simultaneamente por uma única operação de correlação. Este método é referido como MX-FPC. Para avaliar a performance do MX-PFC com uma função de suporte, utilizou-se a intensidade de correlação normalizada no plano de saída e a taxa máxima de intensidade de correlação do lóbulo lateral nos resultados obtidos no processo de reconhecimento de uma máscara.

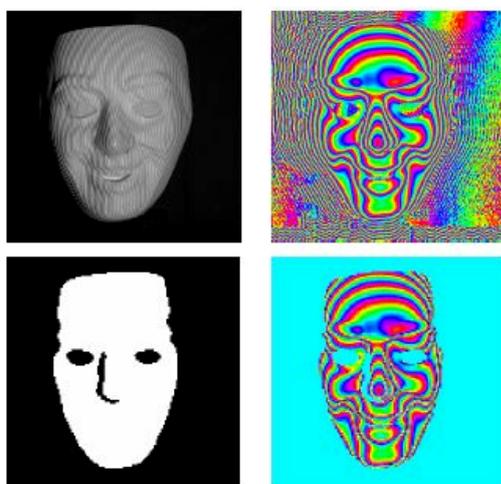


Fig. 216. Representação do método proposto por Yoshikawa. Figura extraída de [623].

10) *Partes planas*: Um *framework* para detecção de múltiplos objetos partindo de uma simples imagem e para a avaliação sobre oclusões entre objetos é proposto por Yu Xiang [624]. Este modelo apresentado é denominado Spatial Layout Model (SLM) e possui a habilidade de modelar a interação entre objetos, cena 3D e visão da câmera, especialmente a oclusão entre objetos. Inicialmente, propõe-se um nova representação de objetos 3D utilizando partes planas. Esta representação por partes planas, introduzida em [625], é adequada para a avaliação de oclusões no sentido que elas podem ser aproximadas como visíveis ou não-visíveis. Depois, agrupam-se as partes planares em 3D para representar porções de um objeto.

Cada grupo é denominado como um 3D *aspectlet*, o qual é gerado automaticamente. Os 3D *aspectlets* podem fornecer evidências mais robustas de observações parciais em oposição as partes planares por si só. Por fim, hipóteses de localização e pose em 3D dos objetos e da câmera são geradas e verificadas através da combinação do conhecimento a priori e evidências dos 3D *aspectlets*. Esta etapa é realizada através de uma estratégia de amostragem de Markov Chain Monte Carlo, onde diferentes tipos de movimentos são projetados para explorar o espaço de hipóteses de forma eficiente. Neste processo, os 3D *aspectlets* são pesados de acordo com os padrões de oclusão induzidos pela hipótese 3D. Consequentemente, combina-se a evidência dos 3D *aspectlets* com avaliação de oclusão para auxiliar na detecção dos objetos. Os experimentos foram conduzidos em duas bases de dados, uma com cenas externas com carros e outra com cenas internas com móveis, onde múltiplos objetos são observados sob vários graus de oclusão. E os resultados do SLM foram comparados com outros dois métodos, Deformable Part Model (DPM) e Aspect Layout Model (ALM), em termos de performance na detecção de objetos com base na precisão média obtida.

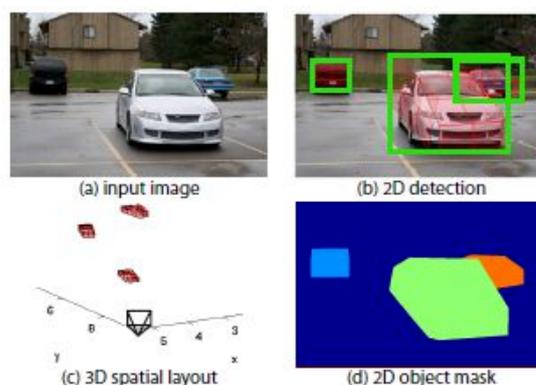


Fig. 217. Representação do método proposto por Yu Xiang. Figura extraída de [624].

11) *Pseudo-Random Binary Sequences*: Ravari [626] apresenta um método para comparar objetos 3D por meio de uma medida de similaridade normalizada. Inicialmente, a nuvem de pontos ou imagens de profundidades equivalentes adquiridas são pré-processadas para a remoção de *outliers* e observações inválidas do sensor. Então, tendo a nuvem de pontos em mão, um método de ajuste por least-squares NURBS é realizado para construir um modelo Non Uniform Rational Basis Spline (NURBS) [627]. Depois, uma representação com base na complexidade é construída do modelo NURBS. Os seguintes passos são realizados para a construção desta representação: Primeiro os os pontos centrais de controle são computados e as coordenadas são armazenadas em um sistema de coordenadas esférico, após a conversão dos mesmos para tal sistema. Em seguida, os vetores paramétricos são armazenados para ter uma descrição completa do modelo NURBS. Desta representação o modelo NURBS original pode ser construído sem perda de informação. Uma vez que uma representação com base na complexidade do NURBS é feita das coordenadas dos pontos de controle central, sua complexidade Kolmogorov

[628] pode ser representada pelas somas das complexidades dos pontos centrais de controle, dos pontos de controle e dos valores de parâmetros. Logo, o modelo NURBS das nuvens de pontos pode ser denotado por uma série de *strings* pseudo-random binary sequences (PRBS) [629], de forma similar ao trabalho anterior realizado pelo autor no qual ele realiza a classificação não supervisionada de objetos 3D em imagens de profundidade [630], com relevante Kolmogorov complexidade. Por fim, o resultado de reconhecimento do objeto 3D é determinado pela comparação da medida de similaridade com um limiar. Para distâncias de similaridade abaixo de um limiar especificado, o reconhecimento é positivo caso contrário o objeto é reportado como desconhecido. O experimento realizado utilizou uma base de dados com 300 objetos com 51 características diferentes [457] e os resultados obtidos com o método proposto foram comparados com os resultados obtidos com um método de classificação de objetos com base em características apresentado em [631]

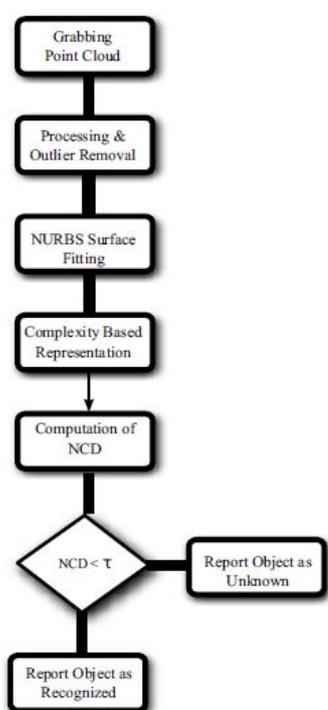


Fig. 218. Representação do método proposto por Ravari. Figura extraída de [626].

12) *Sliced Curvature Scale Space*: Billy Okal [632], explora o uso da teoria espaço escala [633] via o espaço de curvatura do espaço e o estende para representar objetos 3D no *framework* Sliced Curvature Scale Space (SCSS). A representação via SCSS é uma extensão do curvature scale space (CSS) [634], a qual foi aplicada com sucesso em casos 2D. A representação por CSS é desenvolvida envolvendo repetidamente um sinal com um kernel gaussiano. Dado uma curva 2D parametrizada, a curvatura em um dado nível é computa através de um processo conhecido como evolução. A imagem CSS é então feita pela plotagem da localização do *zero-crossings* do sinal da curvatura para todos os níveis. Para

estender o CSS adotou-se um mecanismo de fatiamento através do qual se vê o objeto 3D como um conjunto de fatias de placas finas infinitamente próximas empacotadas junto. Pode-se então aplicar a técnica comum do CSS para cada uma das fatias e utilizar uma forma de combinar os resultados em uma características coerente a qual denomina-se SCSS. Para utilizar a representação SCSS no reconhecimento de objetos e configurações de registro, a quantidade de informação utilizada para codificar um único objeto necessita ser reduzida. Para tanto, explora-se o uso de técnicas padrão de redução de dimensionalidade (Kernel principal component analysis (KPCA) [635] e Nonnegative Matrix Factorization (NMF) [636]). Para fins de reconhecimento de objetos utilizando a representação proposta, emprega-se o Support Vector machine para a classificação de objetos 3D em uma configuração de classificação de múltiplas classes. Para demonstrar a performance da representação proposta utilizou-se um conjunto de dados de cinco classes de objetos e empregou-se uma estratégia de um contra o resto para cada experimento de classificação e repetiu-se esse experimento para cada método de redução de dimensionalidade mostrando como resultado as taxa de classificação obtidas.

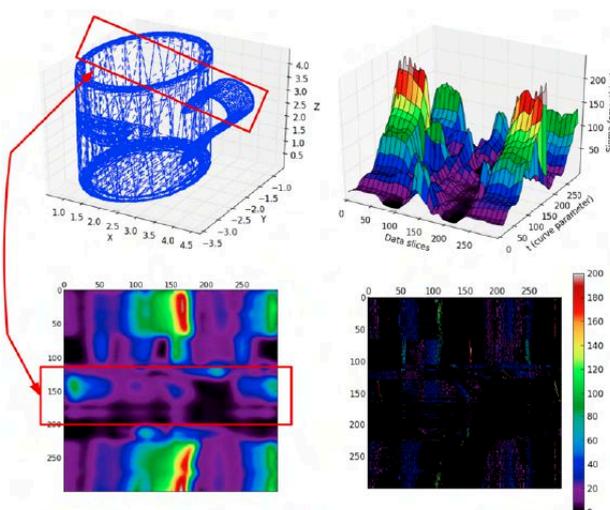


Fig. 219. Representação do método proposto por Billy Okal. Figura extraída de [632].

13) *Templates*: Byung-Gook Lee [637], propõe um sistema computacional melhorado para imageamento integral através da eliminação de oclusões na imagens gravadas do objeto 3D parcialmente ocluído e da recuperação integral das imagens do objeto 3D. Primeiramente, são obtidas as imagens do objeto parcialmente ocluído utilizando um sistema computacional de imageamento integral e estas são transformadas para sub-imagens. Depois, são eliminadas as oclusões com uma técnica de remoção de oclusões e para compensar a parte removida devido as oclusões, utiliza-se uma aplicação recursiva da reconstrução por principal component analysis e da compensação do erro. Por fim, gera-se a imagem como um todo e executa-se o processo de reconhecimento do objeto. O sistema proposto é composto de duas fases. A primeira fase é a fase offline, na qual os templates de referência

para o objeto 3D a ser reconhecido são gerados e seus autovetores são calculados a partir de um conjunto de sub-imagens de treinamento. A outra parte é o processo online, onde a oclusão é removida (utilizando um algoritmo de block matching, bastante conhecido na área de visão estereó), o erro da remoção da oclusão é compensado recursivamente (utilizando os autovetores gerados no processo offline) e o reconhecimento do objeto é executado, através de uma correlação cruzada entre a imagem reconstruída sem a oclusão e os templates de referência. A técnica proposta foi testada em experimentos para reconhecimento de faces e avaliada com base em um cálculo de PSNR, o qual tem como base a média dos erros ao quadrado.

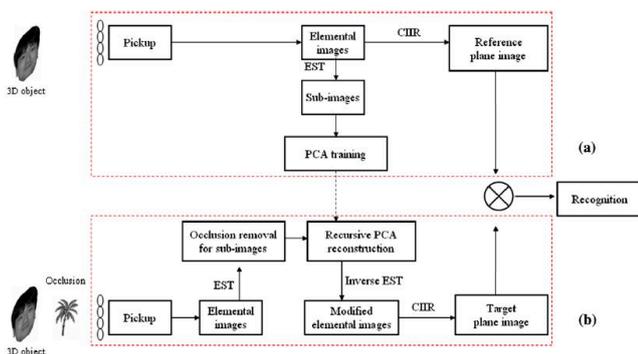


Fig. 220. Representação do método proposto por Byung-Gook Lee. Figura extraída de [637].

Chuantao Zang[638] propõe uma estratégia de rastreamento visual para o reconhecimento de objetos 3D e rastreamento. A estratégia utilizada é uma estratégia de estimativa grosseira e refinamento. Inicialmente, adota-se um método de correspondência com base em *templates* para estimar um valor inicial de pose para o objeto. Este passo é obtido através da comparação da imagem atual (inquerida), com as imagens *templates* renderizadas utilizando OpenGL com diferentes poses virtuais da câmera. O processo de renderização utiliza um modelo CAD de um objeto conhecido e os parâmetros da câmera (parâmetros intrínsecos e extrínsecos da câmera), para retirar uma foto do objeto de uma câmera virtual. Quando os parâmetros intrínsecos e o objeto estão fixos, a imagem renderizada pode ser expressa como $I(E)$ dado que ela muda com os parâmetros extrínsecos E . Uma imagem é capturada com uma câmera real com uma câmera de pose desconhecida E^* . Portanto, a estimativa de pose inicial da câmera consiste em encontrar o parâmetro E que minimiza a diferença entre as imagens renderizada $E(I)$ da câmera virtual e a imagem atual I . No geral este problema é solucionado com uma minimização da diferença dos mínimos quadrados. Com esta estimativa inicial, um método de rastreamento de bordas é explorado para obter uma estimativa final de pose. Adota-se o método de rastreamento com base em bordas proposto por [639]. Nesta etapa, o problema de computação de pose consiste em estimar o movimento entre *frames* entre a pose atual e a próxima pose. Como o erro inicial já foi reduzido no passo grosseiro, este passo de refinamento pode assegurar uma solução local otimizada mais exata. Na aplicação OpenGL pro-

posta, a imagem renderizada do objeto é preenchida com um algoritmo de *flood-fill*, portanto apenas o modelo geométrico do objeto é necessário. Para comparar a imagem atual com a imagem renderizada, um pré-processamento da imagem atual da câmera real é conduzido. Esse pré-processamento inclui subtração de plano de fundo, binarização da imagem, dilatação, extração do contorno do objeto e *flood fill*. Dois experimentos foram realizados para avaliar o método proposto. No primeiro experimento um objeto regular e sem textura foi utilizado e no segundo experimento utilizou-se uma impressora como objeto, estimando as poses dos dois objetos. Um terceiro experimento testou o método proposto contra outro método que utiliza apenas o contorno. Os resultados foram apresentados em termos de erro de estimativa e resultados visuais.



Fig. 221. Representação do método proposto por Chuantao Zang. Figura extraída de [638].

14) *Várias formas de representação*: Muja [640] apresenta a implementação de uma infra-estrutura, ReIn (REcognition INfrastructure), capaz de combinar várias técnicas de reconhecimento de objetos 3D/2D e estimativa de pose em paralelo como *plugins* dinamicamente carregáveis. Ela também fornece uma arquitetura eficiente para passagem de dados e oferece a possibilidade de mudar os parâmetros e configurações iniciais destas técnicas durante suas execuções. Adicionalmente, dois classificadores projetados para a percepção dos robôs, BiGGPy (Binarized Gradient Grid Pyramids) e VFH (Viewpoint Feature Histograms) [145], foram introduzidos. Estes dois classificadores podem ser facilmente combinados utilizando o ReIn para resolver o problemas de reconhecimento de objetos e identificação de pose. Na ReIn um algoritmo é visto como uma caixa preta, com uma interface bem definida que utiliza um conjunto de entradas, produz algumas saídas e está configurado por um conjunto de parâmetros. A arquitetura geral do ReIn é composta pelas seguintes partes: Operador de atenção: tem como entrada uma imagem e/ou uma nuvem de pontos e produz uma máscara de saída, uma região de interesse na imagem ou a segmentação da nuvem de pontos. O operador de atenção é usualmente posto na frente de um detector para encontrar regiões interessantes na imagem/nuvem de pontos para realizar a detecção, reduzindo, deste modo, o espaço de procura do detector. O detector: tem como entrada uma imagem, uma nuvem de pontos, uma lista de máscaras/regiões de interesse ou uma lista de detecções, e produz como saída uma lista de detecções e potencialmente uma lista de poses. Dado que diferentes algoritmos de detecção podem requisitar apenas algumas das entradas, as entradas podem ser utilizadas em qualquer combinação, configurável por um conjunto de parâmetros. O estimador de pose: tem como entrada uma imagem e/ou uma nuvem de pontos e um conjunto de detecções e computa a pose dos objetos detectados. Um estimador de pose é utilizado quando a pose é requerida para tarefas de pegar, mas os algoritmos de detecção utilizados não são capazes de computar

as poses dos objetos detectados. Adicionar algoritmos de atenção, detecção ou estimativa de pose a esta infra-estrutura é realizado através da adição de um invólucro a estes algoritmos para que estes implementem as interfaces descritas. Uma vez com o invólucro eles podem ser livremente combinados em diferentes configurações pelo redirecionamento de suas saídas e entradas. Uma vantagem adicional de envolver os algoritmos existentes nesse invólucro na infra-estrutura apresentada, é o fato de que eles automaticamente se tornam *plugins*, capazes de serem dinamicamente carregados/descarregados do sistema. Dado que ReIn é construída em uma arquitetura distribuída de passagem de mensagem, é simples de configurar os papéis que classificadores vão ter. A configuração do BiGGPy como classificador e o VFH como filtro é mostrado no trabalho em questão. Mostra-se também como pode-se reverter os papéis de forma que o VFH aja como classificador principal e o BiGGPy como filtro. Para validar o ReIn implementou-se uma aplicação exemplo utilizando o BiGG (Binarized Gradient Grids) e sua extensão piramidal (BiGGPy) e o VFH. Mostra-se, também como o BiGGPy funciona e os resultados obtidos da combinação entre o BiGGPy e o VFH em termos de valores de precisão e lembrança no processo de reconhecimento de objetos.

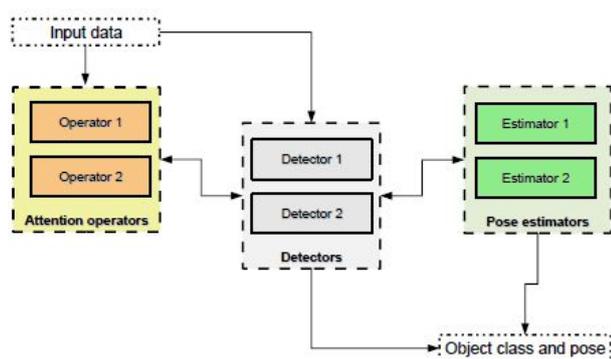


Fig. 222. Representação do método proposto por Muja. Figura extraída de [640].

I. Outros trabalhos

1) *Método de correspondência 3D - 2D*: Marques [641] propõe uma solução para o problema de correspondência entre os pontos 3D de um objeto e suas projeções 2D. Para tanto, define os princípios gerais sob os quais uma solução única e ótima é obtida para o problema de correspondência 3D-2D. Estes princípios permitem construir um problema convexo de otimização (sem mínimo local). A solução possui um custo computacional menor que [642], pois a restrição possui uma formulação linear e os métodos simplex ou *interior-point* podem ser utilizados. O algoritmo proposto para solucionar o problema 3D-2D foi denominado LAMP. O LAMP age de forma global logo a influência de ruídos locais tem menos impacto no resultado. Com este algoritmo o autor demonstra uma forma simples de implementar a solução teórica proposta. Para avaliar o LAMP compararam-se os resultados obtidos por ele com os resultados obtidos pelo método proposto em [642]

em termos de correspondências erradas em relação ao número de pontos.

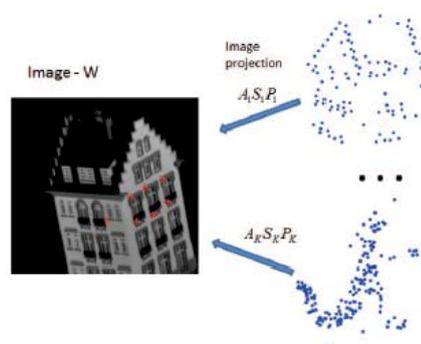


Fig. 223. Representação do método proposto por Marques. Figura extraída de [641].

2) *Montagem de base de dados*: Em [204], Wohlkinger apresenta o funcionamento, a metodologia e todas as características bem como a construção da base de dados 3DNet.

3) *Revisões e análises*: O estudo apresentado por Salti [643] mostra uma avaliação de performance de detectores de pontos 3D chave em relação a tarefa de reconhecimento de objetos 3D. O autor divide, inicialmente, os detectores 3D em duas categorias: detectores com escala variante e detectores invariantes a escala. Para as duas categorias são fornecidas definições básicas as quais descritores pertencentes a cada categoria possuem e alguns exemplos são citados explicando de forma breve o funcionamento de alguns descritores. Como etapa seguinte, o autor explica a metodologia utilizada para avaliar os descritores citando as bases de dados utilizadas nos experimentos (Stanford Repository [151] e as bases de dados usadas na validação em [150] e [138]), as características sendo avaliadas nos descritores (repetibilidade relativa e absoluta), os descritores selecionados para avaliação (Local Surface Patches (LSP), Intrinsic Shape Signatures (ISS) e KeyPoint Quality (KPQ) para os de escala fixa e MeshDoG, Laplace-Beltrami Scale-Space e KeyPoint Quality scale-invariant (KPQ-SI) para os invariantes a escala) e os parâmetros utilizados. Por fim na parte de avaliação dos métodos foram mostrados os resultados obtidos para cada método nas diferentes bases de dados utilizadas em termos de tempo de execução e repetibilidade absoluta e relativa com e sem a presença de ruído.

O estudo apresentado por Elizabeth González [644], mostra uma análise qualitativa e quantitativa sobre a performance de métodos de reconhecimento de formato 2D quando utilizados em problemas de reconhecimento de objetos 3D. Inicialmente, o estudo faz uma descrição das características desejáveis que um modelo de representação de formatos 2D deve ter quando aplicado a um sistema de reconhecimento de objetos 3D. Com base nestas características, o estudo mostra uma tabela relacionando alguns trabalhos, seus tipos de descritores e as características que eles possuem. O estudo também cita como identificação o processo em que um formato consultado é classificado em uma base de dados e define que este processo pode ser alcançado com base em técnicas de correspondência ou de classificação. Neste tópico, são citadas algumas das

propriedades que as medidas de similaridade deveriam ter, finalizando o tópico com uma tabela com varias medidas de similaridade. Outro tópico também relacionado no início do estudo é o de estimativa de posição do formato, no qual são citadas três metodologia utilizadas (com base em modelos, com base em aprendizagem e com base em transformações geométricas) e justificada o porquê da escolha de analisar métodos com base em modelos. Depois, o estudo passa para a descrição de quais formas de representação de formato 2D foram escolhidas (Fourier Descriptors (FD), Boundary Moments (HM), Integral Invariants (II), Shape Context (SC), Zernike Moments (ZM), Generic Fourier Descriptors (GFD), Complex Moments (CM)), quais medidas de similaridade foram utilizadas (Euclidean (ED), City Block (CB), Cosine of the Angle (C), Mahalanobis (MD), Battacharyyya (BD), Support Vector Machine (SVM)), quais as plataformas experimentais utilizadas (Amsterdam Library of Object Images (ALOI) [13] e 3D Synthetic Library (3DSL)) e quais as medidas de avaliação foram utilizadas para avaliar os resultados de reconhecimento (Hard Recognition Rate (Hr), Weak Recognition Rate (Wr) and Ambiguous Recognition Rate (Ar)). E, por fim, faz um compilado dos resultados obtidos com base no *setup* previamente descrito, comparando em termos taxa de reconhecimento com e sem ruído, fazendo análises dos resultados obtidos pelos métodos de representação e similaridade escolhidos e analisando o custo computacional dos métodos nas duas plataformas utilizadas no estudo.

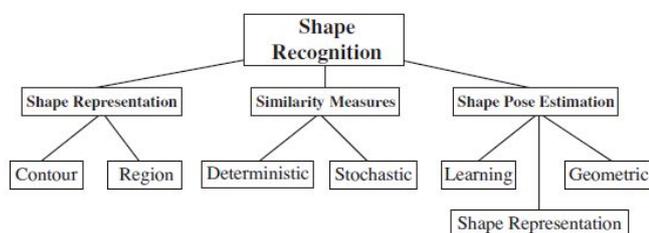


Fig. 224. Representação do método proposto por Elizabeth González. Figura extraída de [644].

Yulan Guo [645] apresenta um *survey* com métodos de reconhecimento de objetos 3D com base em características locais da superfície. Estes métodos geralmente incluem três fases: detecção de pontos chave 3D, descrição da característica local da superfície e correspondência entre superfícies. O autor apresenta uma análise da literatura para cada etapa do processo e também apresenta uma lista com bases de dados populares e contemporâneas evidenciando os atributos relevantes de cada base. Para cada etapa, são analisados os métodos existentes montando secções com as análise e finalizando com um resumo e uma tabela relacionado os métodos analisados. O autor resume sua análise em 29 métodos de detecção de pontos chave, 38 métodos de descrição de superfícies e 17 métodos para correspondência de superfícies. Por fim, o *survey* da uma prospectiva de possíveis futuras direções que a área de reconhecimento de objetos 3D terá.

Um estudo avaliativo sobre descritores 3D é proposto por Yulan Guo[646]. O estudo em questão avalia 10 descritores de características (Spin Image (SI) [647], 3D Shape Context

(3DSC)[98], Unique Shape Context (USC) [149], Rotational Projection Statistics (RoPS) [289], Tri-Spin-Image (TriSI) [364], Local Surface Patch (LSP) [82], THRIFT [292], Point Feature Histogram (PFH)[406], Fast Point Feature Histogram (FPFH)[137], Signature of Histogram of Orientations (SHOT) [150]) no contexto de reconhecimento de objetos, recuperação de formatos 3D e modelagem 3D. Inicialmente, avaliou-se o potencial descritivo dos descritores em 8 base de dados (Retrieval e Random views construídos sinteticamente utilizando modelos da base de dados de stanford [151]; Laser scanner e 2.5D views [81] adquiridos com um scanner a laser; LIDAR [648] adquirido também com um escâner, mas com uma qualidade de imagem relativamente inferior; Dense estéreo coletado com uma câmera estéreo bumblebee de baixa resolução [648]; Space time obtido utilizando a técnica de aquisição estéreo SpaceTime [649]; Kinect [649] gerado utilizando um scanner comercial de baixo custo). Depois, foi a analisado a compacidade utilizando a taxa de lembrança da correspondência de características por cada valor de ponto flutuante no descritor. Também foi avaliada a robustez dos descritores com respeito a variação do raio de suporte, ruído, variação na resolução da malha, distância ao limite da malha, erro na localização do ponto chave, oclusão, desordem e tamanho da base de dados. Além disso, é apresentada a performance do resultado dos descritores quando combinados com diferentes métodos de detecção de pontos chave 3D (Uniform sampling [404], Harris3D [650], ISS [651], e ISS com boundary point removal (ISS-BR)). Por fim, analisa-se a eficiência computacional para a geração de cada descritor.

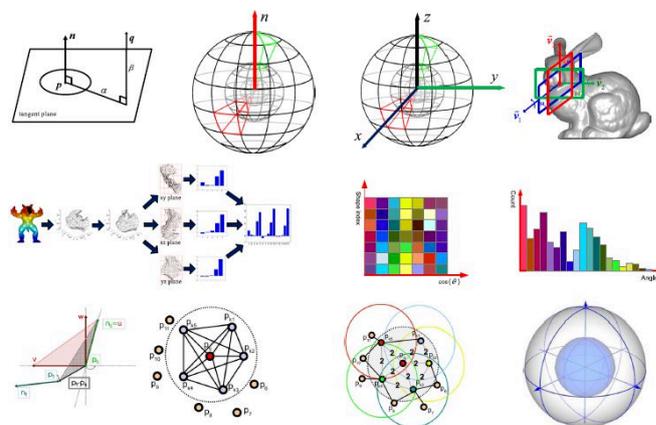


Fig. 225. Representação do estudo avaliativo proposto por Yulan Guo. Figura extraída de [646].

Sharman [652] apresenta uma análise sistemática de vários algoritmos de segmentação e reconhecimento de objetos 3D. Com o objetivo de analisar sistematicamente estes métodos, foram formuladas, inicialmente algumas perguntas de pesquisa. Um número de 150 artigos foram analisados, com base em título, *abstract* e conclusão, e destes 20 foram identificados e analisados na integra. A revisão dividiu os trabalhos em duas categorias, vídeo e imagens. O processo de revisão sistemática foi conduzido de forma similar a [653]. No caso dos algoritmos de segmentação e reconhecimento voltados para imagens, as técnicas mais utilizadas são evoluções de

modelos e algoritmos matemáticos. Um caso similar acontece para os vídeos, onde as técnicas mais utilizadas tem como base em histogramas e algoritmos matemáticos. Com base na análise executada, nota-se que algoritmos mais utilizados para imagem foram: Sparse Convex Combination of Shape Prior Information, Label Transfer, Selective Search Strategy, Boosting Algorithm. Para vídeos as seguintes técnicas foram as mais utilizadas: Unifies Bayesian Perspective, Spatio-Temporal Fusion, Spatially Accurate and Temporal Dense, Bottom-up motion, Path-planning Algorithm. O autor relatou alguns aspectos positivos e negativos sobre essas técnicas, e forneceu uma perspectiva de trabalhos futuros.

J. Livros

Na revisão realizada foram encontrados livros, os quais buscam fornecer desde a parte introdutória do reconhecimento de objetos (*An Introduction to Object Recognition*[654]) até o estudo de métodos e técnicas para o reconhecimento, representação e rastreamento de objetos 3D (*Representations and Techniques for 3D Object Recognition and Scene Interpretation*[655] e *Visual Perception and Robotic Manipulation: 3D Object Recognition, Tracking and Hand-Eye Coordination* [656]).

V. DISCUSSÃO SOBRE OS TRABALHOS ANALISADOS

Com base em todos os trabalhos analisados podemos visualizar duas formas para o reconhecimento/classificação dos objetos, presentes na grande maioria deles. Tanto a primeira forma, ilustrada pela figura 226, quanto a segunda forma, ilustrada pela figura 227, possuem etapas similares (aquisição de dados, pré-processamento - opcional e representação) e diferem na parte de como realizam a classificação/reconhecimento do objeto. Na primeira forma, a representação escolhida para os dados de entrada é correspondida por meio de um método de correspondência, ou de cálculo de similaridade, entre a representação do dado de entrada e as representações armazenadas em uma base de dados, as quais foram calculadas anteriormente em uma fase *offline*, elegendo, com base no maior valor de similaridade ou correspondência, a classificação da representação de entrada ou a representação correspondente a ela na base de dados.

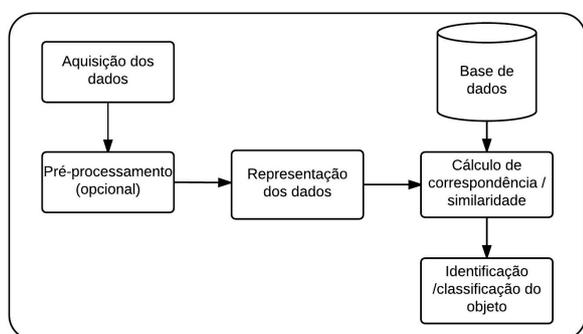


Fig. 226. Pipeline da primeira forma.

Na segunda forma, a representação escolhida serve de entrada para o treinamento de um classificador, realizado

offline ou em alguns casos com treinamento *online*, utilizando posteriormente para classificação/reconhecimento de outros dados de entrada. Cada etapa, dos dois *pipelines* gerais será melhor exemplificada nas seções seguintes.

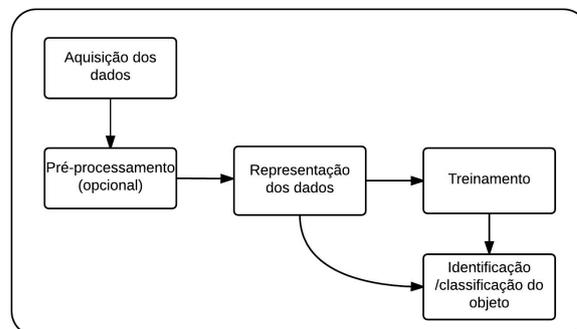


Fig. 227. Pipeline da segunda forma.

A. Aquisição de dados

As técnicas de aquisição dos dados encontradas nos trabalhos analisados partem de aquisição por meio de câmeras [310], par estéreo [18] e Kinect[571] e vão até aquisição por meio de um sensor tátil [553]. Nem todos os trabalhos fizeram aquisições próprias, utilizando nestes casos bases de dados públicas disponíveis [91]. Outros trabalhos realizaram simulações [560], com base no método proposto, ou utilizaram apenas dados sintéticos para os testes [80]. A tabela I mostra para cada trabalho o tipo de dado utilizado.

B. Pré-processamento - opcional

Esta etapa, presente em alguns trabalhos e em outros não, é uma etapa opcional que envolve a possível aplicação dos seguintes métodos:

- Filtragem: para remoção de ruídos [600] ou para alcançar uma melhor amostragem [288];
- Seleção de regiões de interesse: para o foco na área do objeto apenas [640].
- Segmentação: para a simplificação da imagem em termos de segmentos [121].
- Normalização: buscando muitas vezes uma invariância dos dados [258].

C. Representações

As formas de representação são as mais variadas. Dentro de uma determinada categoria apresentada neste trabalho possuímos diversas variações, como por exemplo, nas representações com base em característica existem diversos descritores de característica (SHOT [150], Spin images[16], VFH [145], SIFT [47], entre outros). Outro exemplos estão presentes na forma de representação por modelos, na qual existem modelos de geométricos[534], estatísticos [566], CAD [568], entre outros, e por grafos, no qual existem grafos Reeb [513] e Class Specific Hyper Graph [505]. Como formas de representação mais encontradas nos trabalhos analisados,

temos as com base em características, visões, grafos, modelos, mapas, subespaços e tensores. As representações com menor expressão foram agrupadas em uma categoria que denominamos outras formas de representação. Na tabela VI-A estão relacionados, para cada trabalho, o tipo de representação utilizada. No gráfico 228 mostramos um resumo da quantidade de trabalhos analisados para cada tipo de representação.



Fig. 228. Gráfico do número de trabalhos analisados por tipo de representação.

D. Cálculo de Correspondência/Similaridade

Para o cálculo de correspondência/similaridade, os métodos utilizados nos trabalhos analisados, na sua grande maioria, empregam diretamente um cálculo de correspondência entre descritores definido pelo autor [229] [303] ou empregam medidas de distância para medir a similaridade entre os descritores [201] ou para medir diretamente a distância entre os objetos [305]. Esses métodos de correspondência estão geralmente associados com métodos de votação [191] e/ou alinhamento [229] com a finalidade de gerar e verificar as hipóteses, e eleger efetivamente um modelo objeto.

E. Classificadores

Tanto no reconhecimento/classificação de objetos, os classificadores utilizados são os mais variados. Entre os métodos presentes nos trabalhos analisados destacam-se a utilização dos classificadores por rede neural, na sua grande maioria utilizando o classificador Support Vector Machine [24][173] ou suas variações [28]. Alguns trabalhos utilizam classificadores com base em modelos de *deep learning* [254][310] [424], outros utilizam o K-Nearest Neighbour associado com a distância euclidiana [452] [346] [163]. Há ainda os que utilizam modelo Fuzzy [242] [107], Fuzzy associado com redes Bayesianas [110], Fuzzy associado com redes neurais [117] e modelos probabilísticos [342]. Os classificadores utilizados na parte de reconhecimento geralmente passam por mais uma etapa para eleição do modelo efetiva do modelo.

F. Conjuntos de dados/Bases de dados disponíveis

As bases de dados públicas utilizadas também são as mais variadas. Algumas delas são compostas por modelos 3D de objetos, outras possuem imagens de câmeras RGB-D e tem as que são compostas apenas por diversas imagens 2D de um objeto adquiridas por câmeras em várias posições. Algumas das bases de dados utilizadas nos trabalhos analisados são as seguintes: NYUv2 RGBD [435], TOSCA[153], SHREC 2010 [595], Amsterdam Library of Object Images [13], VidTIMIT [578], ETH-80 [105], base de dados pública com cenas de ruas [569], 3D-REAL-ENV [567], NORB [478], Princeton Shape Benchmark [39], PASCAL VOC Challenge 2005 Cars Test 1, o SUN primitive dataset [544], UCI-Car dataset [542] [550], Weizmann Cars ViewPoint dataset [522], COIL [23], MNIST [511], TUW[499], Challenge[497], 3D Keypoint detection benchmark [81], UWA 3D Object Dataset [138], CMU Oakland 3D Point Cloud Dataset [494], Washington Urban Scenes 3D point Cloud Dataset [495], CMU PIE [473], RGB-D Object Dataset [299], BigBIRD ((Big) Berkeley Instance Recognition Dataset [459], Amazon Picking Challenge at ICRA 2015 [460], stanford [151]; Laser scanner e 2.5D views [81], LIDAR [648], Dense estéreo [648], Space time [649]; Kinect [649], UW-RGBD Object Dataset [457], JHUScene-50 Dataset [382], RGB-D Scenes Dataset v2 [445], UCF YouTube dataset [431], Caltech-101 dataset [432], SHREC 2014 benchmark [433], HMDB51 dataset [434], Bologna e SpaceTime [150], 3Ddata [387], Descriptor Matching Dataset 1 [380], Grand Challenge v2.0 database (FRGC v2.0) [10], Stanford 3D Scanning Repository [351], Bologna Dataset VI [324], Texas 3D Face Recognition Database [307], Berkeley 3D Object (B3DO) [300], McGill 3D Shape Benchmark [170], KIT ObjectModels Web Database [239], PASCAL Challenge Visual Object Class Dataset 2006 [199], base de dados de formatos 3D [516], SHREC'09 [154], base de dados da universidade de Notre Dame (UND) [112], University of California at Riverside (UCR) [113], Purdue Engineering Benchmark [103], base de imagens de profundidade disponível em [68], Protein Data Bank (PDB) [37] e NTU 3D Model Benchmark Dataset [38]. Adicionalmente, em [645], Yulan Guo mostra uma tabela com 16 bases de dados de imagens de profundidade populares.

G. Forma de validação e demonstração dos resultados

Grande parte dos trabalhos analisados utilizaram como métrica de avaliação dos resultados as taxas de reconhecimento e/ou classificação. Alguns trabalhos demonstraram seus resultados em termos de curvas Receiver Operating Characteristics (ROC) e Cumulative Match Characteristic (CMC)[594]. Outros demonstraram com base nos erros de classificação/reconhecimento e no Equal Error Rate (EER) [9]. Na sua grande maioria, sempre comparando estas formas de demonstração dos resultados, dos métodos propostos, com os resultados de outros métodos considerados estado da arte naquele momento.

VI. CONCLUSÃO E PERSPECTIVAS

Com base nos trabalhos analisados nota-se que há um estudo crescente sobre técnicas de reconhecimento/classificação e

representação de objetos 3D. Estes estudos, motivados pela popularização de equipamentos/sensores para captura de dados 3D, mostram aplicações em diversas áreas de estudo, como por exemplo as áreas da saúde, segurança e industrial. Os trabalhos analisados mostram também estudos comparativos entre métodos de classificação/reconhecimento e de formas de representação de objetos. Este relatório procurou condensar os estudos publicados entre 2006 e 2016 presentes em 3 bases de dados de trabalhos científicos, com o propósito de entendimento sobre os métodos e técnicas para reconhecimento/classificação de objetos 3D.

A. *Perspectivas para a área*

Da análise e condensação dos trabalhos selecionados publicados entre 2006 e 2016 pode-se estimar algumas perspectivas para a área:

- Grande parte dos métodos utilizam representações com base em características confirmando a tendência de continuar com este tipo de representação, incluindo mais tipos de características nos descritores;
- Alguns trabalhos realizam a combinação de vários descritores, fornecendo o indicativo que a combinação seria uma possível solução para o aumento da discretização da representação dos objetos;
- Um aumento significativo na quantidade de bases de dados com modelos 3D disponíveis;
- Uma exploração cada vez maior de técnicas para aceleração dos cálculos e redução da dimensionalidade visando a aplicação em problemas que exigem solução em tempo real;
- Um uso cada vez maior de técnicas combinadas e de modelos de aprendizagem como o *deep learning*;
- Variações com os descritores já existentes utilizando diferentes métricas de similaridade, combinação e de comparação dos descritores.

TABLE I
TABELA COM TODOS OS TRABALHOS ANALISADOS.

Autor	Título	Ano	Tipo de representação	Resumo	Tipo de imagem
Pichler	Towards more Agility in Robot Painting through 3D Object Recognition	2006	Características locais	Representação em forma de características locais e posterior correspondência entre modelo e cena	Imagens 2D e CAD
Pichler	RECOGNITION AND 6DOF LOCALISATION OF PARTS FOR LOTSIZE1 AUTOMATION	2007	Características locais	representação em forma de características locais e posterior correspondência entre modelo e cena	Imagens 2D e CAD
Wei-Yang Lin	ROBUST GEOMETRICALLY INVARIANT FEATURES FOR 2D SHAPE MATCHING AND 3D FACE RECOGNITION	2006	Características locais	Aprendizagem de modelo descrito em termos de características locais, rastreamento no tempo, montagem de modelo para treinamento de um classificador e posterior classificação de novos objetos	Imagens de profundidade
Arnaud	Trains of keypoints for 3D object recognition	2006	Características locais	Aprendizagem de modelo descrito em termos de características locais, rastreamento no tempo e montagem de modelo para treinamento de um classificador	Imagens 2D
Shivaswamy	Permutation Invariant SVMs	2006	Padrões	Aprendizagem de modelos com base em padrões e posterior classificação de outras imagens de entrada	Imagens 2D
Yi Tan	Exploiting Model Similarity for Indexing and Matching to a Large Model Database	2006	Características locais	Alinhamento dos objetos modelo e da cena e cálculo de similaridade	Nuvem de pontos
Sukhan Lee	3D Object Recognition using Multiple Features for Robotic Manipulation	2006	Características geométricas e de aparência	Correspondência de objetos em uma sequência de cenas utilizando características locais	Nuvem de pontos
Stasse	3D object recognition using spin-images for a humanoid estéreoscopic vision system	2006	Características locais	Descrição utilizando Spin images com multi-resolução e posterior algoritmo de busca do modelo descrito	Nuvem de pontos
Chen Zhe	Geometric Hashing Using 3D Aspects and Constrained Structures	2006	Características locais - aspectos 3D e estruturas geométricas 3D restritas	Extração de características do modelo e da imagem 2D e posterior correspondência através de um esquema de votação nos modelos	Modelos 3D
Raptis	Bayesian Fusion of Contour Descriptions: Application to 3-D Object and Face Recognition	2006	Características de formato - contorno	Extração de características de contorno, montando uma caricatura 2D do objeto/face com base na fusão de contornos de diferentes visões e posterior cálculo de probabilidade para estimativa da face/objeto	Imagens 2D
Delponte	Analysis on a Local Approach to 3D Object Recognition	2006	Características locais	Descrição de objeto com base em características locais e posterior treinamento e classificação com um SVM	Imagens 2D
Weiwei Xing	3D Object Classification by Part Features Fusion	2006	Características geométricas e topológicas	Descrição de uma biblioteca de modelos e posterior correspondência entre modelos da biblioteca e objetos a serem classificados, com base em uma função de similaridade	Modelo 3D
Weiwei Xing	VOLUMETRIC PART BASED 3D OBJECT CLASSIFICATION	2006	Características geométricas e topológicas	Descrição de uma biblioteca de modelos e posterior correspondência entre modelos da biblioteca e objetos a serem classificados, com base em uma função de similaridade	Modelo 3D
Kushal	Modeling 3D Objects from stereo Views and Recognizing Them in Photographs	2006	Características locais	Montagem de modelos partindo de várias imagens de pares estéreos com posterior correspondência entre modelo e imagem de teste	Imagens 2D
Arana-Daniel	MIMO SVMs for 3D object classification	2006	Características do ponto e a distância do mesmo até o eixo principal do objeto	Extração de característica e treinamento do classificador CSVM para posterior classificação de novos dados	Nuvem de pontos
Ramalingam	Curvature-Based Fuzzy Surface Classification	2006	Fuzzy moments	Segmentação utilizando um sistema fuzzy proposto, extração de características globais com base em moments fuzzy e posterior reconhecimento dos objetos com base nestas características globais	Imagens de profundidade
Sunggho Kim	Scalable Representation and Learning for 3D Object Recognition Using Shared Feature-Based View Clustering	2006	Híbrido - visão e característica	Descreve em termos da forma de representação utilizada e depois compara a cena aos objetos modelados	Imagens 2D
Assfalg	3DString: A Feature String Kernel for 3D Object Classification on Voxelized Data	2006	Características locais	Descrição em termos de string de característica e classificação com base no valor de similaridade, utilizando o Nearest Neighbor	Modelos 3D

Leitner	Learning 3D Object Recognition from an Unlabelled and Unordered Training Set	2007	Características locais	Extração de características e posterior aprendizagem das mesmas, seguidas de correspondência entre as características do modelo e da cena	Imagens 2D
Kootstra	Exploring objects for recognition in the real world	2007	Características SIFT	Extraí-se características SIFT dos objetos na base de dados e dos objetos observados na cena. Utiliza-se um método de clusterização para reduzir o número de pontos armazenados na base de dados e por fim utiliza-se um modelo probabilístico para, com base em evidências provenientes da correspondência das características SIFT de diferentes pontos de vista, determinar o objeto mais ativo da base em relação ao objeto observado e, conseqüentemente reconhecer o objeto	Imagens 2D
Okada	Multi-cue 3D Object Recognition in Knowledge-based Vision-guided Humanoid Robot System	2007	Extração de características visuais - histograma de cor, bordas retas e pontos de característica 3D	Extração de características de um objeto de referência e de uma imagem de entrada, cálculo da probabilidade de ocorrência destas características e posterior integração destas probabilidades em um filtro de partículas	Imagens 2D
Delponte	Appearance-based 3D object recognition with time-invariant features	2007	Características invariantes ao tempo	Extração de características invariantes ao tempo das sequências de treinamento e teste e posterior correspondência entre as características invariantes ao tempo	Vídeos
Jun-Hai Zhai	VIEW-BASED 3D OBJECT RECOGNITION USING WAVELET MULTISCALE SINGULAR-VALUE DECOMPOSITION AND SUPPORT VECTOR MACHINE	2007	Vetores de características	Extração de vetores de características de diferentes visões e treinamento de um classificador SVM com esses vetores para posterior classificação de novos objetos	Imagens 2D
Salgian	Using Multiple Patches for 3D Object Recognition	2007	Descritores de característica local	Avaliação da performance de três descritores, através da extração dos mesmos, de imagens na base de dados e da cena e posterior correspondência visando o reconhecimento de objetos 3D	Imagens 2D
Xinju li	3D object recognition from range images using pyramid matching	2007	Descritores de assinatura local	Extração de descritores de assinatura local, construção de um conjunto de descritores e cálculo de similaridade utilizando uma função de correspondência pyramid match kernel	Imagens de profundidade
Assfalg	Content-Based Retrieval of 3-D Objects Using Spin Image Signatures	2007	Características locais - spin images signatures	Extração do descritor proposto para posterior comparação e recuperação do objeto em uma base de dados	Modelos 3D
Aouada	STATISTICAL ANALYSIS OF THE GLOBAL GEODESIC FUNCTION FOR 3D OBJECT CLASSIFICATION	2007	Características de resolução	Definição de características de resolução para classe, com base no máximo das características de resolução dos objetos de treinamento, e para os objetos a serem classificados de forma a medir a similaridade entre as características de resolução da classe e do objeto para realizar a classificação	Modelos 3D
Lee-Yeng Ong	An Approach to 3-D Object Recognition Using Legendre Moment Invariants	2007	Características locais geométricas	Geração de invariantes 3D com base em momentos geométricos e verificação da performance dos mesmos	Modelos 3D
Hui Chen	3D free-form object recognition in range images using local surface patches	2007	Características locais	Descrição de superfícies e correspondência de superfícies com base em modelo descrito	Imagens de profundidade
Weiwei Xing	3D Object Classification System based on Volumetric Parts	2007	Características geométricas e topológicas	Descrição de uma biblioteca de modelos e posterior correspondência entre modelos da biblioteca e objetos a serem classificados, com base em uma função de similaridade	Modelos 3D
Weiwei Xing	An Integrated System for 3D Object Reconstruction and Recognition	2007	Características geométricas e topológicas	Descrição de uma biblioteca de modelos e posterior correspondência entre modelos da biblioteca e objetos a serem classificados, com base em uma função de similaridade	Modelos 3D
Kordelas	RECOGNIZING 3D OBJECTS USING RAY-TRIANGLE INTERSECTION DISTANCES	2007	Vetor de característica	Extração do vetor de características dos modelos de objetos, presentes na biblioteca e da cena para posterior correspondência e reconhecimento dos objetos presentes na cena	Modelos 3D

Sungho Kim	Robust model-based scene interpretation by multilayered context information	2007	Características locais	Estrutura hierárquica associada a representação de características da imagem para aprendizagem e inferência, reconhecendo objetos com base em contextos de lugar e características da imagem	Imagens 2D
Tim K. Lee	3D Object Recognition by Eigen-Scale-Space of Contours	2007	Características de formato	Correspondência entre formatos entre a base de modelos e o objetos alvo	Imagens 2D
Kietzmann	Incremental GRLVQ: Learning relevant features for 3D object recognition	2008	Características locais	Aprendizagem e classificação com base no treinamento realizado	Imagens 2D
Xu Sheng	3D Object Recognition Using Multiple Features and Neural Network	2008	Características locais	Treinamento com características extraídas de imagens 2D e posterior classificação dos objetos com o classificador	Imagens 2D
Yu Zhong	Intrinsic Shape Signatures: A Shape Descriptor for 3D Object Recognition	2008	Intrinsic Shape Signatures - descritor de formato local	Extração do descritor proposto para as nuvens de pontos de referência e de consulta e posterior correspondência entre descritores, através do cálculo de similaridade e estimativa da pose	Nuvem de pontos
Sungho Kim	Scalable representation for 3D object recognition using feature sharing and view clustering	2008	Híbrido - visão e característica	Descreve em termos da forma de representação utilizada e depois compara a cena aos objetos modelados	Imagens 2D
Huy Tho Ho	Multi-scale Feature Extraction for 3D Models using Local Surface Curvature	2008	Características locais	Extração de características locais salientes dos modelos 3D e avaliação destas características	Modelos 3D
Xiaolan Li	3D Part Identification Based on Local Shape Descriptors	2008	Características locais - Dicionário de palavras visuais	Construção de um dicionário de palavras visuais para cada modelo na base de dados e comparação do objeto alvo com os dicionários extraídos dos modelos em termos de uma métrica de similaridade	Nuvem de pontos
Salgian	Combining Local Descriptors for 3D Object Recognition and Categorization	2008	Descritores de característica local	Extração e combinação de descritores de características utilizando-os no reconhecimento e na classificação de objetos	Imagens 2D
Donghui Wang	3D Object Recognition by Fast Spherical Correlation between Combined View EGIs and PFT	2008	Características phase encoded Fourier transform e esfera gaussiana	Representação da imagem de profundidade de entrada por meio de características PFT e posterior correlacionamento com a esfera gaussiana, montada com base em diferentes visões do objeto para cada objeto na base de dados, para computação da correspondência e determinação do reconhecimento e da pose do objeto	Imagens de profundidade
Mavrinac	Fuzzy Associative Approach for Recognition of 3D Objects in Arbitrary Pose	2008	Extração de características do mapa de disparidade - Compactness, First Hu Moment e histograma da distribuição geral da intensidade na imagem	Extração de características do mapa de disparidade e utilização destas para o treinamento e o reconhecimento via um mapa associativo fuzzy	Mapa de disparidade
Atmosukarto	A Learning Approach to 3D Object Representation for Classification	2008	Características globais	Aprendizagem de pontos de saliência, transformação em mapas de assinatura e aprendizagem e classificação com base nos mapas	Modelos 3D
Elizabeth Gonzáles	Active object recognition based on Fourier descriptors clustering	2008	Características de formato - Silhuetas descritas com descritores de fourier	Monta-se uma base com silhuetas descritas por descritores de Fourier em diversos pontos de visão do objeto, compara-se o novo objeto com esta base seguindo o processo de descrição e posterior resolução de ambiguidade quando necessário	Modelos 3D
Li-Li Wan	3D Object Classification by Fuzzy KNN and Bayesian Decision	2009	Características visuais e estruturais	Aprendizagem de características visuais e estruturais e posterior correspondência com base em classificador Fuzzy KNN e Regras Bayesianas	Imagens 2D
Hui chen	Efficient Recognition of Highly Similar 3D Objects in Range Images	2009	Descritores de característica local	Descreve modelo e cena com base em descritores de características, treina um classificador com os modelos, realiza uma correspondência entre modelo e cena e verifica as hipóteses de correspondência com o ICP	Imagens de profundidade

Hyunjun Kim	Environment Adaptive 3D Object Recognition and Pose Estimation by Cognitive Perception Engine	2009	Características locais - SIFT, linha e cor	Seleção e extração de características para posterior correspondência e estimativa de pose dos objetos	Imagens RGB-D
Gee-Sern Hsu	Real-time 3-D Object Recognition Using Scale Invariant Feature Transform and estéreo Vision	2009	Características SIFT - características locais	Descrição do objeto na base de dados através de características SIFT e posterior correspondência e utilização dessas informações, associadas as informações de profundidade e pose do objeto, para tocar com precisão os objetos alvo	Imagens 2D
Osman	3D Object Recognition Using MANFIS Network with Orthogonal and Non-orthogonal Moments	2009	Características Hu's e Zernike's moments	Extração características por meio dos Hu's e Zernike's moments, utilização desses para treinamento da rede neuro-fuzzy e posterior classificação dos objetos com a rede treinada	Imagens 2D
Himmelsbach	Real-time Object Classification in 3D Point Clouds Using Point Feature Histograms	2009	Histograma de características locais	Segmentação com base em um grid de ocupação seguida de extração de características e posterior classificação utilizando um classificador previamente treinado com dados de treinamento rotulados manualmente	Nuvem de pontos
Gibbins	A comparison of terrain classification using local feature measurements of 3-dimensional colour point-cloud data	2009	Extração de características locais - Zernike moments, curvatura, spin images e cor	Extração de características locais, seguido de treinamento de um classificador para posterior classificação das estruturas de terrenos	Nuvem de pontos
Noceti	Spatio-temporal constraints for on-line 3D object recognition in videos	2009	Características locais invariantes ao tempo	Descreve os modelos dos objetos em vídeos em termos das características locais invariantes a tempo e espaço e faz uma correspondência entre os objetos modelo e as cenas	Vídeos
Atmosukarto	3D object classification using salient point patterns with application to craniofacial research	2010	Características globais	Aprendizagem de pontos de saliência, transformação em mapas de assinatura e aprendizagem e classificação com base nos mapas	Modelos 3D
Hsiao	Making specific features less discriminative to improve point-based 3D object recognition	2010	Características locais	Extração e correspondência de descritores de características quantizadas do modelo e da cena mantendo ambiguidade na correspondência	Imagens 2D
Chih-Hong Kao	Study of Feature-Based Image Capturing and Recognition Algorithm	2010	Extração de características geométricas, de formato do objeto e descritores de fourier	Extração das características utilizadas para a representação do objeto e posterior comparação em duas etapas com as imagens na base de dados	Imagens 2D
Drost	Model Globally, Match Locally: Efficient and Robust 3D Object Recognition	2010	Características de pares de pontos	Extração de características de pares de pontos do modelo e da cena e posterior correspondência com base em um esquema de votação para estimativa final da pose	Nuvem de pontos
Yangye Wang	Research On 3D Object Recognition from Wire-frame Based on Edge Moment	2010	Vetor de característica de edge moments	Extração de características e posterior treinamento e reconhecimento apresentando estas características a uma rede neural wavelet	Modelo 3D
Naikal	Towards an efficient distributed object recognition system in wireless smart camera networks	2010	Histograma de características de entropia de múltiplas visões	Extração do histograma de características de uma ou múltiplas câmeras e comparação com um conjunto de histogramas de treinamento que representam o objeto	Imagens 2D
Rusu	Fast 3D Recognition and Pose Using the Viewpoint Feature Histogram	2010	Descritor de Característica - Viewpoint Feature Histogram	Extração do descritor proposto, treinamento e posterior classificação com base nesse descritor	Nuvem de pontos
Mian	On the Repeatability and Quality of Keypoints for Local Feature-based 3D Object Retrieval from Cluttered Scenes	2010	Características locais	Extração e correspondência entre características do modelo e da cena	Modelo 3D
Tombari	Unique Shape Context for 3D Data Description	2010	Descritor com base em características locais	Descrição do modelo por meio de um descritor e posterior correspondência entre cena e modelo	Modelo 3D
Knoop	Orientation Invariant 3D Object Classification using Hough Transform Based Methods	2010	Descritores invariantes a rotação - características locais	Aprendizagem de classes de modelos com base nos formatos de treinamento e depois classificação com base nestas classes aprendidas	Imagens 2D
Seatovic	Automatic Weed Detection and Treatment in Grasslands	2010	Descritores elípticos de Fourier	Extração de características, treinamento de um classificador e posterior classificação com base no classificador treinado	Nuvem de pontos

Jidan Zhou	Histograms of Categorized Shapes for 3D Ear Detection	2010	Características locais - Histograms of Categorized Shape	Extração de características locais para treinamento de um classificador binário e posterior classificação da presença ou não do objeto na imagem	Imagens de profundidade
Wohlkinger	3D Object Classification for Mobile Robots in Home-Environments Using	2010	Descritor de característica - Spherical Harmonics	Extração do descritor e posterior correspondência utilizando a procura dos vizinhos mais próximos na base de dados	Nuvem de pontos
Tombari	Object recognition in 3D scenes with occlusions and clutter by Hough voting	2010	Descritores de característica local	Descrição das características de cena e modelo e posterior correspondência entre elas através de um esquema de voto	Imagens RGB-D
Owechko	Automatic recognition of diverse 3-D objects and analysis of large urban scenes using ground and aerial LIDAR sensors	2010	Características - propriedades geométricas e topológicas	Extração de características geométricas e topológicas e posterior treinamento e classificação utilizando classificadores estatísticos	Nuvem de pontos
Unel	3D object recognition using invariants of 2D projection curves	2010	Invariantes geométricas e algébricas - características locais	Representação com base em vetores de invariantes e comparação dos modelos com os objetos nas imagens em termos de métrica de similaridade	Nuvem de pontos ou Malha de pontos
Westell	3D Object Recognition via Multi-View Inspection in Unknown Environments	2010	Matriz compactando histograma 2D normalizado com características de cor	Extração de características de cor de diferentes pontos de vista e correspondência com imagens armazenadas na base de dados	Imagens 2D e estéreo
Ayoub	3-D Object Recognition Based on SVM and estéreo-Vision : Application in Endoscopic imaging	2010	Dois formas separadamente: características estatísticas e características locais dos pontos	Extração de características, treinamento de um classificador e posterior reconhecimento das imagens de entrada	Nuvem de pontos
Kanezaki	High-speed 3D Object Recognition Using Additive Features in A Linear Subspace	2010	Vetor de característica - Color-CHLAC	Extração das características Color-CHLAC dos objetos na base de dados, montagem de um subespaço com base no vetor, extração de um vetor de características Color-CHLAC de partes da cena e subsequente consulta da correspondência entre os vetores das partes da cena e o subespaço dos objetos armazenados na base de dados	Modelos 3D
Usui	3D Object Recognition Based on Confidence LUT of SIFT Feature Distance	2010	Características locais - ASIFT	Extração e correspondência de características do modelo e da cena, onde a parte de correspondência é calculada com base em valores de confiança	Imagens 2D
Rodner	Multiple Kernel Gaussian Process Classification for Generic 3D Object Recognition	2010	Características de cor e profundidade	Extração de características e posterior treinamento de um processo gaussiano para classificação	Imagens de cor e profundidade
Papazov	An Efficient RANSAC for 3D Object Recognition in Noisy and Occluded Scenes	2010	Descritores geométricos - características locais	Cálculo de descritores da cena e do modelo e posterior correspondência e confirmação de hipóteses no alinhamento do modelo com a cena	Nuvem de pontos
Bariya	Scale-Hierarchical 3D Object Recognition in Cluttered Scenes	2010	Características locais geométricas	Extração e correspondência entre características geométrica do modelo e da cena, gerando hipóteses as quais são avaliadas, escolhendo apenas a hipótese com maior valor de sobreposição	Imagem de profundidade
Akagunduz	3D object recognition from range images using transform invariant object representation	2010	Descritores de características	Extração de descritores e comparação com um esquema de hashing entre vetores de características de modelo e imagem de teste	Imagens de profundidade
Tangruamsub	3D Object Recognition Using a Voting Algorithm in a Real-world Environment	2011	Características locais e espaciais	Aprendizagem e correspondência de modelo com os objetos da imagem, utilizando um técnica de votação para reconhecer o objeto e posterior estimativa de pose	Imagens 2D
Eunyoung Kim	3D Object Recognition in Range Images Using Visibility Context	2011	Características de pares de pontos	Extração de características de pares de pontos de objetos a serem consultados na cena e posterior correspondência entre essas características através de um esquema de votação	Imagens de profundidade
Wohlkinger	Ensemble of shape functions for 3D object classification	2011	Características de nuvens de pontos - descritor Ensemble of Shape Functions (ESF)	Descrição com ESF dos modelos CAD e de nuvens de pontos e posterior correspondência entre os descritores com base em uma métrica de distância	Imagens RGB-D e Modelos CAD
Zarpalas	RECOGNIZING 3D OBJECTS IN CLUTTERED SCENES USING PROJECTION IMAGES	2011	Descritor de características - projection images	Extração do descritor proposto para os modelos na base e a cena e posterior correspondência entre eles através de uma técnica de clusterização	Nuvem de pontos

Wohlkinger	Shape Distributions on Voxel Surfaces for 3D Object Classification From Depth Images	2011	Descritor de formato - adaptação do D2 shape distribution descriptor	Cálculo e correspondência de descritores do modelo e da cena	Nuvem de pontos e modelo CAD
Jindan Zhou	A Computationally Efficient Approach to 3D Ear Recognition Employing Local and Holistic Features	2011	Características locais e holísticas	Extração de características locais dos modelos da galeria e de teste, posterior correspondência entre os descritores dessas características e fusão dos valores de correspondência para obtenção de um valor final de correspondência	Imagens de profundidade
Decker	Dense Statistic Versus Sparse Feature Based Approach for 3D Object Recognition	2011	Características locais	Extração de características, treinamento e classificação de imagens de objetos vistos de vários ângulos	Imagens 2D
Ryo Hanai	Electric appliance parts classification using a measure combining the whole shape and local shape distribution similarities	2011	Descritores de características	Análise de 3 método de extração de características para classificação de uma base de partes elétricas	Malhas 3D
Sukhan Lee	Dependable 3D Object Recognition with Two-layered Particle Filter	2011	Características locais	Estimativas de pose dos objetos com base em probabilidades	Imagens estéreo
Eunyoung Kim	Scalable Object Classification in Range Images	2011	Características local - Visual word	Aprendizagem de modelo e classificação utilizando uma métrica de similaridade e uma estrutura em árvore	Imagens de profundidade
Akbar	Training neural networks using Clonal Selection Algorithm and Particle Swarm Optimization: A Comparisons for 3D Object Recognition	2011	Descritor de características -Hu moments	Extração de características, posterior treinamento de um a rede neural, com duas abordagens de inicialização, e posterior classificação dos objetos com a rede treinada	Imagens 2D
Albarelli	A Non-Cooperative Game for 3D Object Recognition in Cluttered Scenes	2011	Características locais	Correspondência entre modelo e cena com base nos descritores, respeitando uma consistência global	Modelos 3D
Woongji Jeong	Statistical Feature Selection Model for Robust 3D Object Recognition	2011	Características de cor e formato - SIFT, linha e cor	Modelagem probabilística de cada característica em diferentes ambientes para a seleção ótima de características, visando o reconhecimento de objetos e estimativa de pose	Imagens 2D
Aldoma	A Global Hypotheses Verification Method for 3D Object Recognition	2012	Características locais	Utiliza um descritor de características locais para descrever modelo e cena para posterior correspondência e refinamento da hipótese de correspondência	Nuvem de pontos
Papazov	Rigid 3D geometry matching for grasping of known objects in cluttered scenes	2012	Descritores geométricos - características locais	Cálculo de descritores da cena e do modelo e posterior correspondência e confirmação de hipóteses no alinhamento do modelo com a cena	Nuvem de pontos
Petricek	Area-weighted Surface Normals for 3D Object Recognition	2012	Características locais - histograma das características	Extração dos descritores do modelo e das características da cena para posterior correspondência entre elas	Malhas poligonais
Yabushita	A framework of three-dimensional object recognition which needs only a few reference images	2012	Descritor de característica local - SURF	Extração de características da imagem de referência e da cena, posterior correspondência dos pontos e reconhecimento com base em um esquema de votação	Imagens 2D
Maeda	Surface Matching by Curvature Distribution Images Generated via Gaze Modeling	2012	Descritor com base na curvatura da superfície	Extração de descritores com base na curvatura da superfície, tanto do modelo quanto da cena, e posterior comparação dos descritores	Imagem de profundidade
Sheta	3D Object Recognition Using Fuzzy Mathematical Modeling of 2D Images	2012	Características de moments- Hu's, Zernike's e affine moments	Extração de características de <i>moments</i> e posterior classificação utilizando um modelo matemático Fuzzy	Imagens 2D
Bongale	Implementation of 3D Object Recognition and Tracking	2012	Descritor de características locais -VFH	Extração e posterior correspondência do descritor VHF dos objetos da base de dados e dos objetos de teste	Nuvem de pontos
Bariya	3D Geometric Scale Variability in Range Images: Features and Descriptors	2012	Características locais geométricas	Extração e correspondência entre características geométrica do modelo e da cena, gerando hipóteses as quais são avaliadas, escolhendo apenas a hipótese com maior valor de sobreposição	Imagem de profundidade
Socher	Convolutional-Recursive Deep Learning for 3D Object Classification	2012	Vetor de características	Aprende um vetor de características, treina um classificador e utiliza o classificador treinado para a parte de reconhecimento de objetos	Imagens RGB-D

Hao Ding	A new generalized affine moment invariants for shape retrieval and object recognition	2012	Características de momentos-illumination invariant MSA moments	Extração das características de momentos e utilização destes para classificação por meio de um classificador por nearest neighbor	Imagens 2D
Youngkyoon Jang	Local feature descriptors for 3D object recognition in ubiquitous virtual reality	2012	Características locais	Descreve o objeto em termos de características locais e treina um classificador para classificar os objetos em imagens de entrada	Imagens 2D
Yong-Jin Liu	2D-Line-Drawing-Based 3D Object Recognition	2012	Características locais - histograma circular de características	Descreve o modelo em termos da representação proposta, monta-se um livro de códigos e faz a correspondência com uma métrica de similaridade sob os livros de códigos	Imagens 2D
Wang ping	A Recognition Approach of 3-D Objects based on the Tsallis Entropy	2012	Descritor global de imagem - TESI	Extração e posterior cálculo de correlação entre o vetor descritor do modelo e o vetor descritor da imagem	Imagens 2D
Bencharef	Color Objects Recognition System based on Artificial Neural Network with Zernike, Hu & Geodesic Descriptors	2012	Combinação de descritores de cor e formato	Extração de descritores do objeto para treinamento de uma rede neural e posterior reconhecimento	Imagens 2D
Liang-Chia Chen	Automatic object detection employing viewing angle histogram for range images	2012	Combinação de características geométricas e histogramas de ângulo de visão	Extração de características de modelos e dos objetos na nuvem de pontos para posterior cálculo de correspondência	Nuvem de pontos
Megherbi	A COMPARISON OF CLASSIFICATION APPROACHES FOR THREAT DETECTION IN CT BASED BAGGAGE SCREENING	2012	Descritor de características - Zernike e/ou HSI	Extração dos descritores, treinamento de um classificador e posterior classificação de novas imagens	Imagens tomográficas
Shaiek	Fast 3D keypoints detector and descriptor for view-based 3D Objects Recognition	2012	Descritor de característica local	Correspondência entre descritor modelo e descritor do objeto de teste segundo uma métrica de similaridade	Nuvem de pontos
Rodola	A Scale Independent Selection Process for 3D Object Recognition in Cluttered Scenes	2013	Características locais	Correspondência entre modelo e cena com base nos descritores, respeitando uma consistência global	Modelos 3D
Keaikitse	Comparison of active SIFT-based 3D object recognition algorithms	2013	Características locais - SIFT	Comparação de dois sistemas de reconhecimento de objetos de forma ativa	Imagens 2D
Oleari	A Low-Cost stereo System for 3D Object Recognition	2013	Descritor de características de ponto - Fast Point Feature Histogram	Cálculo da correspondência entre a nuvem de pontos do objeto de teste e as nuvens de pontos de objetos conhecidos por meio do alinhamento entre elas e posterior cálculo de aptidão	Nuvem de pontos
Bennamoun	3D Free Form Object Recognition using Rotational Projection Statistics	2013	Características locais	Extração e correspondência de características locais da superfície de modelo e da cena, seguidos de um esquema de votação e verificação das hipóteses	Malha de pontos
Yabushita	Mobile visual search for 3-D objects: Matching user-captured video to single reference image	2013	Descritores de pontos chave	Extração e correspondência de pontos chave entre as imagens armazenadas na base de dados e o vídeo de teste	Vídeo
Byung-soo Kim	Accurate Localization of 3D Objects from RGB-D Data using Segmentation Hypotheses	2013	Características locais de aparência	Correlacionamento entre partes do objeto na imagem 2D e pontos 3D para localização de objetos	Imagens RGB-D
Palossi	GPU-SHOT: parallel optimization for real-time 3D local description	2013	Descritor de características - SHOT	Proposta de otimização do descritor de características SHOT e uso do mesmo em um pipeline de reconhecimento de objetos	Imagens RGB-D
Abdelrahman	Heat Front Propagation Contours for 3D Face Recognition	2013	Características locais com base na propagação do calor	Extração de características locais e representação com base em contornos para posterior correspondência entre faces na base de dados e faces inqueridas	Imagens de profundidade
Jincheng Yu	A Vision-based Robotic Grasping System Using Deep Learning for 3D Object Recognition and Pose Estimation	2013	Características locais	Identificação do objeto, extração de características, treinamento de um modelo de deep learning e posterior reconhecimento do objeto e estimativa da pose	Imagens 2D
Joseph Lam	3D Object Recognition By Surface Registration of Interest Segments	2013	Características dos segmentos	Registro de segmentos do modelo e da cena procurando a melhor correspondência entre os mesmos	Nuvem de pontos

Ekekrantz	Adaptive Iterative Closest Keypoint	2013	Detetores de pontos chave e descritores de característica	Método de registro de dados RGB-D com base em pontos chave e descritores	Imagens RGB-D
Gomes	Efficient 3D object recognition using foveated point clouds	2013	Descritores de característica	Descrição e correspondência de descritores da cena e do modelo	Nuvem de pontos
Yulan Guo	Integrating Shape and Color Cues for Textured 3D Object Recognition	2013	Descritores de características de cor e espacial	Extração de descritores de características para cena e modelo, posterior correspondências entre eles para geração de hipóteses, verificação das hipóteses e segmentação dos objetos na cena	Modelos 3D
Shuang Ma	3D Object Recognition Using Kernel PCA Based on Depth Information for Twist-lock Grasping	2013	Kernel principal features - características	Identificação da posição do objeto, extração de características e obtenção de uma matriz kernel, por meio de treinamento, para posterior classificação do objeto com base em um coeficiente de correlação	Imagens RGB-D
Guo Jing	3D Object Classification Based on Local Keywords and Hidden Markov Model	2013	Característica espacial do modelo - local keyword	Extração de local keyword, treinamento de uma Hidden Markov Model para cada classe e posterior classificação dos objetos de entrada	Modelos 3D
Moreno	SLAM++: Simultaneous Localisation and Mapping at the Level of Objects	2013	Características de ponto - Point-Pair Features	Construção de uma base de modelos de objetos e correspondência por meio de características de pontos de pares, um esquema de votação e um mapeamento do ambiente	Imagens de profundidade
Soysal	RECOGNITION OF 3D OBJECTS FROM UNCONSTRAINED 2D IMAGES BY USING LOCAL APPEARANCE AND AFFINE GEOMETRY	2013	Características locais	Construção de uma biblioteca de modelos com base em características locais e invariantes geométricas e posterior correspondência das imagens de entrada com a biblioteca de modelos	Imagens 2D
Syed Shah	A Novel Local Surface Description for Automatic 3D Object Recognition in Low Resolution Cluttered Scenes	2013	Descritor de característica locais	Montagem de uma base de modelos descrita na forma do descritor proposto seguido do cálculo de correspondência entre modelos e cena e posterior verificação desta correspondência	Modelos 3D
Choi	Angular-partitioned spin image descriptor for robust 3D facial landmark detection	2013	Descritor de formato local	Comparação entre os pontos de referência extraídos com os pontos anotados manualmente de ambos os descritores, com o intuito de medir o erro médio alcançado na localização destes pontos de referência	Dados 3D de faces
Guan Pang	Training-based Object Recognition in Cluttered 3D Point Clouds	2013	Características locais	Treinamento de classificadores com base em características dos modelos de uma biblioteca e rastreamento na nuvem de pontos procurando o objeto alvo	Nuvem de pontos
Nakashika	3D-Object Recognition Based on LLC Using Depth Spatial Pyramid	2014	Características locais	Extração e treinamento com base em características locais e posterior reconhecimento da categoria dos objetos teste	Imagens RGB-D
Takei	3D Object Recognition using Effective Features Selected by Evaluating Performance of Discrimination	2014	Características locais - Vector pairs	Seleção de características com base na avaliação da performance de discriminação em um espaço de característica para posterior correspondência entre modelo e cena	Nuvem de pontos
Mounton	3D OBJECT CLASSIFICATION IN BAGGAGE COMPUTED TOMOGRAPHY IMAGERY USING RANDOMISED CLUSTERING FORESTS	2014	Características esparsa e densas	Extração de características, construção de um codebook e posterior treinamento de um classificador para a identificação das classes dos objetos	Imagens 2D
Xiangfei Qian	3D Object Recognition by Geometric Context and Gaussian-Mixture-Model-Based Plane Classification	2014	Características entre planos	Extração de características entre planos, treinamento de um classificador e posterior classificação de novos objetos	Imagens de profundidade
Mateo	Performance Evaluation of Surface Normals-based Descriptors for Recognition of Objects Using CAD-Models	2014	Descritores de superfície - características geométricas	Avaliação de descritores de superfície	Nuvem de pontos/modelos CAD
Shang-Hung Lee	Moment-Preserving Techniques for 3D Shape Registration and Recognition	2014	Características de formato local	Extração de descritores de características, para posterior treinamento e classificação de ações 3D do corpo humano	Imagens 2D
Yulan Guo	An Accurate and Robust Range Image Registration Algorithm for 3D Object Modeling	2014	Descritor de características - Rops	Registro de imagens de profundidade de múltiplas visões para a construção de modelo 3D	Imagens de profundidade
Salih	COMPRESSED VFH DESCRIPTOR FOR 3D OBJECT CLASSIFICATION	2014	Descritor de característica - compressão do VFH	Extração do descritor proposto e posterior treinamento de um SVM para classificação de objetos 3D	Nuvem de pontos

Xun Yu	Face Recognition using 3D Directional Corner Points	2014	Descritor de formato 3D - 3D DCPs	Descrição das faces na base de dados e a face a ser correspondida com o descritor proposto e posterior correspondência entre descritores	Modelos 3D de face
Ejima	3D Object Recognition based on the Reference Point Ensemble	2014	Características de formato - Labeled-Surflet-Pair	Geração de um conjunto de pontos de referência definidos pelo formato do objeto e posterior uso desse conjunto de pares de referência no reconhecimento e definição da pose do objeto	Imagens 2D
Dong Liang	A 3D Object Recognition and Pose estimation System Using Deep Learning Method	2014	Características locais	Treinamento de um modelo de deep learning, com base em um conjunto de padrões, e posterior reconhecimento e estimativa de pose do objeto inquerido	Imagens 2D
Carrer	Concealed Weapon Detection Using UWB 3-D Radar Imaging and Automatic Target Recognition	2014	Descritor local - SIFT	Extração do descritor local SIFT da biblioteca de objetos e dos segmentos da cena e posterior correspondência entre os descritores	Imagens de micro-ondas
Geetha	An Improved Content Based Image Retrieval in RGBD Images using Point Clouds	2014	Característica de curvaturas principais	Extração de características dos objetos na base de dados e do objeto inquerido e posterior correspondência entre características utilizando uma métrica de distância	Imagens RGBD
Kasaei	An Interactive Open-Ended Learning Approach for 3D Object Recognition	2014	Características locais	Descrição dos objetos detectados em termos de características e inserção na base de dados para posterior classificação de novos objetos, sendo que cada inserção de um novo objeto a classe faz com que as classes sofrem aprimoramentos ou, se não existir a classe, criam-se novas classes	Nuvem de pontos
Mhamdi	A local approach for 3D object recognition through a set of size functions	2014	Características locais com base em funções	Descreve-se o objeto 3D em um conjunto de 18 funções de medidas e compara-se os objetos 3D através da similaridade das 18 funções de medidas com as funções de medidas presentes em uma base de objetos descritos da mesma forma	Modelos 3D
Yulan Guo	A novel local surface feature for 3D object recognition under clutter and occlusion	2015	Características locais - Tri-Spin-Image	Descrição em termos de um descritor de características locais da superfície e comparação com base em modelo e cena, comparando as transformações entre modelo e cena em termos de alinhamento	Nuvem de pontos
Figueredo	Efficient pose estimation of rotationally symmetric objects	2015	Características locais - surflets	Descrição do modelo com base na representação por Surflets e posterior reconhecimento utilizando indexação em uma tabela hash	Nuvem de pontos
Yan Zhuang	Using Scale Coordination and Semantic Information for Robust 3-D Object Recognition by a Service Robot	2015	Características locais	Extração de características, treinamento de um classificador e posterior uso para o reconhecimento de múltiplos objetos, utilizando informação semântica extraída para remover falsas detecções	Nuvem de pontos
Sanguino	Improving 3D object detection and classification based on Kinect sensor and Hough transform	2015	Características de cor e formato	Extração das características de cor e formato da imagem de entrada e de referência para posterior classificação com base na pontuação de aptidão entre elas	Imagens RGBD
Ribeiro	Global Localization by Soft Object Recognition from 3D Partial Views	2015	Descritor de visões parciais - Partial View Heat Kernel	Extraí a representação para visões parciais e as compara com o descritor referência do objeto, integrando uma sequência de observações e leituras odométricas com o mapa do ambiente por meio de um filtro de partículas	Imagens 2D e de profundidade
Sánchez	SCurV: A 3D descriptor for object classification	2015	Descritor com base em características locais e globais	Extração do descritor para o treinamento e posterior classificação de objetos	Nuvem de pontos
Takei	SHORT: A Fast 3D Feature Description based on Estimating Occupancy in Spherical Shell Regions	2015	Descritor de características locais - SHORT	Uso do descritor proposto para descrever o objeto modelo seguido de um cálculo de correspondência para corresponder o objeto modelo e os dados de entrada	Nuvem de pontos
Rocha	Development of a 3D Model based Part Recognition System for Industrial Applications: Main Challenges	2015	Características de formato	Extração de características de formato do objeto para treinamento de um classificador e posterior classificação de novos objetos	Nuvem de pontos / Modelos CAD

Shimamura	The method based on view-directional consistency constraints for robust 3D object recognition	2015	Características locais	Extração de características locais, seguida de uma verificação na base de dados por possíveis candidatos e posterior verificação das correspondências encontradas com uma verificação geométrica com base em restrições de consistência	Imagens 2D
Fuhao Zou	Supervised feature learning via l2-norm regularized logistic regression for 3D object recognition	2015	Vetores de características	Treinamento com base em vetores de característica e teste do tipo de descritor utilizando o KNN	Imagens 2D
Flitton	Object classification in 3D baggage security computed tomography imagery using visual codebooks	2015	Com base em 4 descritores de característica 3D - Density Histogram(DH), Density Gradient Histogram(DGH), Scale Invariant Feature Transform(SIFT) e Rotation Invariant Feature Transform(RIFT)	Descreve os modelos em termos de descritores de características locais e treina uma SVM para classificar outros objetos nas duas classes utilizadas	Tomografia computadorizada
Qingtong Xu	The application of local Features in 3-dimensional object Recognition	2015	Características locais	Demonstração das três etapas desenvolvidas (estimativa da normal, seleção de pontos de característica, extração de descritores de característica), até o dado momento, de um sistema completo de reconhecimento de objetos	Nuvem de pontos
Soysal	Joint utilization of local appearance and geometric invariants for 3D object recognition	2015	Características locais	Construção de uma biblioteca de modelos com base em características locais e invariantes geométricas e posterior correspondência das imagens de entrada com a biblioteca de modelos	Imagens 2D
Beksi	Object Classification Using Dictionary Learning and RGB-D Covariance Descriptors	2015	Conjunto de descritores de covariância	Extração dos descritores de covariância com base em vetores de característica, montagem de dicionários com base nos descritores e iteração sobre os dicionários de uma categoria para classificar uma nova nuvem de pontos, em termos de categoria e instância	Nuvem de pontos
Filipe	BIK-BUS: Biologically Motivated 3D Keypoint Based on Bottom-Up Saliency	2015	Descritores de característica	Detecção de pontos chave, extração de descritores 3D com base nesses pontos chave, para objetos de entrada e armazenados em uma base de dados, e correspondência entre os descritores para determinar, com base na medida de distância, o objeto resultante	Nuvem de pontos
Garstka	Adaptive 3-D Object Classification with Reinforcement Learning	2015	Descritores de características locais	Extração do descritor, treinamento de um classificador para posterior classificação de objetos e acoplamento de um sistema de reforço de aprendizagem para selecionar a combinação que aprimore o resultado de classificação geral	Nuvem de pontos
Naguib	An Adaptive Evidence Structure for Bayesian Recognition of 3D Objects	2015	Características locais	Treinamento de uma rede Bayesiana adaptativa com base em características locais extraídas e posterior classificação das imagens de entrada	Imagens RGBD
Kasaei	Interactive Open-Ended Learning for 3D Object Recognition: An Approach and Experiments	2015	Características locais	Descrição dos objetos detectados em termos de características e inserção na base de dados para posterior classificação de novos objetos, sendo que cada inserção de um novo objeto a classe faz com que as classes sofrem aprimoramentos ou, se não existir a classe, criam-se novas classes	Nuvem de pontos
Jianhua Wang	Convolutional Neural Network for 3D Object Recognition Based on RGB-D Dataset	2015	Características de imagens RGB-D	Extração de características e treinamento de um classificador para classificação de objetos	Imagens RGB-D
Ren C. Luo	A Scalable Modular Architecture for Manufacturing Automation	2015	Descritores de características	Extração de descritores de características da cena e do modelo e posterior correspondência para classificação do objeto e estimativa da pose	Nuvem de pontos e modelo CAD
Ren C. Luo	Model-based 3D Object Recognition and Fetching by a 7-DoF Robot with Online Obstacle Avoidance for Factory Automation	2015	Descritores de características	Extração de descritores de características da cena e do modelo e posterior correspondência para classificação do objeto e estimativa da pose	Nuvem de pontos e modelo CAD

Yingjie Xia	Recognizing multi-view objects with occlusions using a deep architecture	2015	Vetor de características	Aprendizagem do vetor de características por deep learning e posterior uso dos parâmetros para treinamento de um classificador por random forest	Imagens de profundidade
Chaoqun Hong	Multi-view ensemble manifold regularization for 3D object recognition	2015	Vetor de característica	Extração de características de várias visões, treinamento do classificador com base nessas características e reconhecimento dos objetos utilizando o classificador treinado	Imagens 2D
Rangel	Using GNG on 3D Object Recognition in Noisy RGB-D data	2015	Descritor de característica - SHOT	Deteção de pontos chave, uso de um descritor de característica, correspondência entre a nuvem de pontos do modelo e da cena e posterior verificação das possíveis hipóteses	Nuvem de pontos
Syed Shah	A novel feature representation for automatic 3D object recognition in cluttered scenes	2016	Descritor de característica locais	Montagem de uma base de modelos descrita na forma do descritor proposto seguido do cálculo de correspondência entre modelos e cena e posterior verificação desta correspondência	Modelos 3D
Kechagias-Stamatis	Histogram of Distances for Local Surface Description	2016	Descritor local - HoD	Computação do descritor 3D proposto para características da cena e do modelo e posterior correspondência entre eles com base na distância euclidiana e no critério de taxa de distância do NN	Nuvem de pontos
Fan Zhu	Boosted Cross-Domain Dictionary Learning for Visual Categorization	2016	Características locais	Método para melhoria de classificadores fracos	Imagens 2D
Tateno	When 2.5D is not enough: Simultaneous Reconstruction, Segmentation and Recognition on dense SLAM	2016	Descritores 3D - VFH, CVFH e OUR-CVFH	Reconstrução, segmentação e posterior correspondência entre objetos na cena e objetos armazenados na base de dados de objetos	Modelo 3D e Imagens RGB-D
Sukhan Lee	Adaptive Bayesian Recognition and Pose Estimation of 3D Industrial Objects with Optimal Feature Selection	2016	Características globais e locais	Extração de característica do objeto alvo, posterior comparação com os objetos na base de dados através de uma rede Bayesiana e estimativa da pose com base em registro	Nuvem de pontos e imagens 2D
Iyappan	False Mapped Feature Removal In Spin Images Based 3D Ear Recognition	2016	Spin images - características locais	Extração e correspondência dos descritores das imagens alvo e teste seguidos por fusão de características e remoção de correspondências erradas	Nuvem de pontos
Bedkowski	Intelligent Mobile System for Improving Spatial Design Support and Security Inside Buildings	2016	Características locais	Mapeamento do ambiente por aquisição de nuvem de pontos, caracterização de possíveis objetos e treinamento de um classificador para identificação dos mesmo	Nuvem de pontos
Aldoma	A Global Hypothesis Verification Framework for 3D Object Recognition in Clutter	2016	Características locais	Utiliza um descritor de características locais para descrever modelo e cena para posterior correspondência e refinamento da hipótese de correspondência	Nuvem de pontos
Kasai	GOOD: A global orthographic object descriptor for 3D object recognition and manipulation	2016	Descritor de características	Utiliza uma técnica de aprendizagem associado ao descritor apresentado para posterior classificação com base no classificador treinado	Nuvem de pontos
Fuhai Chen	Multimodal learning for view-based 3D object classification	2016	Três representações com base em descritores de características Sift, Outline Fourier descriptor e Zernike Moments descriptor	SVM associado com 3 formas de descritores para treinamento e reconhecimento de objetos 3D	Imagens 2D
Chi Li	Hierarchical semantic parsing for object pose estimation in densely cluttered scenes	2016	Características locais - LAB (cor), SIFT(gradiente) e FPFH(geometria 3D)	Segmentação semântica da cena, particionando a cena em diferentes regiões dos objetos e estimativa da pose dos objetos através de um método de registro de modelo	Imagens RGB-D
Logoglu	CoSPAIR: Colored Histograms of Spatial Concentric Surflet-Pairs for 3D object recognition	2016	Características locais - Histograms of Spatial Concentric Surflet-Pairs (SPAIR) e o Colored SPAIR (CoSPAIR)	Descreve os modelos e as cenas de teste com base no descritor proposto e faz o reconhecimento com base em correspondência e voto	Nuvem de pontos
Naji	Recognition of 3D Objects using Heat Diffusion Equations and Random Forests	2016	Características derivadas das equações heat	Extração de características com base nas equações <i>heat</i> , treinamento de um classificador e posterior classificação dos objetos 3D em classes	Modelos 3D

Syed Shah	Keypoints-based surface representation for 3D modeling and 3D object recognition	2016	Keypoints-based Surface Representation	Montagem de uma base de modelos descrita na forma da representação proposta seguido do cálculo de correspondência entre modelos e cena e verificação desta correspondência	Imagens de profundidade
Deinzer	Aspects of Optimal Viewpoint Selection and Viewpoint Fusion	2006	Visão	Seleção de pontos de visão ótimo e fusão de diferentes pontos de visão	Imagens 2D
Rui Nian	Probabilistic 3D object recognition from 2D invariant view sequence based on similarity	2007	Modelo de visão	Utiliza modelos de visão associado com aprendizagem destes modelos para a estimativa estatística do resultado	Sequência de imagens 2D
Zografos	Pose-invariant 3d object recognition using linear combination of 2d views and evolutionary optimisation	2007	Com base em 2 visões base	Correspondência entre imagens base e alvo em termos de métrica de similaridade	Imagens 2D
Vázquez	3D Object Recognition Based on Low Frequency Response and Random Feature Selection	2007	Visão	Descreve um padrão com base em baixas frequências, seleciona pontos randomicamente com características sutis e treina um DAM para o reconhecimento desse padrões com base na imagem de entrada	Imagens 2D
Luciw	Topographic Class Grouping with Applications to 3D Object Recognition	2008	Visão	Treinamento de um rede neural (MLN) com diferentes visões do objeto para posterior classificação	Imagens 2D
Polat	Genetic optimization of GRNN for pattern recognition without feature extraction	2008	Visão	Uso de um algoritmo genético para otimizar o conjunto de treinamento e, por consequência, otimizar a rede neural com o resultado obtido	Imagens 2D
Dimov	CBIR over Multiple Projections of 3D Objects	2009	Visão	Sistema de busca e correspondência entre imagens em uma biblioteca de imagens	Imagens 2D
Vázquez	The Role of the Infant Vision System in 3D Object Recognition	2009	Visão	Descreve um padrão com base em baixas frequências, seleciona pontos randomicamente com características sutis e treina um DAM para o reconhecimento desse padrões com base na imagem de entrada	Imagens 2D
Urdiales	View planning for efficient contour-based 3D object recognition	2010	Visão	Correspondência entre objetos modelo e o capturado e como o processo proposto, de seleção da melhor sequência de visões, auxilia no reconhecimento	Modelo 3D
Efremova	Inferotemporal Network Model for 3D Object Recognition	2011	Múltiplas Visões do objeto	Treinamento de rede neural e posterior classificação	Imagens 2D
Bo Pang	An Effective Way of 3D Model Representation in Recognition System	2011	Híbrido Característica e Visão	Monta o modelo com base em características de diversa visões representativas e compara os modelos com o objeto de entrada em termos de uma métrica de distância	Imagens 2D
Ulrich	Combining Scale-Space and Similarity-Based Aspect Graphs for Fast 3D Object Recognition	2012	Modelo hierárquico de visões	Correspondência do modelo hierárquico de visões construído com base em modelos CAD e as imagens adquiridas, com base em métrica de similaridade	Imagens 2D e modelo CAD
Elons	A proposed PCNN features quality optimization technique for pose-invariant 3D Arabic sign language recognition	2013	Múltiplas Visões do objeto	Gera características de assinatura 3D e com base nela treina uma rede neural para classificação	Imagens 2D
Domingo Mery	Automated X-ray object recognition using an efficient search algorithm in multiple views	2013	Múltiplas Visões do objeto	Modelagem e correspondência com base em múltiplas visões	Imagens 2D
Efremova	An Inferior Temporal Cortex Model for Object Recognition and Classification	2014	Múltiplas Visões do objeto	Treinamento de rede neural e posterior classificação	Imagens 2D
Guan pang	Fast and Robust Multi-view 3D Object Recognition in Point Clouds	2015	Múltiplas Visões do objeto	Projeção de uma nuvem de pontos em várias imagens de profundidade com diversas visões, detecção 2D e posterior reprojeção 3D para detecção do objeto	Nuvem de pontos
Faulhammer	Multi-view hypotheses transfer for enhanced object recognition in clutter	2015	Múltiplas Visões do objeto	Combinação de múltiplos reconhecimentos por visão única com a finalidade de melhorar o resultado de reconhecimento final por múltiplas visões	Imagens RGB-D
Yi-Chen Chen	Salient views and view-dependent dictionaries for object recognition	2015	Visões salientes	Representação do objeto em um modelo de representação de visões salientes e seguido da classificação da visão do objeto, descrito na mesma forma de representação, com um critério de vizinho mais próximo	Vídeos

Marini	Structural Shape Prototypes for the Automatic Classification of 3D Objects	2007	Grafo representando aspectos topológicos e geométricos	Representação das classes de modelos do conjunto de modelos por meio de grafos que codificam descritores estruturais e posterior reconhecimento de um modelo inquerido por meio do cálculo de semelhança entre grafos	Modelos 3D
Shengping Xia	3D Object Recognition Using Hyper-Graphs and Ranked Local Invariant Features	2008	Grafo - Class Specific Hyper Graph	Monta-se um modelo de grafo para cada objeto e utiliza este modelo na parte de reconhecimento de outros objetos	Imagens 2D
Aouada	MEANINGFUL 3D SHAPE PARTITIONING USING MORSE FUNCTIONS	2009	Grafos Reeb	Particionamento do objeto em um formato primitivo com grafo e posterior cálculo de similaridade entre grafos	Modelos 3D
Noma	Sparse Representations for Efficient Shape Matching	2010	Grafo	Representação esparsa de formato e treinamento de um classificador	Imagens 2D
Huimin Ma	Multi-resolution recognition of 3D objects based on visual resolution limits	2010	Grafos de aspectos de múltiplas resoluções	Base de dados descrita na forma de grafos de aspecto e comparação desta representação com objetos na cena segmentados e pré-processados	Modelos 3D
Bonev	Information-theoretic Feature Selection from Unattributed Graphs	2010	Grafos Reeb	Monta-se um grafo com base na malha de pontos, extrai medidas dos grafos e classifica com base apenas nestas medidas topológicas	Malha de triângulos
Bonev	High-Dimensional Spectral Feature Selection for 3D Object Recognition Based on Reeb Graphs	2011	Grafos Reeb	Monta-se um grafo com base na malha de pontos, extrai medidas dos grafos e classifica com base apenas nestas medidas topológicas	Malha de triângulos
Kuk-Jin	Recognizing 3D Objects with 3D Information from estéreo Vision	2010	Grafo	Extração de características dos objetos da base de dados e do objeto inquerido, construção da representação por grafo e posterior correspondência entre grafos e verificação das hipóteses geradas	Imagens RGB-D
Mengjie Hu	3-D Object Recognition via Aspect Graph Aware 3-D Object Representation	2015	Grafo de aspecto	Representação por grafo de aspecto, cálculo de correspondências das características entre o modelo representado pelo grafo e a imagem de entrada, refinamento dessas correspondências e posterior estimativa de pose e localização do objeto	Nuvem de pontos e modelo 3D
Madi	Graph Edit Distance based on Triangle-Stars Decomposition for Deformable 3D Objects Recognition	2015	Componentes conexas em formato de estrelas triangulares	Descrição em termos de estrelas triangulares e posterior correspondência com base na medida de dissimilaridade das representações por estrelas triangulares	Malha de triângulos
Sen Wang	3D Surface Matching and Recognition Using Conformal Geometry	2006	Conformal Maps	Cálculo de similaridade entre superfícies descritas em termos da representação utilizada	Imagens 2D
Sen Wang	Conformal Geometry and Its Applications on 3D Shape Matching, Recognition, and Stitching	2007	Conformal Maps	Cálculo de similaridade entre superfícies descritas em termos da representação utilizada	Imagens 2D
Kordelas	Viewpoint independent object recognition in cluttered scenes exploiting ray-triangle intersection and SIFT algorithms	2010	Mapas de distância e descritores de característica	Representação de modelos utilizando mapas de distância, imagens em tons de cinza e SIFT frames e posterior correspondência dessas informações com as mesmas informações retiradas da cena	Modelos 3D
Pintilie	Conformal Mapping by Computationally Efficient Methods	2010	Conformal Maps	Calcula-se o conformal map para as superfícies e utiliza as transformações realizadas para classificar os objetos	Superfícies 3D
Rodrigues	Cortical 3D Face and Object Recognition Using 2D Projections	2012	Mapas de saliência e simbólicos	Correspondência entre templates com as representações das imagens de entrada	Imagens 2D e modelos 3D
Anand	A Microscopic Framework For Distributed Object-Recognition & Pose-Estimation	2006	Modelo híbrido - grafo de função de distribuição de probabilidade (FDP) radial, conjunto de FDP direcionais e árvore de nível de Confiança	Reconhecimento de objetos 3D e estimativa de pose com base em múltiplas visões de um time de robôs e no modelo apresentado	Imagens 2D
Kushal	Flexible Object Models for Category-Level 3D Object Recognition	2007	Partial surface models	Aprendizagem de partial surface models da cena e do modelo e posterior construção de um modelo probabilístico com base nas correspondências entre PSMs para a determinação do objeto e de sua localização	Imagens 2D

Grzegorzek	STATISTICAL 3D OBJECT CLASSIFICATION AND LOCALIZATION WITH CONTEXT MODELING	2007	Modelo estatístico	Modelagem estatística dos objetos e do contexto para posterior classificação e localização dos objetos na cena	Imagens 2D
Aouat	3D object indexing and recognition	2008	Modelos com características geométricas	Descrição do objeto 3D em vários pontos de vista, cálculo das características geométricas e comparação destas características com uma base de modelos geométricos	Modelos 3D
Lina	Construction of Appearance Manifold with Embedded View-Dependent Covariance Matrix for 3D Object Recognition	2008	Matriz de covariância e appearance manifold	Constrói um modelo com base nas imagens de treinamento, seguido da comparação da imagem de entrada com o modelo em termos de distância de mahalanobis	Imagens 2D
Truong	Model-based Recognition of 3D Objects using Intersecting Lines	2008	Modelos com características geométricas	Extração de características geométricas e posterior correspondência das mesmas com um modelo paralelepípedo	Imagens 2d e nuvem de pontos
Nair	3D Object Recognition with Deep Belief Nets	2009	Modelo de máquina Boltzman	Treinamento da estrutura de Deep Belief Nets e posterior classificação	Par de imagens estéreo em tons de cinza
Khatun	3D Shape Recognition Using Wavelet Transform Based on Ellipsoid	2010	Representação através de formato elipsoidal	Método de representação de formato e descrição de formato para o reconhecimento e recuperação de formas 3D	Modelos 3D
Lina	Recognition of 3D Objects in Various Capturing Conditions Using Appearance Manifolds	2010	Modelo de Appearance Manifold	Construção dos Appearance Manifold e posterior uso da distância de mahalanobis para reconhecer a imagem de entrada com relação ao manifold construído	Imagens 2D
Raytchev	View-Invariant Object Recognition with Visibility Maps	2010	Modelo com base em características	Extraí características das imagens, monta um modelo e depois com base neste modelo compara novas imagens segundo uma métrica de distância no subespaço montado	Imagens 2D
Sukhan Lee	Probabilistic 3D Object Recognition Based on Multiple Interpretations Generation	2010	Modelo poliédrico em termos de vértices, arestas e superfícies	Calcula a probabilidade entre o modelo e a imagem visando uma estimativa de pose	Imagens estéreo
Sukhan Lee	Probabilistic 3D Object Recognition with Both Positive and Negative Evidences	2011	Modelo poliédrico em termos de vértices, arestas e superfícies	Calcula a probabilidade entre o modelo e a imagem visando uma estimativa de pose	Imagens estéreo
Khatun	The Effectiveness of Ellipsoidal Shape Representation Technique for 3D Object Recognition System	2011	Representação através de formato elipsoidal	Método de representação de formato e descrição de formato para o reconhecimento e recuperação de formas 3D	Modelos 3D
Yong-Hui Xu	LABEL TRANSFER FOR JOINT RECOGNITION AND SEGMENTATION OF 3D OBJECT	2012	Modelo de segmentação com base em octree	Extraí características da imagem RGB-D, faz a correspondência com imagens armazenadas na base de dados, realiza a transferência de rótulos e por fim, realiza a segmentação do objeto	Imagens RGB-D
Ibrayev	Recognition of Curved Surfaces From "One-Dimensional" Tactile Data	2012	Modelo de superfície	Registro de superfícies entre o modelo da superfície armazenado na base e os dados adquiridos pelo toque no objeto	Informação de contato com o objeto
Dahua Lin	Holistic Scene Understanding for 3D Object Detection with RGBD Cameras	2013	Cubóides	Integração das informações utilizando o Conditional Random field e uma função de objetividade para ranqueamento de regiões com os cuboides	Nuvem de pontos
Kent	Crowdsourcing the Construction of a 3D Object Recognition Database for Robotic Grasping	2014	Múltiplos modelos por objeto	Construção de uma representação do objeto com base em múltiplas nuvens de pontos seguido da tentativa de registro de uma nova nuvem de pontos para cada um dos modelos armazenados e cálculo de métricas para classificar a nuvem de pontos do objeto desconhecido com base no menor resultado de erro proveniente das métricas calculadas	Nuvem de pontos
Hejrati	Analysis by Synthesis: 3D Object Recognition by Object Reconstruction	2014	Modelo de possíveis interpretações geométricas	Construção do modelo e seleção do mesmo que melhor concorda com as evidências visuais medidas	Imagens 2D
Muhammad zia	Are Cars Just 3D Boxes? – Jointly Estimating the 3D Shape of Multiple Objects	2014	Modelos CAD	Inferência de probabilidades com base em modelo dos objetos para compreensão de cena	Imagens 2D e modelos 3D
Wentao fan	3D Object Modeling and Recognition via Online Hierarchical Pitman-Yor Process Mixture Learning	2015	Modelo com base em mistura de distribuições	Aprendizagem do modelo descrito e posterior reconhecimento de outros objetos	Modelos 3D

Pengfei Han	CAD-based 3D objects recognition in monocular images for mobile augmented reality	2015	Modelos CAD	Utilização do centro de massa do objeto para correspondência com a imagem adquirida em termos de posicionamento seguida da renderização do modelo CAD buscando o posicionamento final do objeto	Modelos CAD e imagens 2D
Mabel Zhang	A Triangle Histogram for Object Classification by Tactile Sensing	2016	Histograma de triângulos	Utiliza um histograma de triângulos para a representação do objeto e utilização deste descritor para treinamento de um classificador visando a classificação de outros objetos	Informação de contato com o objeto
Fukui	A Framework for 3D Object Recognition Using the Kernel Constrained Mutual Subspace Method	2006	Subespaços	Modelagem com base em subespaços e cálculo de similaridade entre subespaços	Imagens 2D
Fukui	The Kernel Orthogonal Mutual Subspace Method and Its Application to 3D Object Recognition	2007	Subespaços	Modelagem com base em subespaços ortogonais e cálculo de similaridade entre subespaços	Imagens 2D
Jianing Wu	Multiple View Based 3D Object Classification Using Ensemble Learning of Local Subspaces	2008	Subespaços	Geração de subespaços locais para cada classe de imagens e posterior cálculo de similaridade entre um subespaço de uma imagem de entrada e o subespaços locais da classe	Imagens 2D
Hotta	Pose independent object classification from small number of training samples based on kernel principal component analysis of local parts	2009	Subespaços	Representação de classes em termos de subespaço e posterior cálculo de similaridade entre subespaços e amostra de teste	Imagens 2D
Akihiro	Compound Mutual Subspace Method for 3D Object Recognition: A Theoretical Extension of Mutual Subspace Method	2010	Subespaços	Modelagem com base em subespaços e cálculo de similaridade entre subespaços	Imagens 2D
Igarashi	3D Object Recognition Based on Canonical Angles between Shape Subspaces	2010	Subespaços	Cálculo dos ângulos canônicos com base no subespaço de formato dos objetos e medição da similaridade entre os formatos com base no valor de similaridade obtido com estes ângulos canônicos	Imagens 2D
Kise	1.5 Million Subspaces of a Local Feature Space for 3D Object Recognition	2011	Subespaços	Cálculo de similaridade entre característica e subespaço ou subespaço e subespaço, associado a um esquema de votação, para obter o resultado de reconhecimento do objeto	Imagens 2D
Kobayashi	Generalized Mutual Subspace Based Methods for Image Set Classification	2012	Subespaços	Treinamento de um subespaço discriminativo e classificação de subespaços com base em uma medida de similaridade utilizando NN	Imagens 2D
Fukui	Difference Subspace and Its Generalization for Subspace-Based Methods	2015	Subespaço formado por características de formato	Construção do subespaço para todas as classes de objetos e posterior cálculo de similaridade entre estes subespaços e o subespaço de entrada formado por um conjunto de imagens de entrada de um objeto desconhecido para posterior classificação com base no maior valor de similaridade	Imagens 2D
Mian	Three-Dimensional Model-Based Object Recognition and Segmentation in Cluttered Scenes	2010	Tensores	Correspondência através de esquema de votação, entre objetos na biblioteca com objetos na cena, seguido de um cálculo de similaridade do modelo com o maior número de votos e posterior alinhamento e segmentação do objeto na cena	Imagem de profundidade ou modelo 3D
Ben-Yaacov	Recognition of 3D Objects Based on Implicit Polynomials	2010	Tensores de Implicit polynomials	Treinamento de um classificador com a representação de tensor apresentada e posterior classificação	Imagens 2D
Smeets	Inelastic Deformation Invariant Modal Representation for Non-rigid 3D Object Recognition	2010	Tensores de distância e matriz de distância	Extração dos descritores apresentados e comparação com base nestes descritores	Malha de pontos

Yaniv Gur	GENERALIZED HARDI INVARIANTS BY METHOD OF TENSOR CONTRACTION	2014	Tensores	Cálculo de invariantes através de técnicas de contração de tensor e posterior classificação de voxels utilizando as invariantes	Nuvem de pontos e imagens 2D
Orts-Escolano	Real-time 3D semi-local surface patch extraction using GPGPU	2015	Descritor tensor	Extração e correspondência entre descritores do modelo e da cena	Imagens RGBD
Yoshikawa	Three-Dimensional Object Recognition using Multiplex Complex Amplitude Information with Support Function	2006	Height transformed Complex Amplitude	Correlacionamento entre Height transformed Complex Amplitude de referência e a capturada no objeto sendo visualizado	Imagens 2D
Nelleri	Three-dimensional object recognition from digital Fresnel hologram by wavelet matched filtering	2006	Holografia	Descreve o objeto 3D em termos de um holograma e diferencia cada holograma como único através da aplicação de um filtro para aumentar as características discriminatórias de cada objeto, classificando em duas classes	Hologramas 3D dos objetos
Chye-Hwa Loo	Optoelectronic 3-D Object Classification From 2-D Images	2008	Conjuntos de filtros de correlação	Cálculo de filtros de correlação com base em imagens de treinamento e posterior classificação da imagem de entrada utilizando os filtros da imagem de referência por meio do FJTC	Imagens 2D
Kitaaki	High speed 3-D registration using GPU	2008	Depth Aspect Image (DAI)	Correspondência entre a representação por DAI armazenada na base e a cena e posterior alinhamento para estimativa de pose e posicionamento	Nuvens de pontos
Xu Zhang	An Isomap-Eigenanalysis-Regression Pose Estimation Algorithm of Three-Dimensional Object	2008	Eigen-images	Representação por eigen-imagens e comparação, por distância, das eigen-imagens de entrada e do conjunto de treinamento para estimativa da pose	Modelos 3D
Cuong Manh Do	Three-Dimensional Object Recognition With Multiview Photon-Counting Sensing and Imaging	2009	Imagens photon-counting	Treinamento com base na imagens de photon-counting, redimensionadas com o PCA e projetadas no domínio ICA, e posterior classificação com base no cálculo de similaridade entre as imagens de <i>photon-counting</i> de entrada e de treinamento	Imagem integral

REFERENCES

- [1] B. Kitchenham, "Procedures for Performing Systematic Reviews," Joint Technical Report TR/SE-0401, Tech. Rep., 2004.
- [2] N. Health, M. R. C. (Australia), and N. Staff, *How to Review the Evidence: Systematic Identification and Review of the Scientific Literature*, ser. Handbook series on preparing clinical practice guidelines. National Health and Medical Research Council, 2000. [Online]. Available: <https://books.google.com.br/books?id=G8JmAAAACAAJ>
- [3] —, *How to Use the Evidence: Assessment and Application of Scientific Evidence*, ser. Handbook series on preparing clinical practice guidelines. National Health and Medical Research Council, 2000. [Online]. Available: <https://books.google.com.br/books?id=7ioKGQAACAAJ>
- [4] C. Collaboration, *Cochrane Reviewers' Handbook*, ser. Version 4.2.1. National Health and Medical Research Council, 2003.
- [5] U. of York. NHS Centre for Reviews and Dissemination, *Undertaking systematic reviews of research on effectiveness: CRD's guidance for those carrying out or commissioning reviews*, ser. CRD report. NHS Centre for Reviews and Dissemination, University of York, 2001. [Online]. Available: <https://books.google.com.br/books?id=2rhpAAAAMAAJ>
- [6] M. A. A. Mhamdi and D. Ziou, "A local approach for 3d object recognition through a set of size functions," *Image and Vision Computing*, vol. 32, no. 12, pp. 1030 – 1044, 2014. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0262885614001462>
- [7] I. Atmosukarto, K. Wilamowska, C. Heike, and L. G. Shapiro, "3d object classification using salient point patterns with application to craniofacial research," *Pattern Recognition*, vol. 43, no. 4, pp. 1502 – 1517, 2010. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0031320309004178>
- [8] A. Pichler, H. Bauer, C. Eberst, C. Heindl, and J. Minichberger, "Towards more agility in robot painting through 3d object recognition," in *Intelligent Production Machines and Systems*, D. Pham, E. Eldukhri, and A. Soroka, Eds. Oxford: Elsevier Science Ltd, 2006, pp. 608 – 613. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/B9780080451572501061>
- [9] W.-Y. Lin, "Robust geometrically invariant features for two-dimensional shape matching and three-dimensional face recognition," Ph.D. dissertation, University of Wisconsin at Madison, Madison, WI, USA, 2006, aAI3234848.
- [10] P. J. Phillips, P. J. Flynn, T. Scruggs, K. W. Bowyer, J. Chang, K. Hoffman, J. Marques, J. Min, and W. Worek, "Overview of the face recognition grand challenge," in *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, vol. 1, June 2005, pp. 947–954 vol. 1.
- [11] P. K. Shivaswamy and T. Jebara, "Permutation invariant svms," in *Proceedings of the 23rd International Conference on Machine Learning*, ser. ICML '06. New York, NY, USA: ACM, 2006, pp. 817–824. [Online]. Available: <http://doi.acm.org/10.1145/1143844.1143947>
- [12] H. W. Kuhn, *The Hungarian Method for the Assignment Problem*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2010, pp. 29–47.
- [13] J.-M. Geusebroek, G. J. Burghouts, and A. W. Smeulders, "The amsterdam library of object images," *International Journal of Computer Vision*, vol. 61, no. 1, pp. 103–112, 2005. [Online]. Available: <http://dx.doi.org/10.1023/B:VISI.0000042993.50813.60>
- [14] Y. Tan, B. C. Matei, and H. Sawhney, "Exploiting model similarity for indexing and matching to a large model database," in *Proceedings of the 9th European Conference on Computer Vision - Volume Part II*, ser. ECCV'06. Berlin, Heidelberg: Springer-Verlag, 2006, pp. 536–548.
- [15] P. J. Besl and N. D. McKay, "A method for registration of 3-d shapes," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 14, no. 2, pp. 239–256, Feb 1992.
- [16] A. E. Johnson and M. Hebert, "Using spin images for efficient object recognition in cluttered 3d scenes," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 21, no. 5, pp. 433–449, May 1999.
- [17] Y. Shan, B. Matei, H. S. Sawhney, R. Kumar, D. Huber, and M. Hebert, "Linear model hashing and batch ransac for rapid and accurate object recognition," in *Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2004. CVPR 2004.*, vol. 2, June 2004, pp. II–121–II–128 Vol.2.
- [18] S. Lee, E. Kim, and Y. Park, "3d object recognition using multiple features for robotic manipulation," in *Proceedings 2006 IEEE International Conference on Robotics and Automation, 2006. ICRA 2006.*, May 2006, pp. 3768–3774.
- [19] O. Stasse, S. Dupitier, and K. Yokoi, "3d object recognition using spin-images for a humanoid stereoscopic vision system," in *2006 IEEE/RSJ International Conference on Intelligent Robots and Systems*, Oct 2006, pp. 2955–2960.
- [20] Z. Chen, R. Zhao, and Y. Zhang, "Geometric hashing using 3d aspects and constrained structures," in *2006 8th international Conference on Signal Processing*, vol. 2, Nov 2006.
- [21] K. S. Roh and I. S. Kweon, "3-d object recognition using a new invariant relationship by single-view," *Pattern Recognition*, vol. 33, no. 5, pp. 741 – 754, 2000. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0031320399000928>
- [22] S. N. Raptis and D. Koutsouris, "Bayesian fusion of contour descriptions: Application to 3-d object and face recognition," in *2006 IET Conference on Crime and Security*, June 2006, pp. 438–444.
- [23] S. A. Nene, S. K. Nayar, and H. Murase, "Columbia object image library (coil-100)," Computer Vision Laboratory - Department of Computer Science - Columbia University, Tech. Rep., 1996.
- [24] E. Delponte, E. Arnaud, F. Odone, and A. Verri, "Analysis on a local approach to 3d object recognition," in *Proceedings of the 28th Conference on Pattern Recognition*, ser. DAGM'06. Berlin, Heidelberg: Springer-Verlag, 2006, pp. 253–262.
- [25] E. Arnaud, F. Odone, and A. Verri, "Trains of keypoints for 3d object recognition," in *Proceedings of the 18th International Conference on Pattern Recognition - Volume 02*, ser. ICPR '06. Washington, DC, USA: IEEE Computer Society, 2006, pp. 1014–1017. [Online]. Available: <http://dx.doi.org/10.1109/ICPR.2006.1133>
- [26] A. Kushal and J. Ponce, "Modeling 3d objects from stereo views and recognizing them in photographs," in *Proceedings of the 9th European Conference on Computer Vision - Volume Part II*, ser. ECCV'06. Berlin, Heidelberg: Springer-Verlag, 2006, pp. 563–574.
- [27] K. Mikolajczyk and C. Schmid, "An affine invariant interest point detector," in *Proceedings of the 7th European Conference on Computer Vision-Part I*, ser. ECCV '02. London, UK, UK: Springer-Verlag, 2002, pp. 128–142. [Online]. Available: <http://dl.acm.org/citation.cfm?id=645315.649184>
- [28] N. Arana-Daniel and E. Bayro-Corrochano, "Mimo svms for 3d object classification," in *The 2006 IEEE International Joint Conference on Neural Network Proceedings*, 2006, pp. 1628–1635.
- [29] D. Hestenes and G. Sobczyk, *Clifford algebra to geometric calculus : a unified language for mathematics and physics*. D. Reidel ; Distributed in the U.S.A. and Canada by Kluwer Academic Publishers, 1984.
- [30] S. Ramalingam, Z.-Q. Liu, and D. Iourinski, "Curvature-based fuzzy surface classification," *IEEE Transactions on Fuzzy Systems*, vol. 14, no. 4, pp. 573–589, Aug 2006.
- [31] R. Soodamani and Z. Q. Liu, "Fuzzy surface descriptions for 3-d machine vision," in *1995 IEEE International Conference on Systems, Man and Cybernetics. Intelligent Systems for the 21st Century*, vol. 4, Oct 1995, pp. 3238–3243 vol.4.
- [32] T. Caelli and A. Dreier, "Variations on the evidence-based object recognition theme," *Pattern Recognition*, vol. 27, no. 2, pp. 185 – 204, 1994. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/0031320394900531>
- [33] T.-J. Fan, *Describing and recognizing 3-D objects using surface properties*. Springer Science & Business Media, 1990.
- [34] P. J. Besl and R. C. Jain, "Invariant surface characteristics for 3d object recognition in range images," *Computer Vision, Graphics, and Image Processing*, vol. 33, no. 1, pp. 33 – 80, 1986. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/0734189X86902203>
- [35] R. Soodamani and Z. Q. Liu, "Object recognition using fuzzy modelling and fuzzy matching," in *1998 IEEE International Conference on Fuzzy Systems Proceedings. IEEE World Congress on Computational Intelligence (Cat. No.98CH36228)*, vol. 1, May 1998, pp. 165–170 vol.1.
- [36] J. Assfalg, K. M. Borgwardt, and H.-P. Kriegel, "3dstring: A feature string kernel for 3d object classification on voxelized data," in *Proceedings of the 15th ACM International Conference on Information and Knowledge Management*, ser. CIKM '06. New York, NY, USA: ACM, 2006, pp. 198–207. [Online]. Available: <http://doi.acm.org/10.1145/1183614.1183647>
- [37] H. M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T. N. Bhat, H. Weissig, I. N. Shindyalov, and P. E. Bourne, "The protein data bank," *Nucleic Acids Res*, vol. 28, pp. 235–242, 2000.
- [38] D.-Y. Chen, X.-P. Tian, Y.-T. Shen, and M. Ouhyoung, "On visual similarity based 3d model retrieval," *Computer Graphics Forum*, vol. 22, no. 3, pp. 223–232, 2003. [Online]. Available: <http://dx.doi.org/10.1111/1467-8659.00669>

- [39] P. Shilane, P. Min, M. Kazhdan, and T. Funkhouser, "The princeton shape benchmark," in *In Shape Modeling International*, 2004, pp. 167–178.
- [40] R. Leitner, "Learning 3d object recognition from an unlabelled and unordered training set," in *Proceedings of the 3rd International Conference on Advances in Visual Computing - Volume Part I*, ser. ISVC'07. Berlin, Heidelberg: Springer-Verlag, 2007, pp. 644–651. [Online]. Available: <http://dl.acm.org/citation.cfm?id=1779178.1779251>
- [41] J. Matas, O. Chum, M. Urban, and T. Pajdla, "Robust wide-baseline stereo from maximally stable extremal regions," *Image and Vision Computing*, vol. 22, no. 10, pp. 761 – 767, 2004, british Machine Vision Computing 2002. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0262885604000435>
- [42] D. G. Lowe, "Object recognition from local scale-invariant features," in *Proceedings of the International Conference on Computer Vision - Volume 2 - Volume 2*, ser. ICCV '99. Washington, DC, USA: IEEE Computer Society, 1999, pp. 1150–. [Online]. Available: <http://dl.acm.org/citation.cfm?id=850924.851523>
- [43] Y. Rui, A. C. She, and T. S. Huang, "Modified fourier descriptors for shape representation – a practical approach," in *PROC OF FIRST INTERNATIONAL WORKSHOP ON IMAGE DATABASES AND MULTI MEDIA SEARCH*, 1996.
- [44] D. G. Lowe, "Local feature view clustering for 3d object recognition," in *CVPR*, 2001.
- [45] R. Leitner and H. Bischof, "Recognition of 3d objects by learning from correspondences in a sequence of unlabeled training images," in *Proceedings of the 27th DAGM Conference on Pattern Recognition*, ser. PR '05. Berlin, Heidelberg: Springer-Verlag, 2005, pp. 369–376.
- [46] G. Kootstra, J. Ypma, and B. de Boer, "Exploring objects for recognition in the real world," in *2007 IEEE International Conference on Robotics and Biomimetics (ROBIO)*, Dec 2007, pp. 429–434.
- [47] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *International Journal of Computer Vision*, vol. 60, no. 2, pp. 91–110, 2004. [Online]. Available: <http://dx.doi.org/10.1023/B:VISI.0000029664.99615.94>
- [48] S. Marsland, J. Shapiro, and U. Nehmzow, "A self-organising network that grows when required," *Neural Networks*, vol. 15, pp. 1041 – 1058, 2002. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0893608002000783>
- [49] K. Okada, M. Kojima, S. Tokutsu, T. Maki, Y. Mori, and M. Inaba, "Multi-cue 3d object recognition in knowledge-based vision-guided humanoid robot system," in *2007 IEEE/RSJ International Conference on Intelligent Robots and Systems*, Oct 2007, pp. 3217–3222.
- [50] C. Tomasi and T. Kanade, "Detection and tracking of point features," *International Journal of Computer Vision*, Tech. Rep., 1991.
- [51] P. Pérez, C. Hue, J. Vermaak, and M. Gangnet, "Color-based probabilistic tracking," in *Proceedings of the 7th European Conference on Computer Vision-Part I*, ser. ECCV '02. London, UK, UK: Springer-Verlag, 2002, pp. 661–675. [Online]. Available: <http://dl.acm.org/citation.cfm?id=645315.649166>
- [52] G. Kitagawa, "Monte carlo filter and smoother for non-gaussian non-linear state space models," *Journal of computational and graphical statistics*, vol. 5, no. 1, pp. 1–25, 1996.
- [53] E. Delponte, N. Noceti, F. Odone, and A. Verri, "Appearance-based 3d object recognition with time-invariant features," in *14th International Conference on Image Analysis and Processing (ICIAP 2007)*, Sept 2007, pp. 467–474.
- [54] T. Lindeberg, "Feature detection with automatic scale selection," *International Journal of Computer Vision*, vol. 30, no. 2, pp. 79–116, 1998. [Online]. Available: <http://dx.doi.org/10.1023/A:1008045108935>
- [55] E. A. Wan and R. V. D. Merwe, "The unscented kalman filter for non-linear estimation," in *Proceedings of the IEEE 2000 Adaptive Systems for Signal Processing, Communications, and Control Symposium (Cat. No.00EX373)*, 2000, pp. 153–158.
- [56] J.-H. Zhai, X.-Z. Wang, S.-F. Zhang, and J. Li, "View-based 3d object recognition using wavelet multiscale singular-value decomposition and support vector machine," in *2007 International Conference on Wavelet Analysis and Pattern Recognition*, vol. 3, Nov 2007, pp. 1428–1432.
- [57] T. Li, Q. Li, S. Zhu, and M. Oghihara, "A survey on wavelet applications in data mining," *SIGKDD Explor. Newsl.*, vol. 4, no. 2, pp. 49–68, Dec. 2002. [Online]. Available: <http://doi.acm.org/10.1145/772862.772870>
- [58] R. W. Swiniarski and L. Hargis, "Rough sets as a front end of neural-networks texture classifiers," *Neurocomputing*, vol. 36, pp. 85 – 102, 2001, rough-neuro computing. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0925231200003374>
- [59] G. H. John, R. Kohavi, and K. Pfleger, "Irrelevant features and the subset selection problem," in *MACHINE LEARNING: PROCEEDINGS OF THE ELEVENTH INTERNATIONAL*. Morgan Kaufmann, 1994, pp. 121–129.
- [60] W. Ziarko, "Variable precision rough set model," *Journal of Computer and System Sciences*, vol. 46, no. 1, pp. 39 – 59, 1993. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0022000093900482>
- [61] A. S. Salgian, "Using multiple patches for 3d object recognition," in *2007 IEEE Conference on Computer Vision and Pattern Recognition*, June 2007, pp. 1–6.
- [62] Y. Ke and R. Sukthankar, "Pca-sift: a more distinctive representation for local image descriptors," in *Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2004. CVPR 2004.*, vol. 2, June 2004, pp. II–506–II–513 Vol.2.
- [63] A. Selinger and R. C. Nelson, "A perceptual grouping hierarchy for appearance-based 3d object recognition," *Computer Vision and Image Understanding*, vol. 76, no. 1, pp. 83 – 92, 1999. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S1077314299907889>
- [64] X. Li and I. Guskov, "3d object recognition from range images using pyramid matching," in *2007 IEEE 11th International Conference on Computer Vision*, Oct 2007, pp. 1–6.
- [65] —, "Multi-scale features for approximate alignment of point-based surfaces," in *Proceedings of the Third Eurographics Symposium on Geometry Processing*, ser. SGP '05. Aire-la-Ville, Switzerland, Switzerland: Eurographics Association, 2005. [Online]. Available: <http://dl.acm.org/citation.cfm?id=1281920.1281955>
- [66] K. Grauman and T. Darrell, "The pyramid match kernel: discriminative classification with sets of image features," in *Tenth IEEE International Conference on Computer Vision (ICCV'05) Volume 1*, vol. 2, Oct 2005, pp. 1458–1465 Vol. 2.
- [67] K. I. Chang, K. W. Bowyer, and P. J. Flynn, "An evaluation of multimodal 2d+3d face biometrics," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, no. 4, pp. 619–624, April 2005.
- [68] T. U. of Stuttgart, "Stuttgart range image database," <http://range.informatik.uni-stuttgart.de>, acessado em 2017-04-08.
- [69] J. Assfalg, M. Bertini, A. D. Bimbo, and P. Pala, "Content-based retrieval of 3-d objects using spin image signatures," *IEEE Transactions on Multimedia*, vol. 9, no. 3, pp. 589–599, April 2007.
- [70] J. C. Bezdek, J. Keller, R. Krisnapuram, and N. Pal, *Fuzzy models and algorithms for pattern recognition and image processing*. Springer Science & Business Media, 2006, vol. 4.
- [71] K. Do-Jong, P. Yong-Woon, and P. Dong-Jo, "A novel validity index for determination of the optimal number of clusters," *IEICE Transactions on Information and Systems*, vol. 84, no. 2, pp. 281–285, 2001.
- [72] J. P. Vandeborste, V. Couillet, and M. Daoudi, "A practical approach for 3d model indexing by combining local and global invariants," in *Proceedings. First International Symposium on 3D Data Processing Visualization and Transmission*, 2002, pp. 644–647.
- [73] R. Osada, T. Funkhouser, B. Chazelle, and D. Dobkin, "Shape distributions," *ACM Trans. Graph.*, vol. 21, no. 4, pp. 807–832, Oct. 2002. [Online]. Available: <http://doi.acm.org/10.1145/571647.571648>
- [74] M. Elad, A. Tal, and S. Ar, *Content Based Retrieval of VRML Objects – An Iterative and Interactive Approach*. Vienna: Springer Vienna, 2002, pp. 107–118.
- [75] D. Aouada, S. Feng, and H. Krim, "Statistical analysis of the global geodesic function for 3d object classification," in *2007 IEEE International Conference on Acoustics, Speech and Signal Processing - ICASSP '07*, vol. 1, April 2007, pp. I–645–I–648.
- [76] I. Grosse, P. B. Galván, P. Carpena, R. R. Roldán, J. Oliver, and H. E. Stanley, "Analysis of symbolic sequences using the Jensen-Shannon divergence," *Physical Review E*, vol. 65, no. 4, pp. 041905+, Mar. 2002. [Online]. Available: <http://dx.doi.org/10.1103/physreve.65.041905>
- [77] S. A. G. MCGill, "McGill 3d shape benchmark," <http://www.cim.mcgill.ca/shape/?page=downloads>, acessado em 2017-04-12.
- [78] L.-Y. Ong, C.-W. Chong, and R. Besar, "An approach to 3-d object recognition using legendre moment invariants," in *2007 International Conference on Intelligent and Advanced Systems*, Nov 2007, pp. 671–674.
- [79] L. Yang, F. Albrechtsen, and T. Taxt, "Fast computation of three-dimensional geometric moments using a discrete divergence theorem and a generalization to higher dimensions," *Graphical Models and Image Processing*, vol. 59, no. 2, pp. 97 – 108, 1997. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S1077316997904184>

- [80] G. Kordelas and P. Daras, "Recognizing 3d objects using ray-triangle intersection distances," in *2007 IEEE International Conference on Image Processing*, vol. 6, Sept 2007, pp. VI – 173–VI – 176.
- [81] A. S. Mian, M. Bennamoun, and R. Owens, "Three-dimensional model-based object recognition and segmentation in cluttered scenes," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, no. 10, pp. 1584–1601, Oct. 2006. [Online]. Available: <http://dx.doi.org/10.1109/TPAMI.2006.213>
- [82] H. Chen and B. Bhanu, "3d free-form object recognition in range images using local surface patches," *Pattern Recognition Letters*, vol. 28, no. 10, pp. 1252 – 1262, 2007. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0167865507000621>
- [83] W. Xing, W. Liu, B. Yuan, and W. Lu, "An integrated system for 3d object reconstruction and recognition," in *Proceedings of the 7th WSEAS International Conference on Simulation, Modelling and Optimization*, ser. SMO'07. Stevens Point, Wisconsin, USA: World Scientific and Engineering Academy and Society (WSEAS), 2007, pp. 281–284. [Online]. Available: <http://dl.acm.org/citation.cfm?id=1353862.1353911>
- [84] W. Xing, W. Liu, and B. Yuan, "3d object classification system based on volumetric parts," in *2007 IEEE International Conference on Systems, Man and Cybernetics*, Oct 2007, pp. 984–990.
- [85] W. Xing, B. Yuan, M. Liu, and X. Tang, "3d object classification by part features fusion," in *2006 8th international Conference on Signal Processing*, vol. 2, 2006.
- [86] W. Xing, W. Liu, and B. Yuan, "Volumetric part based 3d object classification," in *2006 5th IEEE International Conference on Cognitive Informatics*, vol. 1, July 2006, pp. 405–412.
- [87] S. Kim and I. S. Kweon, "Robust model-based scene interpretation by multilayered context information," *Computer Vision and Image Understanding*, vol. 105, no. 3, pp. 167 – 187, 2007. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S1077314206001561>
- [88] C. Sutton and A. McCallum, "Piecewise training for undirected models," in *Proceedings of the Twenty-First Conference on Uncertainty in Artificial Intelligence*, ser. UAI'05. Arlington, Virginia, United States: AUAI Press, 2005, pp. 568–575. [Online]. Available: <http://dl.acm.org/citation.cfm?id=3020336.3020405>
- [89] T. K. Lee and M. S. Drew, *3D Object Recognition by Eigen-Scale-Space of Contours*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2007, pp. 883–894.
- [90] T. M.J., "The object databank," <http://wiki.cncb.cmu.edu/images/TheObjectDatabank.zip>, acessado em 2017-02-02.
- [91] T. C. Kietzmann, S. Lange, and M. Riedmiller, "Incremental grlvq: Learning relevant features for 3d object recognition," *Neurocomputing*, vol. 71, no. 13–15, pp. 2868 – 2879, 2008, artificial Neural Networks (ICANN 2006) / Engineering of Intelligent Systems (ICEIS 2006). [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0925231207002652>
- [92] S. Xu and Q. c. Peng, "3d object recognition using multiple features and neural network," in *2008 IEEE Conference on Cybernetics and Intelligent Systems*, Sept 2008, pp. 434–439.
- [93] M. Stricker and M. Orengo, "Similarity of color images," in *Storage and Retrieval for Image and Video Databases (SPIE)*, 1995, pp. 381–392.
- [94] R. M. Haralick, K. Shanmugam, and I. Dinstein, "Textural features for image classification," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. SMC-3, no. 6, pp. 610–621, Nov 1973.
- [95] M.-K. Hu, "Visual pattern recognition by moment invariants," *IRE Transactions on Information Theory*, vol. 8, no. 2, pp. 179–187, February 1962.
- [96] J. Flusser and T. Suk, "Pattern recognition by affine moment invariants," *Pattern Recognition*, vol. 26, no. 1, pp. 167 – 174, 1993. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/003132039390098H>
- [97] Y. Zhong, "Intrinsic shape signatures: A shape descriptor for 3d object recognition," in *2009 IEEE 12th International Conference on Computer Vision Workshops, ICCV Workshops*, Sept 2009, pp. 689–696.
- [98] A. Frome, D. Huber, R. Kolluri, T. Bülow, and J. Malik, *Recognizing Objects in Range Data Using Regional Point Descriptors*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2004, pp. 224–237.
- [99] S. Kim and I. S. Kweon, "Scalable representation and learning for 3d object recognition using shared feature-based view clustering," in *Proceedings of the 7th Asian Conference on Computer Vision - Volume Part II*, ser. ACCV'06. Berlin, Heidelberg: Springer-Verlag, 2006, pp. 561–570.
- [100] —, "Scalable representation for 3d object recognition using feature sharing and view clustering," *Pattern Recognition*, vol. 41, no. 2, pp. 754 – 773, 2008. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0031320307001410>
- [101] H. T. Ho and D. Gibbins, "Multi-scale feature extraction for 3d models using local surface curvature," in *Proceedings of the 2008 Digital Image Computing: Techniques and Applications*, ser. DICTA '08. Washington, DC, USA: IEEE Computer Society, 2008, pp. 16–23. [Online]. Available: <http://dx.doi.org/10.1109/DICTA.2008.64>
- [102] X. Li, A. Godil, and A. Wagan, "3d part identification based on local shape descriptors," in *Proceedings of the 8th Workshop on Performance Metrics for Intelligent Systems*, ser. PerMIS '08. New York, NY, USA: ACM, 2008, pp. 162–166. [Online]. Available: <http://doi.acm.org/10.1145/1774674.1774700>
- [103] S. Jayanti, Y. Kalyanaraman, N. Iyer, and K. Ramani, "Developing an engineering shape benchmark for {CAD} models," *Computer-Aided Design*, vol. 38, no. 9, pp. 939 – 953, 2006, shape Similarity Detection and Search for CAD/CAE Applications Shape Similarity Detection and Search for CAD/CAE Applications. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S001044850600100X>
- [104] A. S. Salgian, "Combining local descriptors for 3d object recognition and categorization," in *2008 19th International Conference on Pattern Recognition*, Dec 2008, pp. 1–4.
- [105] B. Leibe and B. Schiele, "Analyzing appearance and contour based methods for object categorization," in *2003 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2003. Proceedings.*, vol. 2, June 2003, pp. II–409–15 vol.2.
- [106] D. Wang and H. Qian, "3d object recognition by fast spherical correlation between combined view egis and pft," in *2008 19th International Conference on Pattern Recognition*, Dec 2008, pp. 1–4.
- [107] A. Mavrinac, A. Shawky, and X. Chen, "A fuzzy associative approach for recognition of 3d objects in arbitrary pose," in *2008 IEEE International Conference on Fuzzy Systems (IEEE World Congress on Computational Intelligence)*, June 2008, pp. 710–715.
- [108] S. Shahir, X. Chen, and M. Ahmadi, "Fuzzy associative database for multiple planar object recognition," in *Circuits and Systems, 2003. ISCAS '03. Proceedings of the 2003 International Symposium on*, vol. 5, May 2003, pp. V–805–V–808 vol.5.
- [109] E. González, A. Adán, V. Feliú, and L. Sánchez, "Active object recognition based on fourier descriptors clustering," *Pattern Recognition Letters*, vol. 29, no. 8, pp. 1060 – 1071, 2008, pattern Recognition in Interdisciplinary Perception and IntelligencePRIntPerIntel. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0167865507002103>
- [110] L.-L. Wan and Z.-J. Miao, "3d object classification by fuzzy knn and bayesian decision," in *Proceedings of the 2009 Fifth International Conference on Intelligent Information Hiding and Multimedia Signal Processing*, ser. IHH-MSP '09. Washington, DC, USA: IEEE Computer Society, 2009, pp. 455–458. [Online]. Available: <http://dx.doi.org/10.1109/IHH-MSP.2009.233>
- [111] H. Chen and B. Bhanu, "Efficient recognition of highly similar 3d objects in range images," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 31, no. 1, pp. 172–179, Jan. 2009. [Online]. Available: <http://dx.doi.org/10.1109/TPAMI.2008.176>
- [112] P. Yan and K. W. Bowyer, "Biometric recognition using 3d ear shape," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 29, no. 8, pp. 1297–1308, Aug 2007.
- [113] H. Chen and B. Bhanu, "Human ear recognition in 3d," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 29, no. 4, pp. 718–737, Apr. 2007. [Online]. Available: <http://dx.doi.org/10.1109/TPAMI.2007.1005>
- [114] H. Kim, J. Lee, and S. Lee, "Environment adaptive 3d object recognition and pose estimation by cognitive perception engine," in *Proceedings of the 8th IEEE International Conference on Computational Intelligence in Robotics and Automation*, ser. CIRA'09. Piscataway, NJ, USA: IEEE Press, 2009, pp. 532–539. [Online]. Available: <http://dl.acm.org/citation.cfm?id=1811259.1811367>
- [115] S. Lee, S. Lee, J. Lee, D. Moon, E. Kim, and J. Seo, "Robust recognition and pose estimation of 3d objects based on evidence fusion in a sequence of images," in *Proceedings 2007 IEEE International Conference on Robotics and Automation*, April 2007, pp. 3773–3779.
- [116] G.-S. Hsu, C.-Y. Lin, and J.-S. Wu, "Real-time 3-d object recognition using scale invariant feature transform and stereo vision," in *2009 4th International Conference on Autonomous Robots and Agents*, Feb 2009, pp. 239–244.
- [117] M. K. Osman, M. Y. Mashor, M. R. Arshad, and Z. Saad, "3d object recognition using manfis network with orthogonal and non-orthogonal moments," in *2009 5th International Colloquium on Signal Processing Its Applications*, March 2009, pp. 302–306.

- [118] P. Pejnovic, L. Buturovic, and Z. Stojiljkovic, "Object recognition by invariants," in *Pattern Recognition, 1992. Vol. II. Conference B: Pattern Recognition Methodology and Systems, Proceedings., 11th IAPR International Conference on*. IEEE, 1992, pp. 434–437.
- [119] M. K. Osman, M. Y. Mashor, and M. R. Arshad, "An optimised camera-object setup for 3d object recognition system." *Proc. of the 2nd Int. Conf. on Artificial Intell. in Engineering and Technology, ICAIET 04*, pp. 267–273, 2004.
- [120] —, "Multi-view technique for 3-d object recognition using neuro-fuzzy system," *AIAI 2003*, 2003.
- [121] M. Himmelsbach, T. Luettel, and H. J. Wuensche, "Real-time object classification in 3d point clouds using point feature histograms," in *2009 IEEE/RSSJ International Conference on Intelligent Robots and Systems*, Oct 2009, pp. 994–1000.
- [122] D. Gibbins and L. Swierkowski, "A comparison of terrain classification using local feature measurements of 3-dimensional colour point-cloud data," in *2009 24th International Conference Image and Vision Computing New Zealand*, Nov 2009, pp. 293–298.
- [123] N. Werghi, "A discriminative 3d wavelet-based descriptors: Application to the recognition of human body postures," *Pattern Recognition Letters*, vol. 26, no. 5, pp. 663 – 677, 2005. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0167865504002429>
- [124] M. Novotni and R. Klein, "3d zernike descriptors for content based shape retrieval," in *Proceedings of the Eighth ACM Symposium on Solid Modeling and Applications*, ser. SM '03. New York, NY, USA: ACM, 2003, pp. 216–225. [Online]. Available: <http://doi.acm.org/10.1145/781606.781639>
- [125] B. Sofman, J. A. Bagnell, A. Stentz, and N. Vandapel, "Terrain classification from aerial data to support ground vehicle navigation," 2006.
- [126] S. K. Lodha, D. M. Fitzpatrick, and D. P. Helmbold, "Aerial lidar data classification using adaboost," in *Sixth International Conference on 3-D Digital Imaging and Modeling (3DIM 2007)*, Aug 2007, pp. 435–442.
- [127] N. Noceti, E. Delponte, and F. Odone, "Spatio-temporal constraints for on-line 3d object recognition in videos," *Computer Vision and Image Understanding*, vol. 113, no. 12, pp. 1198 – 1209, 2009, special issue on 3D Representation for Object and Scene Recognition. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S1077314209001088>
- [128] I. Atmosukarto and L. G. Shapiro, "A learning approach to 3d object representation for classification," in *Proceedings of the 2008 Joint IAPR International Workshop on Structural, Syntactic, and Statistical Pattern Recognition*, ser. SSPR & SPR '08. Berlin, Heidelberg: Springer-Verlag, 2008, pp. 267–276.
- [129] E. Hsiao, A. Collet, and M. Hebert, "Making specific features less discriminative to improve point-based 3d object recognition," in *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, June 2010, pp. 2653–2660.
- [130] I. Gordon and D. G. Lowe, *What and Where: 3D Object Recognition with Accurate Pose*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2006, pp. 67–82.
- [131] V. Lepetit, F. Moreno-Noguer, and P. Fua, "Eppnp: An accurate o(n) solution to the pnp problem," *International Journal of Computer Vision*, vol. 81, no. 2, p. 155, 2008. [Online]. Available: <http://dx.doi.org/10.1007/s11263-008-0152-6>
- [132] A. Collet, D. Berenson, S. S. Srinivasa, and D. Ferguson, "Object recognition and full pose registration from a single image for robotic manipulation," in *2009 IEEE International Conference on Robotics and Automation*, May 2009, pp. 48–55.
- [133] C. M. U. S. of Computer Science, "3d recognition dataset," <http://www.cs.cmu.edu/~ehsiao/3drecognition/>, acessado em 2017-04-07.
- [134] C. H. Kao, S. P. Hsieh, and C. C. Peng, "Study of feature-based image capturing and recognition algorithm," in *ICCA5 2010*, Oct 2010, pp. 1855–1861.
- [135] C. Xu and J. L. Prince, "Snakes, shapes, and gradient vector flow," *IEEE Transactions on Image Processing*, vol. 7, no. 3, pp. 359–369, Mar 1998.
- [136] B. Drost, M. Ulrich, N. Navab, and S. Ilic, "Model globally, match locally: Efficient and robust 3d object recognition," in *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, June 2010, pp. 998–1005.
- [137] R. B. Rusu, N. Blodow, and M. Beetz, "Fast point feature histograms (fpfh) for 3d registration," in *2009 IEEE International Conference on Robotics and Automation*, May 2009, pp. 3212–3217.
- [138] A. Mian, M. Bennamoun, and R. Owens, "On the repeatability and quality of keypoints for local feature-based 3d object retrieval from cluttered scenes," *International Journal of Computer Vision*, vol. 89, no. 2, pp. 348–361, 2010. [Online]. Available: <http://dx.doi.org/10.1007/s11263-009-0296-z>
- [139] Y. Wang, G. Sun, C. Wang, and D. Han, "Research on 3d object recognition from wire-frame based on edge moment," in *2010 3rd International Conference on Advanced Computer Theory and Engineering (ICACTE)*, vol. 1, Aug 2010, pp. V1–78–V1–82.
- [140] C.-C. Chen, "Improved moment invariants for shape discrimination," *Pattern Recognition*, vol. 26, no. 5, pp. 683 – 686, 1993. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/003132039390121C>
- [141] N. Naikal, A. Y. Yang, and S. S. Sastry, "Towards an efficient distributed object recognition system in wireless smart camera networks," in *2010 13th International Conference on Information Fusion*, July 2010, pp. 1–8.
- [142] D. Nister and H. Stewenius, "Scalable recognition with a vocabulary tree," in *Proceedings of the 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition - Volume 2*, ser. CVPR '06. Washington, DC, USA: IEEE Computer Society, 2006, pp. 2161–2168. [Online]. Available: <http://dx.doi.org/10.1109/CVPR.2006.264>
- [143] H. Bay, A. Ess, T. Tuytelaars, and L. V. Gool, "Speeded-up robust features (surf)," *Computer Vision and Image Understanding*, vol. 110, no. 3, pp. 346 – 359, 2008, similarity Matching in Computer Vision and Multimedia. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S1077314207001555>
- [144] V. Chandrasekhar, G. Takacs, D. M. Chen, S. S. Tsai, Y. Reznik, R. Grzeszczuk, and B. Girod, "Compressed histogram of gradients: A low-bitrate descriptor," *International Journal of Computer Vision*, vol. 96, no. 3, pp. 384–399, 2012. [Online]. Available: <http://dx.doi.org/10.1007/s11263-011-0453-z>
- [145] R. B. Rusu, G. Bradski, R. Thibaux, and J. Hsu, "Fast 3d recognition and pose using the viewpoint feature histogram," in *2010 IEEE/RSSJ International Conference on Intelligent Robots and Systems*, Oct 2010, pp. 2155–2162.
- [146] M. Muja and D. G. Lowe, "Fast approximate nearest neighbors with automatic algorithm configuration," in *In VISAPP International Conference on Computer Vision Theory and Applications*, 2009, pp. 331–340.
- [147] A. S. Mian, M. Bennamoun, and R. Owens, "Keypoint detection and local feature matching for textured 3d face recognition," *International Journal of Computer Vision*, vol. 79, no. 1, pp. 1–12, 2008. [Online]. Available: <http://dx.doi.org/10.1007/s11263-007-0085-5>
- [148] J. D'Erico, "Surface fitting using gridfit," <https://www.mathworks.com/matlabcentral/fileexchange/8998-surface-fitting-using-gridfit?requestedDomain=www.mathworks.com>, acessado em 2017-04-04.
- [149] F. Tombari, S. Salti, and L. Di Stefano, "Unique shape context for 3d data description," in *Proceedings of the ACM Workshop on 3D Object Retrieval*, ser. 3DOR '10. New York, NY, USA: ACM, 2010, pp. 57–62. [Online]. Available: <http://doi.acm.org/10.1145/1877808.1877821>
- [150] —, *Unique Signatures of Histograms for Local Surface Description*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2010, pp. 356–369.
- [151] S. University, "The stanford 3d scanning repository," <http://www.graphics.stanford.edu/data/3Dscanrep/>, acessado em 2017-02-08.
- [152] J. Knopp, M. Prasad, and L. Van Gool, "Orientation invariant 3d object classification using hough transform based methods," in *Proceedings of the ACM Workshop on 3D Object Retrieval*, ser. 3DOR '10. New York, NY, USA: ACM, 2010, pp. 15–20. [Online]. Available: <http://doi.acm.org/10.1145/1877808.1877813>
- [153] A. Bronstein, M. Bronstein, and R. Kimmel, *Numerical Geometry of Non-Rigid Shapes*, 1st ed. Springer Publishing Company, Incorporated, 2008.
- [154] R. C. Veltkamp and F. B. Ter Haar, "Shrec 2009 - shape retrieval contest," in *Proceedings of the 2Nd Eurographics Conference on 3D Object Retrieval*, ser. 3DOR '09. Aire-la-Ville, Switzerland, Switzerland: Eurographics Association, 2009, pp. 57–59. [Online]. Available: <http://dx.doi.org/10.2312/3DOR/3DOR09/057-059>
- [155] D. Šeatović, H. Kutterer, and T. Anken, "Automatic weed detection and treatment in grasslands," in *Proceedings ELMAR-2010*, Sept 2010, pp. 65–68.
- [156] D. Šeatović, "A segmentation approach in novel real time 3d plant recognition system," *Computer Vision Systems*, pp. 363–372, 2008.
- [157] —, "3d-object recognition, localization and treatment of rumex obtusifolius in its natural environment," in *1st International Conference on Machine Control and Guidance*, 2008.

- [158] J. C. Neto, G. E. Meyer, D. D. Jones, and A. K. Samal, "Plant species identification using elliptic fourier leaf shape analysis," *Computers and Electronics in Agriculture*, vol. 50, no. 2, pp. 121–134, 2006. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0168169905001560>
- [159] J. Zhou, S. Cadavid, and M. Abdel-Mottaleb, "Histograms of categorized shapes for 3d ear detection," in *2010 Fourth IEEE International Conference on Biometrics: Theory, Applications and Systems (BTAS)*, Sept 2010, pp. 1–6.
- [160] P. Viola and M. Jones, "Robust real-time object detection," *International Journal of Computer Vision*, vol. 4, no. 34–47, 2001.
- [161] N. Dalal, "Finding people in images and videos," Ph.D. dissertation, Institut National Polytechnique de Grenoble-INPG, 2006.
- [162] S. M. S. Islam, M. Bennamoun, and R. Davies, "Fast and fully automatic ear detection using cascaded adaboost," in *2008 IEEE Workshop on Applications of Computer Vision*, Jan 2008, pp. 1–6.
- [163] W. Wohlkinger and M. Vincze, "3d object classification for mobile robots in home-environments using web-data," in *19th International Workshop on Robotics in Alpe-Adria-Danube Region (RAAD 2010)*, June 2010, pp. 247–252.
- [164] K. Lai and D. Fox, "3d laser scan classification using web data and domain adaptation," in *Robotics: Science and Systems*, vol. 2, 2009.
- [165] T. Funkhouser, P. Min, M. Kazhdan, J. Chen, A. Halderman, D. Dobkin, and D. Jacobs, "A search engine for 3d models," *ACM Trans. Graph.*, vol. 22, no. 1, pp. 83–105, Jan. 2003. [Online]. Available: <http://doi.acm.org/10.1145/588272.588279>
- [166] W. Wohlkinger and M. Vincze, "Ensemble of shape functions for 3d object classification," in *2011 IEEE International Conference on Robotics and Biomimetics*, Dec 2011, pp. 2987–2992.
- [167] F. Tombari and L. Di Stefano, "Object recognition in 3d scenes with occlusions and clutter by hough voting," in *Proceedings of the 2010 Fourth Pacific-Rim Symposium on Image and Video Technology*, ser. PSIVT '10. Washington, DC, USA: IEEE Computer Society, 2010, pp. 349–355. [Online]. Available: <http://dx.doi.org/10.1109/PSIVT.2010.65>
- [168] Y. Owechko, S. Medasani, and T. Korah, "Automatic recognition of diverse 3-d objects and analysis of large urban scenes using ground and aerial lidar sensors," in *CLEO/QELS: 2010 Laser Science to Photonic Applications*, May 2010, pp. 1–2.
- [169] M. Unel, O. Soldea, E. Ozgur, and A. Bassa, "3d object recognition using invariants of 2d projection curves," *Pattern Anal. Appl.*, vol. 13, no. 4, pp. 451–468, Nov. 2010. [Online]. Available: <http://dx.doi.org/10.1007/s10044-010-0179-5>
- [170] C. F. I. Machines and S. of Computer Science McGill University, "McGill 3d shape benchmark," <http://www.cim.mcgill.ca/shape/benchMark/>, acessado em 2017-02-10.
- [171] J. Westell and P. Saeedi, "3d object recognition via multi-view inspection in unknown environments," in *2010 11th International Conference on Control Automation Robotics Vision*, Dec 2010, pp. 2088–2095.
- [172] B. V. Funt, K. Barnard, and L. Martin, "Is machine colour constancy good enough?" in *Proceedings of the 5th European Conference on Computer Vision-Volume I - Volume I*, ser. ECCV '98. London, UK, UK: Springer-Verlag, 1998, pp. 445–459. [Online]. Available: <http://dl.acm.org/citation.cfm?id=645311.649092>
- [173] J. Ayoub, B. Granado, O. Romain, and Y. Mhanna, "3-d object recognition based on svm and stereo-vision: Application in endoscopic imaging," in *2010 International Conference on Soft Computing and Pattern Recognition*, Dec 2010, pp. 198–201.
- [174] A. Kanazaki, T. Harada, and Y. Kuniyoshi, "Partial matching of real textured 3d objects using color cubic higher-order local auto-correlation features," *The Visual Computer*, vol. 26, no. 10, pp. 1269–1281, 2010. [Online]. Available: <http://dx.doi.org/10.1007/s00371-010-0521-3>
- [175] A. Kanazaki, H. Nakayama, T. Harada, and Y. Kuniyoshi, "High-speed 3d object recognition using additive features in a linear subspace," in *2010 IEEE International Conference on Robotics and Automation*, May 2010, pp. 3128–3134.
- [176] E. Oja, *Subspace Methods of Pattern Recognition*, ser. Electronic & Electrical Engineering Research Studies: Pattern Recognition and Image Processing Series. Research Studies Press, 1983, vol. 6.
- [177] G. M. Cortelazzo and N. Orío, "Retrieval of colored 3d models," in *3D Data Processing, Visualization, and Transmission, Third International Symposium on*, June 2006, pp. 986–993.
- [178] Y. Usui and K. Kondo, "3d object recognition based on confidence lut of sift feature distance," in *2010 Second World Congress on Nature and Biologically Inspired Computing (NaBIC)*, Dec 2010, pp. 293–297.
- [179] J.-M. Morel and G. Yu, "Asift: A new framework for fully affine invariant image comparison," *SIAM J. Img. Sci.*, vol. 2, no. 2, pp. 438–469, Apr. 2009. [Online]. Available: <http://dx.doi.org/10.1137/080732730>
- [180] E. Rodner, D. Hegazy, and J. Denzler, "Multiple kernel gaussian process classification for generic 3d object recognition," in *2010 25th International Conference of Image and Vision Computing New Zealand*, Nov 2010, pp. 1–8.
- [181] C. Rasmussen and C. Williams, "Gaussian processes for machine learning," *Gaussian Processes for Machine Learning*, 2006.
- [182] S. Lazebnik, C. Schmid, and J. Ponce, "Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories," in *2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06)*, vol. 2, 2006, pp. 2169–2178.
- [183] K. van de Sande, T. Gevers, and C. Snoek, "Evaluating color descriptors for object and scene recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 32, no. 9, pp. 1582–1596, Sept 2010.
- [184] A. Kapoor, K. Grauman, R. Urtasun, and T. Darrell, "Gaussian processes for object categorization," *International Journal of Computer Vision*, vol. 88, no. 2, pp. 169–188, 2010. [Online]. Available: <http://dx.doi.org/10.1007/s11263-009-0268-3>
- [185] D. Hegazy and J. Denzler, *Combining Appearance and Range Based Information for Multi-class Generic Object Recognition*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2009, pp. 741–748.
- [186] D. A. al-Kareem Mohammed Hegazy, "Boosting for generic 2d/3d object recognition," Ph.D. dissertation, Friedrich Schiller University of Jena, 2009. [Online]. Available: <http://www.db-thueringen.de/servlets/>
- [187] C. Papazov and D. Burschka, "An efficient ransac for 3d object recognition in noisy and occluded scenes," in *Proceedings of the 10th Asian Conference on Computer Vision - Volume Part I*, ser. ACCV'10. Berlin, Heidelberg: Springer-Verlag, 2011, pp. 135–148. [Online]. Available: <http://dl.acm.org/citation.cfm?id=1964320.1964334>
- [188] Y. Lamdan and H. J. Wolfson, "Geometric hashing: A general and efficient model-based recognition scheme," in *[1988 Proceedings] Second International Conference on Computer Vision*, Dec 1988, pp. 238–249.
- [189] C. Papazov, S. Haddadin, S. Parusel, K. Krieger, and D. Burschka, "Rigid 3d geometry matching for grasping of known objects in cluttered scenes," *Int. J. Rob. Res.*, vol. 31, no. 4, pp. 538–553, Apr. 2012. [Online]. Available: <http://dx.doi.org/10.1177/0278364911436019>
- [190] E. Akagunduz and I. Ulusoy, "3d object recognition from range images using transform invariant object representation," *Electronics Letters*, vol. 46, no. 22, pp. 1499–1500, October 2010.
- [191] S. Tangraumb, K. Takada, and O. Hasegawa, "3d object recognition using a voting algorithm in a real-world environment," in *Proceedings of the 2011 IEEE Workshop on Applications of Computer Vision (WACV)*, ser. WACV '11. Washington, DC, USA: IEEE Computer Society, 2011, pp. 153–158. [Online]. Available: <http://dx.doi.org/10.1109/WACV.2011.5711497>
- [192] J. Liebelt, C. Schmid, and K. Schertler, "Viewpoint-independent object class detection using 3d feature maps," in *2008 IEEE Conference on Computer Vision and Pattern Recognition*, June 2008, pp. 1–8.
- [193] H. Su, M. Sun, L. Fei-Fei, and S. Savarese, "Learning a dense multi-view representation for detection, viewpoint classification and synthesis of object categories," in *2009 IEEE 12th International Conference on Computer Vision*, Sept 2009, pp. 213–220.
- [194] M. Sun, H. Su, S. Savarese, and L. Fei-Fei, "A multi-view probabilistic model for 3d object classes," in *2009 IEEE Conference on Computer Vision and Pattern Recognition*, June 2009, pp. 1247–1254.
- [195] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, vol. 1, June 2005, pp. 886–893 vol. 1.
- [196] Y. Abramson, F. Moutarde, B. Stanculescu, and B. Steux, "Combining adaBoost with a Hill-Climbing evolutionary feature search for efficient training of performant visual object detector," in *FLINS2006 on Applied Artificial Intelligence*, Genova, Italy, Aug. 2006, pp. –. [Online]. Available: <https://hal.archives-ouvertes.fr/hal-00435700>
- [197] J. Shotton, J. Winn, C. Rother, and A. Criminisi, *TexonBoost: Joint Appearance, Shape and Context Modeling for Multi-class Object Recognition and Segmentation*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2006, pp. 1–15.
- [198] V. Viitaniemi and J. Laaksonen, *Techniques for Still Image Scene Classification and Object Detection*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2006, pp. 35–44.
- [199] M. Everingham, A. Zisserman, C. K. I. Williams, and L. Van Gool, "The pascal visual object classes challenge 2006," <http://>

- //host.robots.ox.ac.uk/pascal/VOC/voc2006/index.html, acessado em 2017-03-24.
- [200] E. Kim and G. Medioni, "3d object recognition in range images using visibility context," in *2011 IEEE/RSJ International Conference on Intelligent Robots and Systems*, Sept 2011, pp. 3800–3807.
- [201] W. Wohlkinger and M. Vincze, "Shape distributions on voxel surfaces for 3d object classification from depth images," in *2011 IEEE International Conference on Signal and Image Processing Applications (ICSIPA)*, Nov 2011, pp. 115–120.
- [202] Google, "Google warehouse," <https://3dwarehouse.sketchup.com/index.html>, acessado em 2017-04-08.
- [203] C. Fellbaum, *WordNet: An Electronic Lexical Database*, ser. Language, speech, and communication. MIT Press, 1998. [Online]. Available: <https://books.google.fr/books?id=Rehu8OzMMIMC>
- [204] W. Wohlkinger, A. Aldoma, R. B. Rusu, and M. Vincze, "3dnet: Large-scale object class recognition from cad models," in *2012 IEEE International Conference on Robotics and Automation*, May 2012, pp. 5384–5391.
- [205] A. Aldoma, M. Vincze, N. Blodow, D. Gossow, S. Gedikli, R. B. Rusu, and G. Bradski, "Cad-model recognition and 6dof pose estimation using 3d cues," in *2011 IEEE International Conference on Computer Vision Workshops (ICCV Workshops)*, Nov 2011, pp. 585–592.
- [206] D. Zarpalas, G. Kordelas, and P. Daras, "Recognizing 3d objects in cluttered scenes using projection images," in *2011 18th IEEE International Conference on Image Processing*, Sept 2011, pp. 673–676.
- [207] A. S. Mian, "3d model-based object recognition and segmentation in cluttered scene," <http://staf.fhome.ecm.uwa.edu.au/00053650/recognition.html>, acessado em 2017-04-09.
- [208] G. Kordelas and P. Daras, "Viewpoint independent object recognition in cluttered scenes exploiting ray-triangle intersection and {SIFT} algorithms," *Pattern Recognition*, vol. 43, no. 11, pp. 3833 – 3845, 2010. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S003132031000258X>
- [209] R. Osada, T. Funkhouser, B. Chazelle, and D. Dobkin, "Matching 3d models with shape distributions," in *Proceedings International Conference on Shape Modeling and Applications*, May 2001, pp. 154–166.
- [210] J. Zhou, S. Cadavid, and M. Abdel-Mottaleb, "A computationally efficient approach to 3d ear recognition employing local and holistic features," in *CVPR 2011 WORKSHOPS*, June 2011, pp. 98–105.
- [211] S. W. Wang and A. E. Kaufman, "Volume sampled voxelization of geometric primitives," in *Proceedings of the 4th Conference on Visualization '93*, ser. VIS '93. Washington, DC, USA: IEEE Computer Society, 1993, pp. 78–84. [Online]. Available: <http://dl.acm.org/citation.cfm?id=949845.949864>
- [212] A. A. Ross, K. Nandakumar, and A. K. Jain, *Handbook of Multibiometrics (International Series on Biometrics)*. Secaucus, NJ, USA: Springer-Verlag New York, Inc., 2006.
- [213] T. Theoharis, G. Passalis, G. Toderici, and I. A. Kakadiaris, "Unified 3d face and ear recognition using wavelets on geometry images," *Pattern Recognition*, vol. 41, no. 3, pp. 796 – 804, 2008, part Special issue: Feature Generation and Machine Learning for Robust Multimodal Biometrics. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0031320307003214>
- [214] P. Decker, S. Thierfelder, D. Paulus, and M. Grzegorzec, "Dense statistic versus sparse feature-based approach for 3d object recognition," *Pattern Recognit. Image Anal.*, vol. 21, no. 2, pp. 238–241, Jun. 2011. [Online]. Available: <http://dx.doi.org/10.1134/S1054661811020246>
- [215] M. Grzegorzec and H. Niemann, *Statistical Object Recognition Including Color Modeling*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2005, pp. 481–489.
- [216] R. Hanai, K. Yamazaki, H. Yaguchi, K. Okada, and M. Inaba, "Electric appliance parts classification using a measure combining the whole shape and local shape distribution similarities," in *Proceedings of the 2011 International Conference on 3D Imaging, Modeling, Processing, Visualization and Transmission*, ser. 3DIMPVT '11. Washington, DC, USA: IEEE Computer Society, 2011, pp. 296–303. [Online]. Available: <http://dx.doi.org/10.1109/3DIMPVT.2011.44>
- [217] C. B. Akgül, B. Sankur, Y. Yemez, and F. Schmitt, "3d model retrieval using probability density-based shape descriptors," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 31, no. 6, pp. 1117–1133, June 2009.
- [218] P. Papadakis, I. Pratikakis, S. Perantonis, and T. Theoharis, "Efficient 3d shape matching and retrieval using a concrete radialized spherical projection representation," *Pattern Recognition*, vol. 40, no. 9, pp. 2437 – 2452, 2007. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0031320307000106>
- [219] E. Wahl, U. Hillenbrand, and G. Hirzinger, "Surfllet-pair-relation histograms: a statistical 3d-shape representation for rapid classification," in *Fourth International Conference on 3-D Digital Imaging and Modeling, 2003. 3DIM 2003. Proceedings.*, Oct 2003, pp. 474–481.
- [220] S. Lee and Z. Lu, "Dependable 3d object recognition with two-layered particle filter," in *Proceedings of the 5th International Conference on Ubiquitous Information Management and Communication*, ser. ICUIMC '11. New York, NY, USA: ACM, 2011, pp. 37:1–37:8. [Online]. Available: <http://doi.acm.org/10.1145/1968613.1968658>
- [221] E. Kim and G. Medioni, "Scalable object classification using range images," in *Proceedings of the 2011 International Conference on 3D Imaging, Modeling, Processing, Visualization and Transmission*, ser. 3DIMPVT '11. Washington, DC, USA: IEEE Computer Society, 2011, pp. 65–72. [Online]. Available: <http://dx.doi.org/10.1109/3DIMPVT.2011.63>
- [222] E. Bart, I. Porteous, P. Perona, and M. Welling, "Unsupervised learning of visual taxonomies," in *2008 IEEE Conference on Computer Vision and Pattern Recognition*, June 2008, pp. 1–8.
- [223] H. Akbar, N. Suryana, and S. Sahib, "Training neural networks using clonal selection algorithm and particle swarm optimization: A comparisons for 3d object recognition," in *2011 11th International Conference on Hybrid Intelligent Systems (HIS)*, Dec 2011, pp. 692–697.
- [224] L. N. de Castro and F. J. V. Zuben, "Learning and optimization using the clonal selection principle," *IEEE Transactions on Evolutionary Computation*, vol. 6, no. 3, pp. 239–251, Jun 2002.
- [225] J. Kennedy, "Particle swarm optimization," in *Encyclopedia of machine learning*. Springer, 2011, pp. 760–766.
- [226] W. Jeong, S. Lee, and Y. Kim, "Statistical feature selection model for robust 3d object recognition," in *2011 15th International Conference on Advanced Robotics (ICAR)*, June 2011, pp. 402–408.
- [227] Z. Lu, S. Lee, and H. Kim, "Probabilistic 3d object recognition based on multiple interpretations generation," in *Proceedings of the 10th Asian Conference on Computer Vision - Volume Part IV*, ser. ACCV'10. Berlin, Heidelberg: Springer-Verlag, 2011, pp. 333–346. [Online]. Available: <http://dl.acm.org/citation.cfm?id=1966111.1966139>
- [228] J. Lee, S.-M. Baek, C. Choi, and S. Lee, *Particle Filter Based Robust Recognition and Pose Estimation of 3D Objects in a Sequence of Images*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2008, pp. 241–253.
- [229] A. Aldoma, F. Tombari, L. Di Stefano, and M. Vincze, "A global hypotheses verification method for 3d object recognition," in *Proceedings of the 12th European Conference on Computer Vision - Volume Part III*, ser. ECCV'12. Berlin, Heidelberg: Springer-Verlag, 2012, pp. 511–524.
- [230] A. Aldoma, F. Tombari, L. D. Stefano, and M. Vincze, "A global hypothesis verification framework for 3d object recognition in clutter," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 38, no. 7, pp. 1383–1396, July 2016.
- [231] K. S. Arun, T. S. Huang, and S. D. Blostein, "Least-squares fitting of two 3-d point sets," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. PAMI-9, no. 5, pp. 698–700, Sept 1987.
- [232] C. Papazov and D. Burschka, *An Efficient RANSAC for 3D Object Recognition in Noisy and Occluded Scenes*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2011, pp. 135–148.
- [233] T. Petříček and T. Svoboda, "Area-weighted surface normals for 3d object recognition," in *Proceedings of the 21st International Conference on Pattern Recognition (ICPR2012)*, Nov 2012, pp. 1492–1496.
- [234] X. Feng and P. Milanfar, "Multiscale principal components analysis for image local orientation estimation," in *Conference Record of the Thirty-Sixth Asilomar Conference on Signals, Systems and Computers, 2002.*, vol. 1, Nov 2002, pp. 478–482 vol.1.
- [235] R. Bro, E. Acar, and T. G. Kolda, "Resolving the sign ambiguity in the singular value decomposition," *Journal of Chemometrics*, vol. 22, no. 2, pp. 135–140, 2008. [Online]. Available: <http://dx.doi.org/10.1002/cem.1122>
- [236] A. Zaharescu, E. Boyer, K. Varanasi, and R. Horaud, "Surface feature detection and description with applications to mesh matching," in *2009 IEEE Conference on Computer Vision and Pattern Recognition*, June 2009, pp. 373–380.

- [237] P. Bariya and K. Nishino, "Scale-hierarchical 3d object recognition in cluttered scenes," in *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, June 2010, pp. 1657–1664.
- [238] H. Yabushita, J. Shimamura, and M. Morimoto, "A framework of three-dimensional object recognition which needs only a few reference images," in *Proceedings of the 21st International Conference on Pattern Recognition (ICPR2012)*, Nov 2012, pp. 1375–1378.
- [239] I. for Anthropomatics, R. Humanoids, and I. S. Laboratories, "The kit objectmodels web database," <https://h2t-projects.webarchiv.kit.edu/Projects/ObjectModelsWebUI/>, acessado em 2017-04-12.
- [240] M. Maeda, T. Nakamae, and K. Inoue, "Surface matching by curvature distribution images generated via gaze modeling," in *Proceedings of the 21st International Conference on Pattern Recognition (ICPR2012)*, Nov 2012, pp. 2194–2197.
- [241] C. Dorai and A. K. Jain, "Cosmos-a representation scheme for 3d free-form objects," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 19, no. 10, pp. 1115–1130, Oct. 1997. [Online]. Available: <http://dx.doi.org/10.1109/34.625113>
- [242] A. F. Sheta, A. Baareh, and M. Al-Batah, "3d object recognition using fuzzy mathematical modeling of 2d images," in *2012 International Conference on Multimedia Computing and Systems*, May 2012, pp. 278–283.
- [243] N. A. M. Isa, Z. M. Sani, and M. S. Al-Batah, "Automated intelligent real-time system for aggregate classification," *International Journal of Mineral Processing*, vol. 100, pp. 41 – 50, 2011. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S030175161100038X>
- [244] J. unic, K. Hirota, and P. L. Rosin, "A hu moment invariant as a shape circularity measure," *Pattern Recognition*, vol. 43, no. 1, pp. 47 – 57, 2010. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0031320309002660>
- [245] C.-L. Lim, B. Honarvar, K.-H. Thung, and R. Paramesran, "Fast computation of exact zernike moments using cascaded digital filters," *Information Sciences*, vol. 181, no. 17, pp. 3638 – 3651, 2011. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0020025511002039>
- [246] J. Liu, D. Li, W. Tao, and L. Yan, "An automatic method for generating affine moment invariants," *Pattern Recognition Letters*, vol. 28, no. 16, pp. 2295 – 2304, 2007. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0167865507002358>
- [247] T. Suk and J. Flusser, "Affine moment invariants generated by graph method," *Pattern Recognition*, vol. 44, no. 9, pp. 2047 – 2056, 2011, computer Analysis of Images and Patterns. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0031320310002359>
- [248] R. Babuska, J. A. Roubos, and H. B. Verbruggen, "Identification of mimo systems by input-output ts fuzzy models," in *1998 IEEE International Conference on Fuzzy Systems Proceedings. IEEE World Congress on Computational Intelligence (Cat. No.98CH36228)*, vol. 1, May 1998, pp. 657–662 vol.1.
- [249] P. Bongale, A. Ranjan, and S. Anand, "Implementation of 3d object recognition and tracking," in *2012 International Conference on Recent Advances in Computing and Software Systems*, April 2012, pp. 77–79.
- [250] P. Bariya, J. Novatnack, G. Schwartz, and K. Nishino, "3d geometric scale variability in range images: Features and descriptors," *Int. J. Comput. Vision*, vol. 99, no. 2, pp. 232–255, Sep. 2012. [Online]. Available: <http://dx.doi.org/10.1007/s11263-012-0526-7>
- [251] W. E. L. Grimson, *Object Recognition by Computer: The Role of Geometric Constraints*. Cambridge, MA, USA: MIT Press, 1990.
- [252] P. J. Flynn and A. K. Jain, "Bonsai: 3d object recognition using constrained search," in *[1990] Proceedings Third International Conference on Computer Vision*, Dec 1990, pp. 263–267.
- [253] B. Taati, M. Bondy, P. Jasiobedzki, and M. Greenspan, "Variable dimensional local shape descriptors for object recognition in range data," in *2007 IEEE 11th International Conference on Computer Vision*, Oct 2007, pp. 1–8.
- [254] R. Socher, B. Huval, B. Bhat, C. D. Manning, and A. Y. Ng, "Convolutional-recursive deep learning for 3d object classification," in *Proceedings of the 25th International Conference on Neural Information Processing Systems*, ser. NIPS'12. USA: Curran Associates Inc., 2012, pp. 656–664. [Online]. Available: <http://dl.acm.org/citation.cfm?id=2999134.2999208>
- [255] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," in *Proceedings of the IEEE*, 1998, pp. 2278–2324.
- [256] R. Socher, C. C. Lin, A. Y. Ng, and C. D. Manning, "Parsing Natural Scenes and Natural Language with Recursive Neural Networks," in *Proceedings of the 26th International Conference on Machine Learning (ICML)*, 2011.
- [257] K. Lai, L. Bo, X. Ren, and D. Fox, "Rgb-d object dataset," <http://rgbd-dataset.cs.washington.edu/dataset.html>, acessado em 2017-01-25.
- [258] H. Ding, X. Li, H. Zhao, and W. Xiao, "A new generalized affine moment invariants for shape retrieval and object recognition," in *2012 8th IEEE International Symposium on Instrumentation and Control Technology (ISICT) Proceedings*, July 2012, pp. 137–142.
- [259] J. Flusser and T. Suk, "A moment-based approach to registration of images with affine geometric distortion," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 32, no. 2, pp. 382–387, Mar 1994.
- [260] E. Rahtu, M. Salo, and J. Heikkila, "Affine invariant pattern recognition using multiscale autoconvolution," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, no. 6, pp. 908–918, June 2005.
- [261] Y. Jang and W. Woo, "Local feature descriptors for 3d object recognition in ubiquitous virtual reality," in *Proceedings of the 2012 International Symposium on Ubiquitous Virtual Reality*, ser. ISUVR '12. Washington, DC, USA: IEEE Computer Society, 2012, pp. 42–45. [Online]. Available: <http://dx.doi.org/10.1109/ISUVR.2012.20>
- [262] Y.-J. Liu, Q.-F. Fu, Y. Liu, and X.-L. Fu, "2d-line-drawing-based 3d object recognition," in *Proceedings of the First International Conference on Computational Visual Media*, ser. CVM'12. Berlin, Heidelberg: Springer-Verlag, 2012, pp. 146–153.
- [263] H. Kang, S. Lee, and C. K. Chui, "Coherent line drawing," in *Proceedings of the 5th International Symposium on Non-photorealistic Animation and Rendering*, ser. NPAR '07. New York, NY, USA: ACM, 2007, pp. 43–50. [Online]. Available: <http://doi.acm.org/10.1145/1274871.1274878>
- [264] W. Ping, W. Wei, G. Ying-hui, and L. Shi-fei, "A recognition approach of 3-d objects based on the tsallis entropy," in *2012 International Conference on Computer Vision in Remote Sensing*, Dec 2012, pp. 242–245.
- [265] H. Q. Dinh and S. Kropac, "Multi-resolution spin-images," in *2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06)*, vol. 1, June 2006, pp. 863–870.
- [266] Y. Wang, G. Pan, Z. Wu, and S. Han, *Sphere-Spin-Image: A Viewpoint-Invariant Surface Representation for 3D Face Recognition*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2004, pp. 427–434.
- [267] O. Bencharef, M. Fakir, B. Minaoui, A. Hajraoui, and M. Oujaoura, "Color objects recognition system based on artificial neural network with zernike, hu amp; geodesic descriptors," in *2012 6th International Conference on Sciences of Electronics, Technologies of Information and Telecommunications (SETIT)*, March 2012, pp. 338–343.
- [268] L. C. Chen, X. L. Nguyen, and S. T. Lin, "Automatic object detection employing viewing angle histogram for range images," in *2012 IEEE/ASME International Conference on Advanced Intelligent Mechatronics (AIM)*, July 2012, pp. 196–201.
- [269] —, "Automatic object detection employing viewing angle histogram for range images," in *2012 IEEE/ASME International Conference on Advanced Intelligent Mechatronics (AIM)*, July 2012, pp. 196–201.
- [270] N. Megherbi, J. Han, T. P. Breckon, and G. T. Flitton, "A comparison of classification approaches for threat detection in ct based baggage screening," in *2012 19th IEEE International Conference on Image Processing*, Sept 2012, pp. 3109–3112.
- [271] J. K. Udupa and S. Samarasekera, "Fuzzy connectedness and object definition: Theory, algorithms, and applications in image segmentation," *Graphical Models and Image Processing*, vol. 58, no. 3, pp. 246 – 261, 1996. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S1077316996900210>
- [272] M. Novotni and R. Klein, "Shape retrieval using 3d zernike descriptors," *Computer-Aided Design*, vol. 36, no. 11, pp. 1047 – 1062, 2004, solid Modeling Theory and Applications. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0010448504000077>
- [273] C. wei Hsu, C. chung Chang, and C. jen Lin, "A practical guide to support vector classification," 2010.
- [274] C. M. Bishop, *Neural Networks for Pattern Recognition*. New York, NY, USA: Oxford University Press, Inc., 1995.
- [275] L. Breiman, J. Friedman, R. Olshen, and C. Stone, *Classification and Regression Trees*. Monterey, CA: Wadsworth and Brooks, 1984.
- [276] H. Drucker and C. Cortes, "Boosting decision trees," in *Proceedings of the 8th International Conference on Neural Information Processing Systems*. MIT Press, 1995, pp. 479–485.

- [277] L. Breiman, "Random forests," *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001. [Online]. Available: <http://dx.doi.org/10.1023/A:1010933404324>
- [278] G. Flitton, T. Breckon, and N. Megherbi Bouallagu, "Object recognition using 3d sift in complex ct volumes," in *Proceedings of the British Machine Vision Conference*. BMVA Press, 2010, pp. 11.1–11.12, doi:10.5244/C.24.11.
- [279] N. Megherbi, G. T. Flitton, and T. P. Breckon, "A classifier based approach for the detection of potential threats in ct based baggage screening," in *2010 IEEE International Conference on Image Processing*, Sept 2010, pp. 1833–1836.
- [280] A. Shaiek and F. Moutarde, "Fast 3d keypoints detector and descriptor for view-based 3d objects recognition," in *Revised Selected and Invited Papers of the International Workshop on Advances in Depth Image Analysis and Applications - Volume 7854*. New York, NY, USA: Springer-Verlag New York, Inc., 2013, pp. 106–115.
- [281] E. Rodolà, A. Albarelli, F. Bergamasco, and A. Torsello, "A scale independent selection process for 3d object recognition in cluttered scenes," *Int. J. Comput. Vision*, vol. 102, no. 1-3, pp. 129–145, Mar. 2013. [Online]. Available: <http://dx.doi.org/10.1007/s11263-012-0568-x>
- [282] A. Albarelli, E. Rodolà, F. Bergamasco, and A. Torsello, "A non-cooperative game for 3d object recognition in cluttered scenes," in *2011 International Conference on 3D Imaging, Modeling, Processing, Visualization and Transmission*, May 2011, pp. 252–259.
- [283] A. Albarelli, E. Rodolà, and A. Torsello, "A game-theoretic approach to fine surface registration without initial motion estimation," in *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, June 2010, pp. 430–437.
- [284] M. Keaikitse, N. Govender, and J. Warrell, "Comparison of active sift-based 3d object recognition algorithms," in *2013 Africon*, Sept 2013, pp. 1–5.
- [285] G. Kootstra, J. Ypma, and B. de Boer, "Active exploration and keypoint clustering for object recognition," in *Robotics and Automation, 2008. ICRA 2008. IEEE International Conference on*. IEEE, 2008, pp. 1005–1010.
- [286] N. Govender, J. Claassens, F. Nicolls, and J. Warrell, "Active object recognition using vocabulary trees," in *2013 IEEE Workshop on Robot Vision (WORV)*, Jan 2013, pp. 20–26.
- [287] D. Nistér and H. Stewénius, "Scalable recognition with a vocabulary tree," in *IN CVPR*, 2006, pp. 2161–2168.
- [288] F. Oleari, D. L. Rizzini, and S. Caselli, "A low-cost stereo system for 3d object recognition," in *2013 IEEE 9th International Conference on Intelligent Computer Communication and Processing (ICCP)*, Sept 2013, pp. 127–132.
- [289] M. Bennamoun, F. A. Sohel, Y. Guo, M. Lu, and J. Wan, "3d free form object recognition using rotational projection statistics," in *Proceedings of the 2013 IEEE Workshop on Applications of Computer Vision (WACV)*, ser. WACV '13. Washington, DC, USA: IEEE Computer Society, 2013, pp. 1–8. [Online]. Available: <http://dx.doi.org/10.1109/WACV.2013.6474992>
- [290] U. Castellani, M. Cristani, S. Fantoni, and V. Murino, "Sparse points matching by combining 3D mesh saliency with statistical descriptors," *Computer Graphics Forum*, 2008.
- [291] G. Hetzel, B. Leibe, P. Levi, and B. Schiele, "3d object recognition from range images using local feature histograms," in *Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR 2001*, vol. 2, 2001, pp. II–394–II–399 vol.2.
- [292] A. Flint, A. Dick, and A. v. d. Hengel, "Thrift: Local 3d structure recognition," in *9th Biennial Conference of the Australian Pattern Recognition Society on Digital Image Computing Techniques and Applications (DICTA 2007)*, Dec 2007, pp. 182–188.
- [293] H. Yabushita, T. Osawa, J. Shimamura, and Y. Taniguchi, "Mobile visual search for 3-d objects: Matching user-captured video to single reference image," in *2013 IEEE 2nd Global Conference on Consumer Electronics (GCCE)*, Oct 2013, pp. 122–123.
- [294] B. s. Kim, S. Xu, and S. Savarese, "Accurate localization of 3d objects from rgb-d data using segmentation hypotheses," in *2013 IEEE Conference on Computer Vision and Pattern Recognition*, June 2013, pp. 3182–3189.
- [295] P. F. Felzenszwalb, R. B. Girshick, D. McAllester, and D. Ramanan, "Object detection with discriminatively trained part-based models," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 32, no. 9, pp. 1627–1645, Sept 2010.
- [296] Y. Xiang and S. Savarese, "Estimating the aspect layout of object categories," in *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*. IEEE, 2012, pp. 3410–3417.
- [297] B. Pepik, M. Stark, P. Gehler, and B. Schiele, "Teaching 3d geometry to deformable part models," in *2012 IEEE Conference on Computer Vision and Pattern Recognition*, June 2012, pp. 3362–3369.
- [298] T. Joachims, T. Finley, and C.-N. J. Yu, "Cutting-plane training of structural svms," *Machine Learning*, vol. 77, no. 1, pp. 27–59, 2009. [Online]. Available: <http://dx.doi.org/10.1007/s10994-009-5108-8>
- [299] K. Lai, L. Bo, X. Ren, and D. Fox, "A large-scale hierarchical multi-view rgb-d object dataset," in *2011 IEEE International Conference on Robotics and Automation*, May 2011, pp. 1817–1824.
- [300] A. Janoch, S. Karayev, Y. Jia, J. T. Barron, M. Fritz, K. Saenko, and T. Darrell, "A category-level 3-d object dataset: Putting the kinect to work," in *2011 IEEE International Conference on Computer Vision Workshops (ICCV Workshops)*, Nov 2011, pp. 1168–1174.
- [301] K. S. Choi and D. H. Kim, "Angular-partitioned spin image descriptor for robust 3d facial landmark detection," *Electronics Letters*, vol. 49, no. 23, pp. 1454–1455, Nov 2013.
- [302] A. B. Moreno and A. Sánchez, "GavabDB: a 3D Face Database," in *Workshop on Biometrics on the Internet*, Vigo, Mar. 2004, pp. 77–85.
- [303] D. Palossi, F. Tombari, S. Salti, M. Ruggiero, L. D. Stefano, and L. Benini, "Gpu-shot: Parallel optimization for real-time 3d local description," in *2013 IEEE Conference on Computer Vision and Pattern Recognition Workshops*, June 2013, pp. 584–591.
- [304] F. Tombari, S. Salti, and L. Di Stefano, "Rgb-d object recognition and localization with clutter and occlusions," in *Proc. RGB-D Workshop on 3D Perception in Robotics in conj. with euRobotics Forum*, vol. 6, 2011.
- [305] M. Abdelrahman, A. A. Farag, and M. El-Melehy, "Heat front propagation contours for 3d face recognition," in *2013 IEEE Sixth International Conference on Biometrics: Theory, Applications and Systems (BTAS)*, Sept 2013, pp. 1–6.
- [306] M. Everingham, J. Sivic, and A. Zisserman, "Hello! my name is... buffy – automatic naming of characters in tv video," in *In BMVC*, 2006.
- [307] S. Gupta, M. K. Markey, and A. C. Bovik, "Anthropometric 3d face recognition," *International Journal of Computer Vision*, vol. 90, no. 3, pp. 331–349, 2010. [Online]. Available: <http://dx.doi.org/10.1007/s11263-010-0360-8>
- [308] C. BenAbdelkader and P. A. Griffin, "Comparing and combining depth and texture cues for face recognition," *Image and Vision Computing*, vol. 23, no. 3, pp. 339 – 352, 2005. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0262885604001945>
- [309] X. Lu, A. K. Jain, and D. Colbry, "Matching 2.5d face scans to 3d models," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 28, no. 1, pp. 31–43, Jan 2006.
- [310] J. Yu, K. Weng, G. Liang, and G. Xie, "A vision-based robotic grasping system using deep learning for 3d object recognition and pose estimation," in *IEEE International Conference on Robotics and Biomimetics, ROBIO 2013, Shenzhen, China, December 12-14, 2013*, 2013, pp. 1175–1180. [Online]. Available: <http://dx.doi.org/10.1109/ROBIO.2013.6739623>
- [311] J. Lam and M. Greenspan, "3d object recognition by surface registration of interest segments," in *Proceedings of the 2013 International Conference on 3D Vision*, ser. 3DV '13. Washington, DC, USA: IEEE Computer Society, 2013, pp. 199–206. [Online]. Available: <http://dx.doi.org/10.1109/3DV.2013.34>
- [312] D. Aiger, N. J. Mitra, and D. Cohen-Or, "4pointss congruent sets for robust pairwise surface registration," in *ACM SIGGRAPH 2008 Papers*, ser. SIGGRAPH '08. New York, NY, USA: ACM, 2008, pp. 85:1–85:10. [Online]. Available: <http://doi.acm.org/10.1145/1399504.1360684>
- [313] J. Ekekrantz, A. Pronobis, J. Folkesson, and P. Jensfelt, "Adaptive iterative closest keypoint," in *2013 European Conference on Mobile Robots*, Sept 2013, pp. 80–87.
- [314] E. Rublee, V. Rabaud, K. Konolige, and G. Bradski, "Orb: An efficient alternative to sift or surf," in *Proceedings of the 2011 International Conference on Computer Vision*, ser. ICCV '11. Washington, DC, USA: IEEE Computer Society, 2011, pp. 2564–2571. [Online]. Available: <http://dx.doi.org/10.1109/ICCV.2011.6126544>
- [315] J. Sturm, N. Engelhard, F. Endres, W. Burgard, and D. Cremers, "A benchmark for the evaluation of rgb-d slam systems," in *2012 IEEE/RSJ International Conference on Intelligent Robots and Systems*, Oct 2012, pp. 573–580.
- [316] A. Segal, D. Haehnel, and S. Thrun, "Generalized-icp," in *Robotics: science and systems*, 2009.
- [317] M. Magnusson, A. Lilienthal, and T. Duckett, "Scan registration for autonomous mining vehicles using 3d-ndt," *Journal of Field Robotics*, vol. 24, no. 10, pp. 803–827, 2007.

- [318] M. A. Fischler and R. C. Bolles, "Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography," *Commun. ACM*, vol. 24, no. 6, pp. 381–395, Jun. 1981. [Online]. Available: <http://doi.acm.org/10.1145/358669.358692>
- [319] R. B. Gomes, B. M. F. da Silva, L. K. de Medeiros Rocha, R. V. Aroca, L. C. P. R. Velho, and L. M. G. Gonçalves, "Efficient 3d object recognition using foveated point clouds," *Computers & Graphics*, vol. 37, no. 5, pp. 496 – 508, 2013. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0097849313000459>
- [320] A. Petrelli and L. D. Stefano, "On the repeatability of the local reference frame for partial shape matching," in *2011 International Conference on Computer Vision*, Nov 2011, pp. 2244–2251.
- [321] Y. Guo, F. A. Soheli, M. Bennamoun, J. Wan, and M. Lu, "Integrating shape and color cues for textured 3d object recognition," in *2013 IEEE 8th Conference on Industrial Electronics and Applications (ICIEA)*, June 2013, pp. 1614–1619.
- [322] —, "Rops: A local feature descriptor for 3d rigid objects based on rotational projection statistics," in *2013 1st International Conference on Communications, Signal Processing, and their Applications (ICCSPA)*, Feb 2013, pp. 1–6.
- [323] Y. Guo, F. Soheli, M. Bennamoun, M. Lu, and J. Wan, "Rotational projection statistics for 3d local surface description and object recognition," *International journal of computer vision*, vol. 105, no. 1, pp. 63–86, 2013.
- [324] F. Tombari, S. Salti, and L. D. Stefano, "A combined texture-shape descriptor for enhanced 3d feature matching," in *2011 18th IEEE International Conference on Image Processing*, Sept 2011, pp. 809–812.
- [325] S. Ma, C. Zhou, L. Zhang, W. Hong, and Y. Tian, "3d object recognition using kernel pca based on depth information for twist-lock grasping," in *2013 IEEE International Conference on Robotics and Biomimetics (ROBIO)*, Dec 2013, pp. 2667–2672.
- [326] L. Bo, X. Ren, and D. Fox, "Kernel descriptors for visual recognition," in *Advances in neural information processing systems*, 2010, pp. 244–252.
- [327] L. Bo, K. Lai, X. Ren, and D. Fox, "Object recognition with hierarchical kernel descriptors," in *CVPR 2011*, June 2011, pp. 1729–1736.
- [328] G. Jing, Z. Mingquan, and L. Chao, "3d object classification based on local keywords and hidden markov model," in *2013 Fourth International Conference on Digital Manufacturing Automation*, June 2013, pp. 1–4.
- [329] L. Xiao-ming, Y. Jian-wei, F. Zhi-lin, and D. Jin-xiang, "3d model classification based on adaptive-weighted asymmetric adaboost hidden markov models," in *Journal of Zhejiang University*, 2006, pp. 1300–1305.
- [330] R. F. Salas-Moreno, R. A. Newcombe, H. Strasdat, P. H. J. Kelly, and A. J. Davison, "Slam++: Simultaneous localisation and mapping at the level of objects," in *Proceedings of the 2013 IEEE Conference on Computer Vision and Pattern Recognition*, ser. CVPR '13. Washington, DC, USA: IEEE Computer Society, 2013, pp. 1352–1359. [Online]. Available: <http://dx.doi.org/10.1109/CVPR.2013.178>
- [331] G. Pang and U. Neumann, "Training-based object recognition in cluttered 3d point clouds," in *Proceedings of the 2013 International Conference on 3D Vision*, ser. 3DV '13. Washington, DC, USA: IEEE Computer Society, 2013, pp. 87–94. [Online]. Available: <http://dx.doi.org/10.1109/3DV.2013.20>
- [332] Y. Freund and R. E. Schapire, "A decision-theoretic generalization of on-line learning and an application to boosting," *Journal of Computer and System Sciences*, vol. 55, no. 1, pp. 119 – 139, 1997. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S00220009791504X>
- [333] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," in *Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR 2001*, vol. 1, 2001, pp. 1–511–1–518 vol.1.
- [334] T. Nakashika, T. Hori, T. Takiguchi, and Y. Ariki, "3d-object recognition based on llc using depth spatial pyramid," in *Proceedings of the 2014 22Nd International Conference on Pattern Recognition*, ser. ICPR '14. Washington, DC, USA: IEEE Computer Society, 2014, pp. 4224–4228. [Online]. Available: <http://dx.doi.org/10.1109/ICPR.2014.724>
- [335] S. Takei, S. Akizuki, and M. Hashimoto, "3d object recognition using effective features selected by evaluating performance of discrimination," in *2014 13th International Conference on Control Automation Robotics Vision (ICARCV)*, Dec 2014, pp. 70–75.
- [336] O. Maron and T. Lozano-Pérez, "A framework for multiple-instance learning," in *Proceedings of the 1997 Conference on Advances in Neural Information Processing Systems 10*, ser. NIPS '97. Cambridge, MA, USA: MIT Press, 1998, pp. 570–576. [Online]. Available: <http://dl.acm.org/citation.cfm?id=302528.302753>
- [337] S. Akizuki and M. Hashimoto, "Fast and reliable 3-d object recognition based on surface normal distributions," in *Proc. International Symposium on Optomechatronic Technologies*, 2013, pp. 1–9.
- [338] A. Mouton, T. P. Breckon, G. T. Flitton, and N. Megherbi, "3d object classification in baggage computed tomography imagery using randomised clustering forests," in *2014 IEEE International Conference on Image Processing (ICIP)*, Oct 2014, pp. 5202–5206.
- [339] E. Nowak, F. Jurie, and B. Triggs, *Sampling Strategies for Bag-of-Features Image Classification*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2006, pp. 490–503.
- [340] G. Flitton, T. P. Breckon, and N. Megherbi, "A 3d extension to cortex like mechanisms for 3d object class recognition," in *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*. IEEE, 2012, pp. 3634–3641.
- [341] F. Moosmann, B. Triggs, F. Jurie *et al.*, "Fast discriminative visual codebooks using randomized clustering forests," in *NIPS*, vol. 2, 2006, p. 4.
- [342] X. Qian and C. Ye, "3d object recognition by geometric context and gaussian-mixture-model-based plane classification," in *2014 IEEE International Conference on Robotics and Automation (ICRA)*, May 2014, pp. 3910–3915.
- [343] —, "Ncc-ransac: A fast plane extraction method for 3-d range data segmentation," *IEEE Transactions on Cybernetics*, vol. 44, no. 12, pp. 2771–2783, Dec 2014.
- [344] C. M. Mateo, P. Gil, and F. Torres, "A performance evaluation of surface normals-based descriptors for recognition of objects using cad-models," in *2014 11th International Conference on Informatics in Control, Automation and Robotics (ICINCO)*, vol. 02, Sept 2014, pp. 428–435.
- [345] R. B. Rusu, "Semantic 3d object maps for everyday manipulation in human living environments," *KI - Künstliche Intelligenz*, vol. 24, no. 4, pp. 345–348, 2010. [Online]. Available: <http://dx.doi.org/10.1007/s13218-010-0059-6>
- [346] S. H. Lee, S. C. Cheng, and C. C. Chang, "Moment-preserving techniques for 3d shape registration and recognition," in *2014 International Symposium on Computer, Consumer and Control*, June 2014, pp. 516–519.
- [347] J. C. Niebles, H. Wang, and L. Fei-Fei, "Unsupervised learning of human action categories using spatial-temporal words," *Int. J. Comput. Vision*, vol. 79, no. 3, pp. 299–318, Sep. 2008. [Online]. Available: <http://dx.doi.org/10.1007/s11263-007-0122-4>
- [348] Y. Guo, F. Soheli, M. Bennamoun, J. Wan, and M. Lu, "An accurate and robust range image registration algorithm for 3d object modeling," *IEEE Transactions on Multimedia*, vol. 16, no. 5, pp. 1377–1390, Aug 2014.
- [349] D. S. Alexiadis, D. Zarpalas, and P. Daras, "Real-time, full 3-d reconstruction of moving foreground objects from multiple consumer depth cameras," *IEEE Transactions on Multimedia*, vol. 15, no. 2, pp. 339–358, Feb 2013.
- [350] J. Williams and M. Bennamoun, "Simultaneous registration of multiple corresponding point sets," *Computer Vision and Image Understanding*, vol. 81, no. 1, pp. 117 – 142, 2001. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S1077314200908841>
- [351] B. Curless and M. Levoy, "A volumetric method for building complex models from range images," in *Proceedings of the 23rd Annual Conference on Computer Graphics and Interactive Techniques*, ser. SIGGRAPH '96. New York, NY, USA: ACM, 1996, pp. 303–312. [Online]. Available: <http://doi.acm.org/10.1145/237170.237269>
- [352] T. Masuda, "Log-polar height maps for multiple range image registration," *Computer Vision and Image Understanding*, vol. 113, no. 11, pp. 1158 – 1169, 2009, special issue on New Advances in 3D Imaging and Modeling. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S1077314209001015>
- [353] Y. Salih, A. S. Malik, D. Sidibé, M. T. Simsim, N. Saad, and F. Meriaudeau, "Compressed vfh descriptor for 3d object classification," in *2014 3DTV-Conference: The True Vision - Capture, Transmission and Display of 3D Video (3DTV-CON)*, July 2014, pp. 1–4.
- [354] X. Yu, Y. Gao, and J. Zhou, "Face recognition using 3d directional corner points," in *2014 22nd International Conference on Pattern Recognition*, Aug 2014, pp. 2802–2807.
- [355] D. H. Douglas and T. K. Peucker, "Algorithms for the reduction of the number of points required to represent a digitized line or its caricature,"

- Cartographica: The International Journal for Geographic Information and Geovisualization*, vol. 10, no. 2, pp. 112–122, 1973.
- [356] M. H. Mahoor and M. Abdel-Mottaleb, “Face recognition based on 3d ridge images obtained from range data,” *Pattern Recognition*, vol. 42, no. 3, pp. 445 – 451, 2009. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S003132030800321X>
- [357] T. Ejima, S. Enokida, T. Kouno, H. Ideguchi, and T. Horiuchi, “3d object recognition based on the reference point ensemble,” in *2014 International Conference on Computer Vision Theory and Applications (VISAPP)*, vol. 3, Jan 2014, pp. 261–269.
- [358] D. Liang, K. Weng, C. Wang, G. Liang, H. Chen, and X. Wu, “A 3d object recognition and pose estimation system using deep learning method,” in *2014 4th IEEE International Conference on Information Science and Technology*, April 2014, pp. 401–404.
- [359] L. Carrer and A. G. Yarovoy, “Concealed weapon detection using uwb 3-d radar imaging and automatic target recognition,” in *The 8th European Conference on Antennas and Propagation (EuCAP 2014)*, April 2014, pp. 2786–2790.
- [360] K. Zuiderveld, “Graphics gems iv,” P. S. Heckbert, Ed. San Diego, CA, USA: Academic Press Professional, Inc., 1994, ch. Contrast Limited Adaptive Histogram Equalization, pp. 474–485. [Online]. Available: <http://dl.acm.org/citation.cfm?id=180895.180940>
- [361] G. M. M. P. Paul, and M. R. Kaimal, “An improved content based image retrieval in rgbd images using point clouds,” in *2014 International Conference on Communication and Signal Processing*, April 2014, pp. 828–832.
- [362] K. Demarsin, D. Vanderstraeten, T. Volodine, and D. Roose, “Detection of closed sharp edges in point clouds using normal estimation and graph theory,” *Computer-Aided Design*, vol. 39, no. 4, pp. 276 – 283, 2007. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0010448506002260>
- [363] R. Schnabel, R. Wessel, R. Wahl, and R. Klein, “Shape recognition in 3d point-clouds,” in *The 16-th International Conference in Central Europe on Computer Graphics, Visualization and Computer Vision*, vol. 8. Citeseer, 2008.
- [364] Y. Guo, F. Sohnel, M. Bennamoun, J. Wan, and M. Lu, “A novel local surface feature for 3d object recognition under clutter and occlusion,” *Information Sciences*, vol. 293, pp. 196 – 213, 2015. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0020025514009219>
- [365] R. P. de Figueiredo, P. Moreno, and A. Bernardino, “Efficient pose estimation of rotationally symmetric objects,” *Neurocomputing*, vol. 150, Part A, pp. 126 – 135, 2015, bioinspired and knowledge based techniques and applications The Vitality of Pattern Recognition and Image Analysis Data Stream Classification and Big Data Analytics Selected papers from the 16th International Conference on Knowledge-Based and Intelligent Information; Engineering Systems (KES 2012) Selected papers from the 6th Iberian Conference on Pattern Recognition and Image Analysis (IbPRIA 2013). [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S092523121401265X>
- [366] A. Torralba, K. P. Murphy, and W. T. Freeman, “Sharing visual features for multiclass and multiview object detection,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 29, no. 5, pp. 854–869, May 2007.
- [367] Y. Zhuang, X. Lin, H. Hu, and G. Guo, “Using scale coordination and semantic information for robust 3-d object recognition by a service robot,” *IEEE Sensors Journal*, vol. 15, no. 1, pp. 37–47, Jan 2015.
- [368] T. J. M. Sanguino and F. P. Gómez, “Improving 3d object detection and classification based on kinect sensor and hough transform,” in *2015 International Symposium on Innovations in Intelligent Systems and Applications (INISTA)*, Sept 2015, pp. 1–8.
- [369] F. Ribeiro, S. Brandão, J. P. Costeira, and M. Veloso, “Global localization by soft object recognition from 3d partial views,” in *2015 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, Sept 2015, pp. 3709–3714.
- [370] S. Brandão, J. P. Costeira, and M. Veloso, “The partial view heat kernel descriptor for 3d object representation,” in *2014 IEEE International Conference on Robotics and Automation (ICRA)*, May 2014, pp. 1054–1059.
- [371] S. Brandão, M. Veloso, and J. P. Costeira, “Multiple hypothesis for object class disambiguation from multiple observations,” in *2014 2nd International Conference on 3D Vision*, vol. 1, Dec 2014, pp. 91–98.
- [372] A. J. Rodríguez-Sánchez, S. Szedmak, and J. Piater, “Scurv: A 3d descriptor for object classification,” in *2015 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, Sept 2015, pp. 1320–1327.
- [373] A. Aldoma, F. Tombari, R. B. Rusu, and M. Vincze, *OUR-CVFH – Oriented, Unique and Repeatable Clustered Viewpoint Feature Histogram for Object Recognition and 6DOF Pose Estimation*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2012, pp. 113–122.
- [374] R. B. Rusu, Z. C. Marton, N. Blodow, M. Dolha, and M. Beetz, “Towards 3d point cloud based object maps for household environments,” *Robotics and Autonomous Systems*, vol. 56, no. 11, pp. 927 – 941, 2008, semantic Knowledge in Robotics. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0921889008001140>
- [375] M. Körtgen, M. Novotni, and R. Klein, “3d shape matching with 3d shape contexts,” in *In The 7th Central European Seminar on Computer Graphics*, 2003.
- [376] K. Astikainen, L. Holm, E. Pitkänen, S. Szedmak, and J. Rousu, “Towards structured output prediction of enzyme function,” *BMC Proceedings*, vol. 2, no. 4, p. S2, 2008. [Online]. Available: <http://dx.doi.org/10.1186/1753-6561-2-s4-s2>
- [377] S. Takei, S. Akizuki, and M. Hashimoto, “Short: A fast 3d feature description based on estimating occupancy in spherical shell regions,” in *2015 International Conference on Image and Vision Computing New Zealand (IVCNZ)*, Nov 2015, pp. 1–5.
- [378] Y. Zhong, “Intrinsic shape signatures: A shape descriptor for 3d object recognition,” in *Computer Vision Workshops (ICCV Workshops), 2009 IEEE 12th International Conference on*. IEEE, 2009, pp. 689–696.
- [379] S. Salti, F. Tombari, and L. D. Stefano, “Shot: Unique signatures of histograms for surface and texture description,” *Computer Vision and Image Understanding*, vol. 125, pp. 251 – 264, 2014. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S1077314214000988>
- [380] D. of Computer Science and E. D. U. of Bologna, “Descriptor matching dataset,” <http://vision.deis.unibo.it/research/>, acessado em 2017-04-15.
- [381] L. F. Rocha, P. Malaca, J. Silva, A. P. Moreira, and G. Veiga, “Development of a 3d model based part recognition system for industrial applications: Main challenges,” in *2015 IEEE International Conference on Industrial Technology (ICIT)*, March 2015, pp. 3296–3301.
- [382] S. Hinterstoisser, V. Lepetit, S. Ilic, S. Holzer, G. Bradski, K. Konolige, and N. Navab, “Model based training, detection and pose estimation of texture-less 3d objects in heavily cluttered scenes,” in *Proceedings of the 11th Asian Conference on Computer Vision - Volume Part I*, ser. ACCV’12. Berlin, Heidelberg: Springer-Verlag, 2013, pp. 548–562.
- [383] C.-C. Chang and C.-J. Lin, “Libsvm: A library for support vector machines,” *ACM Trans. Intell. Syst. Technol.*, vol. 2, no. 3, pp. 27:1–27:27, May 2011. [Online]. Available: <http://doi.acm.org/10.1145/1961189.1961199>
- [384] J. Shimamura, T. Yoshida, Y. Taniguchi, H. Yabushita, K. Sudo, and K. Murasaki, “The method based on view-directional consistency constraints for robust 3d object recognition,” in *2015 14th IAPR International Conference on Machine Vision Applications (MVA)*, May 2015, pp. 455–458.
- [385] V. R. Chandrasekhar, D. M. Chen, S. S. Tsai, N.-M. Cheung, H. Chen, G. Takacs, Y. Reznik, R. Vedantham, R. Grzeszczuk, J. Bach, and B. Girod, “The stanford mobile visual search data set,” in *Proceedings of the Second Annual ACM Conference on Multimedia Systems*, ser. MMSys ’11. New York, NY, USA: ACM, 2011, pp. 117–122. [Online]. Available: <http://doi.acm.org/10.1145/1943552.1943568>
- [386] F. Zou, Y. Wang, Y. Yang, K. Zhou, Y. Chen, and J. Song, “Supervised feature learning via l2-norm regularized logistic regression for 3d object recognition,” *Neurocomputing*, vol. 151, Part 2, pp. 603 – 611, 2015. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0925231214013903>
- [387] S. Savarese and L. Fei-Fei, “3d generic object categorization, localization and pose estimation,” in *2007 IEEE 11th International Conference on Computer Vision*, Oct 2007, pp. 1–8.
- [388] G. Flitton, A. Mouton, and T. P. Breckon, “Object classification in 3d baggage security computed tomography imagery using visual codebooks,” *Pattern Recognition*, vol. 48, no. 8, pp. 2489 – 2499, 2015. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0031320315000540>
- [389] Q. Xu, W. Wan, J. Wang, and X. An, “The application of local features in 3-dimensional object recognition,” in *2015 International Conference on Smart and Sustainable City and Big Data (ICSSC)*, July 2015, pp. 96–100.
- [390] M. Soysal and A. A. Alatan, “Joint utilization of local appearance and geometric invariants for 3d object recognition,” *Multimedia Tools Appl.*, vol. 74, no. 8, pp. 2611–2637, Apr. 2015. [Online]. Available: <http://dx.doi.org/10.1007/s11042-013-1622-6>

- [391] I. Weiss and M. Ray, *Model-based recognition of 3D objects from one view*. Berlin, Heidelberg: Springer Berlin Heidelberg, 1998, pp. 716–732. [Online]. Available: <http://dx.doi.org/10.1007/BFb0054775>
- [392] M. Soysal and A. A. Alatan, “Recognition of 3d objects from unconstrained 2d images by using local appearance and affine geometry,” in *2013 IEEE International Conference on Multimedia and Expo (ICME)*, July 2013, pp. 1–6.
- [393] F. Rothganger, S. Lazebnik, C. Schmid, and J. Ponce, “3d object modeling and recognition using local affine-invariant image descriptors and multi-view spatial constraints,” *International Journal of Computer Vision*, vol. 66, no. 3, pp. 231–259, 2006. [Online]. Available: <http://dx.doi.org/10.1007/s11263-005-3674-1>
- [394] T. P. G. C. Vision and U. o. I. a. U.-C. Robotics Beckman Institute, “Datasets for computer vision research,” http://www-cvr.ai.uiuc.edu/ponce_grp/data/, acessado em 2017-04-05.
- [395] W. J. Beksi and N. Papanikolopoulos, “Object classification using dictionary learning and rgb-d covariance descriptors,” in *2015 IEEE International Conference on Robotics and Automation (ICRA)*, May 2015, pp. 1880–1885.
- [396] K. Lai, L. Bo, X. Ren, and D. Fox, “Sparse distance learning for object recognition combining rgb and depth information,” in *2011 IEEE International Conference on Robotics and Automation*, May 2011, pp. 4007–4013.
- [397] L. Bo, X. Ren, and D. Fox, “Depth kernel descriptors for object recognition,” in *2011 IEEE/RSJ International Conference on Intelligent Robots and Systems*, Sept 2011, pp. 821–826.
- [398] M. Blum, J. T. Springenberg, J. Wülfing, and M. Riedmiller, “A learned feature descriptor for object recognition in rgb-d data,” in *2012 IEEE International Conference on Robotics and Automation*, May 2012, pp. 1298–1303.
- [399] L. Bo, X. Ren, and D. Fox, *Unsupervised Feature Learning for RGB-D Based Object Recognition*. Heidelberg: Springer International Publishing, 2013, pp. 387–402.
- [400] D. Fehr, W. J. Beksi, D. Zermas, and N. Papanikolopoulos, “Covariance based point cloud descriptors for object detection and recognition,” *Comput. Vis. Image Underst.*, vol. 142, no. C, pp. 80–93, Jan. 2016. [Online]. Available: <https://doi.org/10.1016/j.cviu.2015.06.008>
- [401] S. Filipe, L. Itti, and L. A. Alexandre, “Bik-bus: Biologically motivated 3d keypoint based on bottom-up saliency,” *IEEE Transactions on Image Processing*, vol. 24, no. 1, pp. 163–175, Jan 2015.
- [402] L. Itti, C. Koch, and E. Niebur, “A model of saliency-based visual attention for rapid scene analysis,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 20, no. 11, pp. 1254–1259, Nov 1998.
- [403] C. Harris and M. Stephens, “A combined corner and edge detector,” in *In Proc. of Fourth Alvey Vision Conference*, 1988, pp. 147–151.
- [404] R. B. Rusu and S. Cousins, “3d is here: Point cloud library (pcl),” in *2011 IEEE International Conference on Robotics and Automation*, May 2011, pp. 1–4.
- [405] S. Filipe and L. A. Alexandre, “A comparative evaluation of 3d keypoint detectors in a rgb-d object dataset,” in *2014 International Conference on Computer Vision Theory and Applications (VISAPP)*, vol. 1, Jan 2014, pp. 476–483.
- [406] R. B. Rusu, N. Blodow, Z. C. Marton, and M. Beetz, “Aligning point cloud views using persistent feature histograms,” in *2008 IEEE/RSJ International Conference on Intelligent Robots and Systems*, Sept 2008, pp. 3384–3391.
- [407] J. G. Daugman, “High confidence visual recognition of persons by a test of statistical independence,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 15, no. 11, pp. 1148–1161, Nov 1993.
- [408] J. Garstka and G. Peters, “Adaptive 3-d object classification with reinforcement learning,” in *2015 12th International Conference on Informatics in Control, Automation and Robotics (ICINCO)*, vol. 02, July 2015, pp. 381–385.
- [409] A. M. Naguib and S. Lee, “An adaptive evidence structure for bayesian recognition of 3d objects,” in *Proceedings of the 9th International Conference on Ubiquitous Information Management and Communication*, ser. IMCOM '15. New York, NY, USA: ACM, 2015, pp. 75:1–75:8. [Online]. Available: <http://doi.acm.org/10.1145/2701126.2701160>
- [410] —, “Adaptive bayesian recognition with multiple evidences,” in *2014 International Conference on Multimedia Computing and Systems (ICMCS)*, April 2014, pp. 337–344.
- [411] S. H. Kasaei, M. Oliveira, G. H. Lim, L. Seabra Lopes, and A. M. Tomé, “Interactive open-ended learning for 3d object recognition: An approach and experiments,” *J. Intell. Robotics Syst.*, vol. 80, no. 3-4, pp. 537–553, Dec. 2015. [Online]. Available: <http://dx.doi.org/10.1007/s10846-015-0189-z>
- [412] S. H. Kasaei, M. Oliveira, G. H. Lim, L. S. Lopes, and A. M. Tomé, “An interactive open-ended learning approach for 3d object recognition,” in *2014 IEEE International Conference on Autonomous Robot Systems and Competitions (ICARSC)*, May 2014, pp. 47–52.
- [413] D. Schulz, W. Burgard, D. Fox, and A. B. Cremers, “Tracking multiple moving targets with a mobile robot using particle filters and statistical data association,” in *Proceedings 2001 ICRA. IEEE International Conference on Robotics and Automation (Cat. No.01CH37164)*, vol. 2, 2001, pp. 1665–1670 vol.2.
- [414] A. Chauhan and L. Seabra Lopes, “Using spoken words to guide open-ended category formation,” *Cognitive Processing*, vol. 12, no. 4, p. 341, 2011. [Online]. Available: <http://dx.doi.org/10.1007/s10339-011-0407-y>
- [415] J. Wang, J. Lu, W. Chen, and X. Wu, “Convolutional neural network for 3d object recognition based on rgb-d dataset,” in *2015 IEEE 10th Conference on Industrial Electronics and Applications (ICIEA)*, June 2015, pp. 34–39.
- [416] Y. Jia, E. Shelhamer, J. Donahue, S. Karayev, J. Long, R. Girshick, S. Guadarrama, and T. Darrell, “Caffe: Convolutional architecture for fast feature embedding,” in *Proceedings of the 22Nd ACM International Conference on Multimedia*, ser. MM '14. New York, NY, USA: ACM, 2014, pp. 675–678. [Online]. Available: <http://doi.acm.org/10.1145/2647868.2654889>
- [417] S. Gupta, R. Girshick, P. Arbeláez, and J. Malik, “Learning rich features from rgb-d images for object detection and segmentation,” in *European Conference on Computer Vision*. Springer, 2014, pp. 345–360.
- [418] R. C. Luo, C.-W. Kuo, and Y. T. Chung, “Model-based 3d object recognition and fetching by a 7-dof robot with online obstacle avoidance for factory automation,” in *2015 IEEE International Conference on Robotics and Automation (ICRA)*, May 2015, pp. 2647–2652.
- [419] R. C. Luo and C. W. Kuo, “A scalable modular architecture of 3d object acquisition for manufacturing automation,” in *2015 IEEE 13th International Conference on Industrial Informatics (INDIN)*, July 2015, pp. 269–274.
- [420] J. C. Rangel, V. Morell, M. Cazorla, S. Orts-Escolano, and J. García-Rodríguez, “Using gng on 3d object recognition in noisy rgb-d data,” in *2015 International Joint Conference on Neural Networks (IJCNN)*, July 2015, pp. 1–7.
- [421] B. Fritzke, “A growing neural gas network learns topologies,” in *Advances in Neural Information Processing Systems 7*. MIT Press, 1995, pp. 625–632.
- [422] M. Muja and D. G. Lowe, “Scalable nearest neighbor algorithms for high dimensional data,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 36, no. 11, pp. 2227–2240, Nov 2014.
- [423] G. Xu, B. Mourrain, R. Duvigneau, and A. Galligo, “Analysis-suitable volume parameterization of multi-block computational domain in isogeometric applications,” *Computer-Aided Design*, vol. 45, no. 2, pp. 395 – 404, 2013, solid and Physical Modeling 2012. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0010448512002278>
- [424] Y. Xia, L. Zhang, W. Xu, Z. Shan, and Y. Liu, “Recognizing multi-view objects with occlusions using a deep architecture,” *Information Sciences*, vol. 320, pp. 333 – 345, 2015. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0020025515000766>
- [425] C. Hong, J. Yu, J. You, X. Chen, and D. Tao, “Multi-view ensemble manifold regularization for 3d object recognition,” *Information Sciences*, vol. 320, pp. 395 – 405, 2015. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0020025515001978>
- [426] S. A. A. Shah, M. Bennamoun, and F. Boussaid, “A novel feature representation for automatic 3d object recognition in cluttered scenes,” *Neurocomputing*, vol. 205, pp. 1 – 15, 2016. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0925231215017385>
- [427] S. A. A. Shah, M. Bennamoun, F. Boussaid, and A. A. El-Sallam, “A novel local surface description for automatic 3d object recognition in low resolution cluttered scenes,” in *2013 IEEE International Conference on Computer Vision Workshops*, Dec 2013, pp. 638–643.
- [428] O. Kechagias-Stamatis and N. Aouf, “Histogram of distances for local surface description,” in *2016 IEEE International Conference on Robotics and Automation (ICRA)*, May 2016, pp. 2487–2493.
- [429] F. Zhu, L. Shao, and Y. Fang, “Boosted cross-domain dictionary learning for visual categorization,” *IEEE Intelligent Systems*, vol. 31, no. 3, pp. 6–18, May 2016.
- [430] F. Zhu and L. Shao, “Weakly-supervised cross-domain dictionary learning for visual recognition,” *International Journal of Computer*

- Vision*, vol. 109, no. 1, pp. 42–59, 2014. [Online]. Available: <http://dx.doi.org/10.1007/s11263-014-0703-y>
- [431] J. Liu, J. Luo, and M. Shah, “Recognizing realistic actions from videos “in the wild”,” in *Computer vision and pattern recognition, 2009. CVPR 2009. IEEE conference on*. IEEE, 2009, pp. 1996–2003.
- [432] L. Fei-Fei, R. Fergus, and P. Perona, “Learning generative visual models from few training examples: An incremental bayesian approach tested on 101 object categories,” *Computer Vision and Image Understanding*, vol. 106, no. 1, pp. 59 – 70, 2007, special issue on Generative Model Based Vision. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S1077314206001688>
- [433] B. Li, Y. Lu, C. Li, A. Godil, T. Schreck, M. Aono, M. Burtcher, H. Fu, T. Furuya, H. Johan *et al.*, “Shrec’14 track: extended large scale sketch-based 3d shape retrieval,” in *Eurographics workshop on 3D object retrieval*, vol. 2014, 2014.
- [434] H. Kuehne, H. Jhuang, E. Garrote, T. Poggio, and T. Serre, “Hmdb: a large video database for human motion recognition,” in *Computer Vision (ICCV), 2011 IEEE International Conference on*. IEEE, 2011, pp. 2556–2563.
- [435] N. Silberman, D. Hoiem, P. Kohli, and R. Fergus, *Indoor Segmentation and Support Inference from RGBD Images*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2012, pp. 746–760.
- [436] J. Wang, J. Yang, K. Yu, F. Lv, T. Huang, and Y. Gong, “Locality-constrained linear coding for image classification,” in *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, June 2010, pp. 3360–3367.
- [437] J. Yang, K. Yu, Y. Gong, and T. Huang, “Linear spatial pyramid matching using sparse coding for image classification,” in *2009 IEEE Conference on Computer Vision and Pattern Recognition*, June 2009, pp. 1794–1801.
- [438] M. Aharon, M. Elad, and A. Bruckstein, “rmk-svd: An algorithm for designing overcomplete dictionaries for sparse representation,” *IEEE Transactions on Signal Processing*, vol. 54, no. 11, pp. 4311–4322, Nov 2006.
- [439] Z. Jiang, Z. Lin, and L. S. Davis, “Label consistent k-svd: Learning a discriminative dictionary for recognition,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 11, pp. 2651–2664, Nov 2013.
- [440] W. Dai, Q. Yang, G.-R. Xue, and Y. Yu, “Boosting for transfer learning,” in *Proceedings of the 24th International Conference on Machine Learning*, ser. ICML ’07. New York, NY, USA: ACM, 2007, pp. 193–200. [Online]. Available: <http://doi.acm.org/10.1145/1273496.1273521>
- [441] K. Tateno, F. Tombari, and N. Navab, “When 2.5d is not enough: Simultaneous reconstruction, segmentation and recognition on dense slam,” in *2016 IEEE International Conference on Robotics and Automation (ICRA)*, May 2016, pp. 2295–2302.
- [442] R. A. Newcombe, S. Izadi, O. Hilliges, D. Molyneux, D. Kim, A. J. Davison, P. Kohi, J. Shotton, S. Hodges, and A. Fitzgibbon, “Kinectfusion: Real-time dense surface mapping and tracking,” in *2011 10th IEEE International Symposium on Mixed and Augmented Reality*, Oct 2011, pp. 127–136.
- [443] K. Tateno, F. Tombari, and N. Navab, “Real-time and scalable incremental segmentation on dense slam,” in *2015 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, Sept 2015, pp. 4465–4472.
- [444] A. Aldoma, F. Tombari, J. Prankl, A. Richtsfeld, L. D. Stefano, and M. Vincze, “Multimodal cue integration through hypotheses verification for rgb-d object recognition and 6dof pose estimation,” in *2013 IEEE International Conference on Robotics and Automation*, May 2013, pp. 2104–2111.
- [445] K. Lai, L. Bo, and D. Fox, “Unsupervised feature learning for 3d scene labeling,” in *2014 IEEE International Conference on Robotics and Automation (ICRA)*, May 2014, pp. 3050–3057.
- [446] S. Lee, L. Wei, and A. M. Naguib, “Adaptive bayesian recognition and pose estimation of 3d industrial objects with optimal feature selection,” in *2016 IEEE International Symposium on Assembly and Manufacturing (ISAM)*, Aug 2016, pp. 50–55.
- [447] Y. Lee, S. Lee, D. Kim, and J.-K. Oh, “Improved industrial part pose determination based on 3d closed-loop boundaries,” in *IEEE ISR 2013*, Oct 2013, pp. 1–3.
- [448] H. Murase and S. K. Nayar, “Visual learning and recognition of 3-d objects from appearance,” *International Journal of Computer Vision*, vol. 14, no. 1, pp. 5–24, 1995. [Online]. Available: <http://dx.doi.org/10.1007/BF01421486>
- [449] I. I. G and S. Prakash, “False mapped feature removal in spin images based 3d ear recognition,” in *2016 3rd International Conference on Signal Processing and Integrated Networks (SPIN)*, Feb 2016, pp. 620–623.
- [450] J. Bedkowski, K. Majek, P. Majek, P. Musialik, M. Pelka, and A. Nüchter, “Intelligent mobile system for improving spatial design support and security inside buildings,” *Mob. Netw. Appl.*, vol. 21, no. 2, pp. 313–326, Apr. 2016. [Online]. Available: <http://dx.doi.org/10.1007/s11036-015-0654-8>
- [451] D. Borrmann, J. Elseberg, K. Lingemann, A. Nachter, and J. Hertzberg, “Globally consistent 3d mapping with scan matching,” *Robotics and Autonomous Systems*, vol. 56, no. 2, pp. 130 – 142, 2008. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0921889007000863>
- [452] S. H. Kasaei, A. M. Tomé, L. S. Lopes, and M. Oliveira, “Good: A global orthographic object descriptor for 3d object recognition and manipulation,” *Pattern Recognition Letters*, vol. 83, Part 3, pp. 312 – 320, 2016, efficient Shape Representation, Matching, Ranking, and its Applications. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0167865516301684>
- [453] F. Chen, R. Ji, and L. Cao, “Multimodal learning for view-based 3d object classification,” *Neurocomputing*, vol. 195, pp. 23 – 29, 2016, learning for Medical Imaging. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S092523121600103X>
- [454] C. Li, J. Bohren, E. Carlson, and G. D. Hager, “Hierarchical semantic parsing for object pose estimation in densely cluttered scenes,” in *2016 IEEE International Conference on Robotics and Automation (ICRA)*, May 2016, pp. 5068–5075.
- [455] R. Girshick, J. Donahue, T. Darrell, and J. Malik, “Rich feature hierarchies for accurate object detection and semantic segmentation,” in *Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition*, ser. CVPR ’14. Washington, DC, USA: IEEE Computer Society, 2014, pp. 580–587. [Online]. Available: <http://dx.doi.org/10.1109/CVPR.2014.81>
- [456] C. Li, J. Bohren, and G. D. Hager, “Bridging the robot perception gap with mid-level vision,” in *International Symposium on Robotics Research (ISRR)*, 2015.
- [457] K. Lai, L. Bo, X. Ren, and D. Fox, “A large-scale hierarchical multi-view rgb-d object dataset,” in *2011 IEEE International Conference on Robotics and Automation*, May 2011, pp. 1817–1824.
- [458] K. B. Logoglu, S. Kalkan, and A. Temizel, “Cospair: Colored histograms of spatial concentric surflet-pairs for 3d object recognition,” *Robotics and Autonomous Systems*, vol. 75, Part B, pp. 558 – 570, 2016. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0921889015002225>
- [459] A. Singh, J. Sha, K. S. Narayan, T. Achim, and P. Abbeel, “Bigbird: A large-scale 3d database of object instances,” in *2014 IEEE International Conference on Robotics and Automation (ICRA)*, May 2014, pp. 509–516.
- [460] A. Singh, K. Narayan, B. Kehoe, S. Patil, K. Goldberg, and P. Abbeel, “Amazon picking challenge object scans,” http://rll.berkeley.edu/amazon_picking_challenge/, accessed em 2017-01-17.
- [461] D. Naji, M. Fakir, B. Bouikhalene, and R. Elayachi, “Recognition of 3d objects using heat diffusion equations and random forests,” in *2016 13th International Conference on Computer Graphics, Imaging and Visualization (CGiV)*, March 2016, pp. 161–166.
- [462] D. Giorgi, S. Biasotti, and L. Paraboschi, “Shape retrieval contest 2007: Watertight models track,” 2007.
- [463] S. R. S. Varadhan, “On the behavior of the fundamental solution of the heat equation with variable coefficients,” *Comm. Pure Appl. Math.*, vol. 20, no. 2, pp. 431–455, May 1967, mR:0208191. Zbl:0155.16503.
- [464] K. Crane, C. Weischedel, and M. Wardetzky, “Geodesics in heat: A new approach to computing distance based on heat flow,” *ACM Trans. Graph.*, vol. 32, no. 5, pp. 152:1–152:11, Oct. 2013. [Online]. Available: <http://doi.acm.org/10.1145/2516971.2516977>
- [465] D. Naji, M. Fakir, and O. Bencharef, “3d objects retrieval using geodesic distance based on eikonal equation,” *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 15, no. 3, pp. 597–604, 2015.
- [466] S. A. A. Shah, M. Bennamoun, and F. Boussaid, “Keypoints-based surface representation for 3d modeling and 3d object recognition,” *Pattern Recognition*, vol. 64, pp. 29 – 38, 2017. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0031320316303429>
- [467] F. Deinzer, J. Denzler, C. Derichs, and H. Niemann, “Aspects of optimal viewpoint selection and viewpoint fusion,” in *Proceedings of the 7th Asian Conference on Computer Vision - Volume Part II*, ser. ACCV’06. Berlin, Heidelberg: Springer-Verlag, 2006, pp. 902–912.

- [468] R. S. Sutton and A. G. Barto, *Reinforcement Learning: An Introduction*, B. Book, Ed. Cambridge, MA: MIT Press, 1998.
- [469] F. G. Callari and F. P. Ferrie, "Active object recognition: Looking for differences," *International Journal of Computer Vision*, vol. 43, no. 3, pp. 189–204, 2001. [Online]. Available: <http://dx.doi.org/10.1023/A:1011135513777>
- [470] A. Torn and A. Zilinskas, *Global Optimization*. New York, NY, USA: Springer-Verlag New York, Inc., 1989.
- [471] R. Nian, G. Ji, W. Zhao, and C. Feng, "Probabilistic 3d object recognition from 2d invariant view sequence based on similarity," *Neurocomputing*, vol. 70, no. 4–6, pp. 785 – 793, 2007, advanced Neurocomputing Theory and Methodology Selected papers from the International Conference on Intelligent Computing 2005 (ICIC 2005) International Conference on Intelligent Computing 2005. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0925231206002839>
- [472] V. Zografos and B. F. Buxton, "Pose-invariant 3d object recognition using linear combination of 2d views and evolutionary optimisation," in *Proceedings of the International Conference on Computing: Theory and Applications*, ser. ICCTA '07. Washington, DC, USA: IEEE Computer Society, 2007, pp. 645–649. [Online]. Available: <http://dx.doi.org/10.1109/ICCTA.2007.105>
- [473] T. Sim, S. Baker, and M. Bsat, "The cmu pose, illumination, and expression (pie) database," in *Proceedings of Fifth IEEE International Conference on Automatic Face Gesture Recognition*, May 2002, pp. 46–51.
- [474] R. A. Vázquez, H. Sossa, and B. A. Garro, "3d object recognition based on low frequency response and random feature selection," in *Proceedings of the Artificial Intelligence 6th Mexican International Conference on Advances in Artificial Intelligence*, ser. MICAI'07. Berlin, Heidelberg: Springer-Verlag, 2007, pp. 694–704.
- [475] —, "The role of the infant vision system in 3d object recognition," in *Advances in Neuro-Information Processing*, M. Köppen, N. Kasabov, and G. Coghill, Eds. Berlin, Heidelberg: Springer-Verlag, 2009, pp. 800–807.
- [476] M. D. Luciw and J. Weng, "Topographic class grouping with applications to 3d object recognition," in *2008 IEEE International Joint Conference on Neural Networks (IEEE World Congress on Computational Intelligence)*, June 2008, pp. 3987–3994.
- [477] J. Weng and M. D. Luciw, "Optimal in-place self-organization for cortical development: Limited cells, sparse coding and cortical topography," in *Proc. IEEE 5th International Conf. on Development and Learning (ICDL 2006)*, 2006.
- [478] Y. LeCun, F. J. Huang, and L. Bottou, "Learning methods for generic object recognition with invariance to pose and lighting," in *Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2004. CVPR 2004.*, vol. 2, June 2004, pp. II-97–104 Vol.2.
- [479] Övünç Polat and T. Yildirim, "Genetic optimization of {GRNN} for pattern recognition without feature extraction," *Expert Systems with Applications*, vol. 34, no. 4, pp. 2444 – 2448, 2008. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0957417407001376>
- [480] —, "Recognition of patterns without feature extraction by grnn," in *Proceedings of the 8th International Conference on Adaptive and Natural Computing Algorithms, Part II*, ser. ICANNGA '07. Berlin, Heidelberg: Springer-Verlag, 2007, pp. 161–168.
- [481] D. Dimov, N. Zlateva, and A. Marinov, "Cbir over multiple projections of 3d objects," in *Proceedings of the 2009 Joint COST 2101 and 2102 International Conference on Biometric ID Management and Multimodal Communication*. Berlin, Heidelberg: Springer-Verlag, 2009, pp. 146–153.
- [482] D. Dimov, "A polar-fourier-wavelet transform for effective cbir," *Proc. of the ADMKD*, vol. 7, no. 02.10, p. 07, 2007.
- [483] C. Urdiales, C. de Trazegnies, J. Pacheco, and F. Sandoval, "View planning for efficient contour-based 3d object recognition," in *Melecon 2010 - 2010 15th IEEE Mediterranean Electrotechnical Conference*, April 2010, pp. 206–211.
- [484] B. Pang and H. Ma, "An effective way of 3d model representation in recognition system," in *Proceedings of the 2011 International Conference on Multimedia and Signal Processing - Volume 01*, ser. CMSP '11. Washington, DC, USA: IEEE Computer Society, 2011, pp. 107–111. [Online]. Available: <http://dx.doi.org/10.1109/CMSP.2011.28>
- [485] J. B. Tenenbaum, V. de Silva, and J. C. Langford, "A global geometric framework for nonlinear dimensionality reduction," *Science*, no. 5500, p. 2319, 2000.
- [486] M. Ulrich, C. Wiedemann, and C. Steger, "Combining scale-space and similarity-based aspect graphs for fast 3d object recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 10, pp. 1902–1914, Oct. 2012. [Online]. Available: <http://dx.doi.org/10.1109/TPAMI.2011.266>
- [487] D. W. Eggert, K. W. Bowyer, C. R. Dyer, H. I. Christensen, and D. B. Goldgof, "The scale space aspect graph," in *Proceedings 1992 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, June 1992, pp. 335–340.
- [488] C. Steger, "Occlusion, clutter, and illumination invariant object recognition," 2002.
- [489] A. S. Elons, M. Abull-ela, and M. Tolba, "A proposed {PCNN} features quality optimization technique for pose-invariant 3d arabic sign language recognition," *Applied Soft Computing*, vol. 13, no. 4, pp. 1646 – 1660, 2013. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S1568494612005200>
- [490] D. Mery, V. Riffo, I. Zuccar, and C. Pieringer, "Automated x-ray object recognition using an efficient search algorithm in multiple views," in *Proceedings of the 2013 IEEE Conference on Computer Vision and Pattern Recognition Workshops*, ser. CVPRW '13. Washington, DC, USA: IEEE Computer Society, 2013, pp. 368–374. [Online]. Available: <http://dx.doi.org/10.1109/CVPRW.2013.62>
- [491] N. A. Efremova and T. Inui, "An inferior temporal cortex model for object recognition and classification," *Sci. Tech. Inf. Process.*, vol. 41, no. 6, pp. 362–369, Dec. 2014. [Online]. Available: <http://dx.doi.org/10.3103/S0147688214060045>
- [492] N. Efremova, N. Asakura, T. Inui, and N. Abdikeev, "Inferotemporal network model for 3d object recognition," in *The 2011 IEEE/ICME International Conference on Complex Medical Engineering*, May 2011, pp. 555–560.
- [493] G. Pang and U. Neumann, "Fast and robust multi-view 3d object recognition in point clouds," in *Proceedings of the 2015 International Conference on 3D Vision*, ser. 3DV '15. Washington, DC, USA: IEEE Computer Society, 2015, pp. 171–179. [Online]. Available: <http://dx.doi.org/10.1109/3DV.2015.27>
- [494] D. Munoz, J. A. Bagnell, N. Vandapel, and M. Hebert, "Contextual classification with functional max-margin markov networks," in *2009 IEEE Conference on Computer Vision and Pattern Recognition*, June 2009, pp. 975–982.
- [495] K. Lai and D. Fox, "Object recognition in 3d point clouds using web data and domain adaptation," *Int. J. Rob. Res.*, vol. 29, no. 8, pp. 1019–1037, Jul. 2010. [Online]. Available: <http://dx.doi.org/10.1177/0278364910369190>
- [496] T. Fülhammer, M. Zillich, and M. Vincze, "Multi-view hypotheses transfer for enhanced object recognition in clutter," in *2015 14th IAPR International Conference on Machine Vision Applications (MVA)*, May 2015, pp. 10–13.
- [497] A. Aldoma, T. Fülhammer, and M. Vincze, "Automation of "ground truth" annotation for multi-view rgb-d object instance recognition datasets," in *Intelligent Robots and Systems (IROS 2014), 2014 IEEE/RSJ International Conference on*. IEEE, 2014, pp. 5016–5023.
- [498] A. Collet, M. Martinez, and S. S. Srinivasa, "The moped framework: Object recognition and pose estimation for manipulation," *The International Journal of Robotics Research*, vol. 30, no. 10, pp. 1284–1306, 2011.
- [499] Z. Xie, A. Singh, J. Uang, K. S. Narayan, and P. Abbeel, "Multimodal blending for high-accuracy instance recognition," in *Proceedings of the 26th IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2013.
- [500] —, "Multimodal blending for high-accuracy instance recognition," in *2013 IEEE/RSJ International Conference on Intelligent Robots and Systems*, Nov 2013, pp. 2214–2221.
- [501] J. Tang, S. Miller, A. Singh, and P. Abbeel, "A textured object recognition pipeline for color and depth image data," in *2012 IEEE International Conference on Robotics and Automation*, May 2012, pp. 3467–3474.
- [502] Y.-C. Chen, V. M. Patel, R. Chellappa, and P. J. Phillips, "Salient views and view-dependent dictionaries for object recognition," *Pattern Recognition*, vol. 48, no. 10, pp. 3053 – 3066, 2015, discriminative Feature Learning from Big Data for Visual Recognition. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0031320315000321>
- [503] S. Marini, M. Spagnuolo, and B. Falcidieno, "Structural shape prototypes for the automatic classification of 3d objects," *IEEE Computer Graphics and Applications*, vol. 27, no. 4, pp. 28–37, July 2007.
- [504] B. T. Messmer and H. Bunke, "A new algorithm for error-tolerant sub-graph isomorphism detection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 20, no. 5, pp. 493–504, May 1998.
- [505] S. Xia and E. R. Hancock, "3d object recognition using hyper-graphs and ranked local invariant features," in *Proceedings of the 2008 Joint*

- IAPR International Workshop on Structural, Syntactic, and Statistical Pattern Recognition*, ser. SSPR & SPR '08. Berlin, Heidelberg: Springer-Verlag, 2008, pp. 117–126.
- [506] D. Aouada and H. Krim, “Meaningful 3d shape partitioning using morse functions,” in *2009 16th IEEE International Conference on Image Processing (ICIP)*, Nov 2009, pp. 417–420.
- [507] J. Milnor, *Morse Theory:(AM-51)*. Princeton university press, 2016, vol. 51.
- [508] P. N. Klein, T. B. Sebastian, and B. B. Kimia, “Shape matching using edit-distance: An implementation,” in *SODA*, 2001, pp. 781–790.
- [509] A. Noma and R. M. Cesar Jr., “Sparse representations for efficient shape matching,” in *Proceedings of the 2010 23rd SIBGRAPI Conference on Graphics, Patterns and Images*, ser. SIBGRAPI '10. Washington, DC, USA: IEEE Computer Society, 2010, pp. 186–192. [Online]. Available: <http://dx.doi.org/10.1109/SIBGRAPI.2010.33>
- [510] P. F. Felzenszwalb, “Representation and detection of deformable shapes,” in *2003 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2003. Proceedings.*, vol. 1, June 2003, pp. I-102–I-108 vol.1.
- [511] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner, “Gradient-based learning applied to document recognition,” *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278–2324, Nov 1998.
- [512] H. Ma, T. Huang, and Y. Wang, “Multi-resolution recognition of 3d objects based on visual resolution limits,” *Pattern Recognition Letters*, vol. 31, no. 3, pp. 259 – 266, 2010. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0167865509002256>
- [513] B. Bonev, F. Escolano, D. Giorgi, and S. Biasotti, “Information-theoretic feature selection from unattributed graphs,” in *Proceedings of the 2010 20th International Conference on Pattern Recognition*, ser. ICPR '10. Washington, DC, USA: IEEE Computer Society, 2010, pp. 930–933. [Online]. Available: <http://dx.doi.org/10.1109/ICPR.2010.233>
- [514] —, “High-dimensional spectral feature selection for 3d object recognition based on reeb graphs,” in *Proceedings of the 2010 Joint IAPR International Conference on Structural, Syntactic, and Statistical Pattern Recognition*, ser. SSPR&SPR'10. Berlin, Heidelberg: Springer-Verlag, 2010, pp. 119–128. [Online]. Available: <http://dl.acm.org/citation.cfm?id=1887003.1887018>
- [515] G. Reeb, “Sur les points singuliers d’une forme de Pfaff complètement intégrable ou d’une fonction numérique,” *Comptes Rendus Acad. Sciences*, vol. 222, pp. 847–849, 1946.
- [516] S. Biasotti and M. Attene, “Shape retrieval contest 2008: Stability of watertight models,” in *2008 IEEE International Conference on Shape Modeling and Applications*, June 2008.
- [517] K. J. Yoon, M. G. Shin, and J. H. Lee, “Recognizing 3d objects with 3d information from stereo vision,” in *2010 20th International Conference on Pattern Recognition*, Aug 2010, pp. 4020–4023.
- [518] S. Kim, K.-J. Yoon, and I. S. Kweon, “Object recognition using a generalized robust invariant feature and gestalt’s law of proximity and similarity,” *Pattern Recognition*, vol. 41, no. 2, pp. 726 – 741, 2008. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0031320307002567>
- [519] S. Gold and A. Rangarajan, “A graduated assignment algorithm for graph matching,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 18, no. 4, pp. 377–388, Apr. 1996. [Online]. Available: <http://dx.doi.org/10.1109/34.491619>
- [520] M. Hu, Z. Wei, M. Shao, and G. Zhang, “3-d object recognition via aspect graph aware 3-d object representation,” *IEEE Signal Processing Letters*, vol. 22, no. 12, pp. 2359–2363, Dec 2015.
- [521] N. Snavely, S. M. Seitz, and R. Szeliski, “Photo tourism: exploring photo collections in 3d,” in *ACM transactions on graphics (TOG)*, vol. 25, no. 3. ACM, 2006, pp. 835–846.
- [522] D. Glasner, M. Galun, S. Alpert, R. Basri, and G. Shakhnarovich, “Viewpoint-aware object detection and pose estimation,” in *2011 International Conference on Computer Vision*, Nov 2011, pp. 1275–1282.
- [523] K. Madi, E. Paquet, H. Seba, and H. Kheddouci, “Graph edit distance based on triangle-stars decomposition for deformable 3d objects recognition,” in *Proceedings of the 2015 International Conference on 3D Vision*, ser. 3DV '15. Washington, DC, USA: IEEE Computer Society, 2015, pp. 55–63. [Online]. Available: <http://dx.doi.org/10.1109/3DV.2015.14>
- [524] A. M. Bronstein, M. M. Bronstein, and R. Kimmel, “Calculus of nonrigid surfaces for geometry and texture manipulation,” *IEEE Transactions on Visualization and Computer Graphics*, vol. 13, no. 5, pp. 902–913, Sept 2007.
- [525] H. Tabia, M. Daoudi, J. P. Vandeborre, and O. Colot, “A new 3d-matching method of nonrigid and partially similar models using curve analysis,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 33, no. 4, pp. 852–858, April 2011.
- [526] D. Saupe and D. V. Vranic, “3d model retrieval with spherical harmonics and moments,” in *Proceedings of the 23rd DAGM-Symposium on Pattern Recognition*. London, UK, UK: Springer-Verlag, 2001, pp. 392–397. [Online]. Available: <http://dl.acm.org/citation.cfm?id=648286.756262>
- [527] S. Sorlin, C. Solnon, and J.-M. Jolion, *A Generic Graph Distance Measure Based on Multivalent Matchings*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2007, pp. 151–181.
- [528] S. Wang, Y. Wang, M. Jin, X. Gu, and D. Samaras, “3d surface matching and recognition using conformal geometry,” in *Proceedings of the 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition - Volume 2*, ser. CVPR '06. Washington, DC, USA: IEEE Computer Society, 2006, pp. 2453–2460. [Online]. Available: <http://dx.doi.org/10.1109/CVPR.2006.17>
- [529] S. Wang, Y. Wang, M. Jin, X. D. Gu, and D. Samaras, “Conformal geometry and its applications on 3d shape matching, recognition, and stitching,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 29, no. 7, pp. 1209–1220, July 2007.
- [530] S. Pintilie and A. Ghodsi, “Conformal mapping by computationally efficient methods,” in *Proceedings of the Twenty-Fourth AAAI Conference on Artificial Intelligence*, ser. AAAI'10. AAAI Press, 2010, pp. 557–562. [Online]. Available: <http://dl.acm.org/citation.cfm?id=2898607.2898697>
- [531] J. a. Rodrigues, R. Lam, and H. du Buf, “Cortical 3d face and object recognition using 2d projections,” *Int. J. Creat. Interaces Comput. Graph.*, vol. 3, no. 1, pp. 45–62, Jan. 2012. [Online]. Available: <http://dx.doi.org/10.4018/jcicg.2012010104>
- [532] J. Rodrigues and J. H. du Buf, “Multi-scale lines and edges in {V1} and beyond: Brightness, object categorization and recognition, and consciousness,” *Biosystems*, vol. 95, no. 3, pp. 206 – 226, 2009. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0303264708002372>
- [533] J. Rodrigues and J. M. Hans du Buf, “A cortical framework for invariant object categorization and recognition,” *Cognitive Processing*, vol. 10, no. 3, pp. 243–261, 2009. [Online]. Available: <http://dx.doi.org/10.1007/s10339-009-0262-2>
- [534] S. Aouat, N. Laiche, F. Souami, and S. Larabi, “3d object indexing and recognition,” *Applied Mathematics and Computation*, vol. 196, no. 1, pp. 318 – 332, 2008. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0096300307006741>
- [535] H. Q. Truong, S. Lee, and S.-W. Jang, “Model-based recognition of 3d objects using intersecting lines,” in *2008 IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems*, Aug 2008, pp. 656–660.
- [536] A. Khatun, Y. C. Wang, M. R. Islam, and D. A. Iskandar, “3d shape recognition using wavelet transform based on ellipsoid,” in *2010 International Conference on Intelligent and Advanced Systems*, June 2010, pp. 1–6.
- [537] A. Khatun, W. Y. Chai, D. A. Iskandar, and M. R. Islam, “The effectiveness of ellipsoidal shape representation technique for 3d object recognition system,” in *2011 7th International Conference on Information Technology in Asia*, July 2011, pp. 1–6.
- [538] S. Lee, Z. Lu, and H. Kim, “Probabilistic 3d object recognition with both positive and negative evidences,” in *Proceedings of the 2011 International Conference on Computer Vision*, ser. ICCV '11. Washington, DC, USA: IEEE Computer Society, 2011, pp. 2360–2367. [Online]. Available: <http://dx.doi.org/10.1109/ICCV.2011.6126518>
- [539] D. Lin, S. Fidler, and R. Urtasun, “Holistic scene understanding for 3d object detection with rgbd cameras,” in *Proceedings of the 2013 IEEE International Conference on Computer Vision*, ser. ICCV '13. Washington, DC, USA: IEEE Computer Society, 2013, pp. 1417–1424. [Online]. Available: <http://dx.doi.org/10.1109/ICCV.2013.179>
- [540] J. Carreira and C. Sminchisescu, “Cpmc: Automatic object segmentation using constrained parametric min-cuts,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 34, no. 7, pp. 1312–1328, July 2012.
- [541] M. Hejrati and D. Ramanan, “Analysis by synthesis: 3d object recognition by object reconstruction,” in *2014 IEEE Conference on Computer Vision and Pattern Recognition*, June 2014, pp. 2449–2456.
- [542] —, “Analyzing 3d objects in cluttered images,” in *Advances in Neural Information Processing Systems*, 2012, pp. 593–601.
- [543] L. Torresani, A. Hertzmann, and C. Bregler, “Learning non-rigid 3d shape from 2d motion,” in *Advances in Neural Information Processing Systems 16*, S. Thrun, L. Saul, and B. Schölkopf, Eds. Cambridge, MA: MIT Press, 2003, p. None.

- [544] J. Xiao, B. C. Russell, and A. Torralba, "Localizing 3d cuboids in single-view images," in *Proceedings of the 25th International Conference on Neural Information Processing Systems*, ser. NIPS'12. USA: Curran Associates Inc., 2012, pp. 746–754. [Online]. Available: <http://dl.acm.org/citation.cfm?id=2999134.2999218>
- [545] M. M. Zhang, M. D. Kennedy, M. A. Hsieh, and K. Daniilidis, "A triangle histogram for object classification by tactile sensing," in *2016 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, Oct 2016, pp. 4931–4938.
- [546] Lina, T. Takahashi, I. Ide, and H. Murase, "Construction of appearance manifold with embedded view-dependent covariance matrix for 3d object recognition," *IEICE - Trans. Inf. Syst.*, vol. E91-D, no. 4, pp. 1091–1100, Apr. 2008. [Online]. Available: <http://dx.doi.org/10.1093/ietisy/e91-d.4.1091>
- [547] Lina, "Recognition of 3d objects in various capturing conditions using appearance manifolds," in *2010 The 2nd International Conference on Computer and Automation Engineering (ICCAE)*, vol. 2, Feb 2010, pp. 349–352.
- [548] A. Kushal, C. Schmid, and J. Ponce, "Flexible object models for category-level 3d object recognition," in *2007 IEEE Conference on Computer Vision and Pattern Recognition*, June 2007, pp. 1–8.
- [549] S. Lazebnik, C. Schmid, and J. Ponce, "A maximum entropy framework for part-based texture and object recognition," in *Tenth IEEE International Conference on Computer Vision (ICCV'05) Volume 1*, vol. 1, Oct 2005, pp. 832–838 Vol. 1.
- [550] M. Everingham, L. Van Gool, C. K. I. Williams, J. Winn, and A. Zisserman, "The pascal visual object classes (voc) challenge," *International Journal of Computer Vision*, vol. 88, no. 2, pp. 303–338, 2010. [Online]. Available: <http://dx.doi.org/10.1007/s11263-009-0275-4>
- [551] B. Leibe, A. Leonardis, and B. Schiele, "Combined object categorization and segmentation with an implicit shape model," in *Workshop on statistical learning in computer vision, ECCV*, vol. 2, no. 5, 2004, p. 7.
- [552] C. Garcia and M. Delakis, "Convolutional face finder: a neural architecture for fast and robust face detection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 26, no. 11, pp. 1408–1423, Nov 2004.
- [553] R. Ibrayev and Y. B. Jia, "Recognition of curved surfaces from one-dimensional tactile data," *IEEE Transactions on Automation Science and Engineering*, vol. 9, no. 3, pp. 613–621, July 2012.
- [554] Y. B. Jia and J. Tian, "Surface patch reconstruction from one-dimensional tactile data," *IEEE Transactions on Automation Science and Engineering*, vol. 7, no. 2, pp. 400–407, April 2010.
- [555] W. Fan, F. R. Al-Osaimi, N. Bouguila, and J. X. Du, "3d object modeling and recognition via online hierarchical pitman-yor process mixture learning," in *2015 IEEE Global Conference on Signal and Information Processing (GlobalSIP)*, Dec 2015, pp. 448–452.
- [556] Y. W. Teh and M. I. Jordan, "Hierarchical bayesian nonparametric models with applications," 2010.
- [557] N. Bouguila, "Infinite liouville mixture models with application to text and texture categorization," *Pattern Recognition Letters*, vol. 33, no. 2, pp. 103 – 110, 2012. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0167865511003205>
- [558] J. Sun, M. Ovsjanikov, and L. Guibas, "A concise and provably informative multi-scale signature based on heat diffusion," in *Proceedings of the Symposium on Geometry Processing*, ser. SGP '09. Aire-la-Ville, Switzerland, Switzerland: Eurographics Association, 2009, pp. 1383–1392. [Online]. Available: <http://dl.acm.org/citation.cfm?id=1735603.1735621>
- [559] A. M. Bronstein, M. M. Bronstein, L. J. Guibas, and M. Ovsjanikov, "Shape google: Geometric words and expressions for invariant shape retrieval," *ACM Trans. Graph.*, vol. 30, no. 1, pp. 1:1–1:20, Feb. 2011. [Online]. Available: <http://doi.acm.org/10.1145/1899404.1899405>
- [560] S. Anand, A. Kirmani, S. Shrivastava, S. Chaudhury, and B. Bhaumik, "A microscopic framework for distributed object-recognition pose-estimation," in *2006 9th International Conference on Control, Automation, Robotics and Vision*, Dec 2006, pp. 1–6.
- [561] "Free software tools for robot and sensor applications," <http://playerstage.sourceforge.net>, accessed em 2017-04-19.
- [562] B. Raychev, T. Mino, T. Tamaki, and K. Kaneda, "View-invariant object recognition with visibility maps," in *Proceedings of the 2010 20th International Conference on Pattern Recognition*, ser. ICPR '10. Washington, DC, USA: IEEE Computer Society, 2010, pp. 1040–1043. [Online]. Available: <http://dx.doi.org/10.1109/ICPR.2010.260>
- [563] D. Kent, M. Behrooz, and S. Chernova, "Crowdsourcing the construction of a 3d object recognition database for robotic grasping," in *2014 IEEE International Conference on Robotics and Automation (ICRA)*, May 2014, pp. 4526–4531.
- [564] J. R. Quinlan, *C4.5: Programs for Machine Learning*. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 1993.
- [565] V. Nair and G. E. Hinton, "3d object recognition with deep belief nets," in *Proceedings of the 22nd International Conference on Neural Information Processing Systems*, ser. NIPS'09. USA: Curran Associates Inc., 2009, pp. 1339–1347. [Online]. Available: <http://dl.acm.org/citation.cfm?id=2984093.2984244>
- [566] M. Grzegorzec and E. Izquierdo, "Statistical 3d object classification and localization with context modeling," in *2007 15th European Signal Processing Conference*, Sept 2007, pp. 1585–1589.
- [567] M. P. Reinhold, M. Grzegorzec, J. Denzler, and H. Niemann, "Appearance-based recognition of 3-d objects by cluttered background and occlusions," *Pattern Recognition*, vol. 38, no. 5, pp. 739 – 753, 2005. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0031320304004327>
- [568] M. Z. Zia, M. Stark, and K. Schindler, "Are cars just 3d boxes? jointly estimating the 3d shape of multiple objects," in *Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition*, ser. CVPR '14. Washington, DC, USA: IEEE Computer Society, 2014, pp. 3678–3685. [Online]. Available: <http://dx.doi.org/10.1109/CVPR.2014.470>
- [569] A. Geiger, P. Lenz, and R. Urtasun, "Are we ready for autonomous driving? the kitti vision benchmark suite," in *2012 IEEE Conference on Computer Vision and Pattern Recognition*, June 2012, pp. 3354–3361.
- [570] P. Han and G. Zhao, "Cad-based 3d objects recognition in monocular images for mobile augmented reality," *Computers & Graphics*, vol. 50, pp. 36 – 46, 2015. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0097849315000680>
- [571] Y.-H. Xu, R.-H. Luo, and H.-Q. Min, "Label transfer for joint recognition and segmentation of 3d object," in *2012 International Conference on Machine Learning and Cybernetics*, vol. 3, July 2012, pp. 1188–1192.
- [572] K. Fukui and O. Yamaguchi, "The kernel orthogonal mutual subspace method and its application to 3d object recognition," in *Proceedings of the 8th Asian Conference on Computer Vision - Volume Part II*, ser. ACCV'07. Berlin, Heidelberg: Springer-Verlag, 2007, pp. 467–476. [Online]. Available: <http://dl.acm.org/citation.cfm?id=1775728.1775784>
- [573] K. Fukui, B. Stenger, and O. Yamaguchi, "A framework for 3d object recognition using the kernel constrained mutual subspace method," in *Proceedings of the 7th Asian Conference on Computer Vision - Volume Part II*, ser. ACCV'06. Berlin, Heidelberg: Springer-Verlag, 2006, pp. 315–324.
- [574] J. Wu and K. Fukui, "Multiple view based 3d object classification using ensemble learning of local subspaces," in *2008 19th International Conference on Pattern Recognition*, Dec 2008, pp. 1–4.
- [575] O. Yamaguchi, K. Fukui, and K. i. Maeda, "Face recognition using temporal image sequence," in *Proceedings Third IEEE International Conference on Automatic Face and Gesture Recognition*, 1998, pp. 318–323.
- [576] K. Hotta, "Pose independent object classification from small number of training samples based on kernel principal component analysis of local parts," *Image and Vision Computing*, vol. 27, no. 9, pp. 1240 – 1251, 2009. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0262885608002461>
- [577] N. Akihiro and K. Fukui, "Compound mutual subspace method for 3d object recognition: A theoretical extension of mutual subspace method," in *Proceedings of the 2010 International Conference on Computer Vision - Volume Part II*, ser. ACCV'10. Berlin, Heidelberg: Springer-Verlag, 2011, pp. 374–383. [Online]. Available: <http://dl.acm.org/citation.cfm?id=2040739.2040782>
- [578] C. Sanderson, *Biometric person recognition: Face, speech and fusion*. VDM Publishing, 2008, vol. 4.
- [579] Y. Igarashi and K. Fukui, "3d object recognition based on canonical angles between shape subspaces," in *Proceedings of the 10th Asian Conference on Computer Vision - Volume Part IV*, ser. ACCV'10. Berlin, Heidelberg: Springer-Verlag, 2011, pp. 580–591. [Online]. Available: <http://dl.acm.org/citation.cfm?id=1966111.1966158>
- [580] C. Tomasi and T. Kanade, "Shape and motion from image streams under orthography: a factorization method," *International Journal of Computer Vision*, vol. 9, no. 2, pp. 137–154, 1992. [Online]. Available: <http://dx.doi.org/10.1007/BF00129684>

- [581] K. Kise and T. Kashiwagi, "1.5 million subspaces of a local feature space for 3d object recognition," in *The First Asian Conference on Pattern Recognition*, Nov 2011, pp. 672–676.
- [582] T. Kobayashi, "Generalized mutual subspace based methods for image set classification," in *Proceedings of the 11th Asian Conference on Computer Vision - Volume Part I*, ser. ACCV'12. Berlin, Heidelberg: Springer-Verlag, 2013, pp. 578–592.
- [583] K. Fukui and O. Yamaguchi, *Face Recognition Using Multi-viewpoint Patterns for Robot Vision*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2005, pp. 192–201.
- [584] K. Fukui and A. Maki, "Difference subspace and its generalization for subspace-based methods," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 37, no. 11, pp. 2164–2177, Nov 2015.
- [585] S. Watanabe and N. Pakvasa, "Subspace method of pattern recognition," in *Proc. 1st. IJCP*, 1973, pp. 25–32.
- [586] K. Maeda and S. Watanabe, "A pattern matching method with local structure," in *Trans. IEICE*, 1985, pp. 345–352.
- [587] B. Schölkopf, A. Smola, and K.-R. Müller, "Nonlinear component analysis as a kernel eigenvalue problem," *Neural Comput.*, vol. 10, no. 5, pp. 1299–1319, Jul. 1998. [Online]. Available: <http://dx.doi.org/10.1162/089976698300017467>
- [588] T. K. Kim, J. Kittler, and R. Cipolla, "Discriminative learning and recognition of image set classes using canonical correlations," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 29, no. 6, pp. 1005–1018, June 2007.
- [589] J. Wright, A. Y. Yang, A. Ganesh, S. S. Sastry, and Y. Ma, "Robust face recognition via sparse representation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 31, no. 2, pp. 210–227, Feb 2009.
- [590] J. Hamm and D. D. Lee, "Grassmann discriminant analysis: A unifying view on subspace-based learning," in *Proceedings of the 25th International Conference on Machine Learning*, ser. ICML '08. New York, NY, USA: ACM, 2008, pp. 376–383. [Online]. Available: <http://doi.acm.org/10.1145/1390156.1390204>
- [591] T. Kawahara, M. Nishiyama, T. Kozakaya, and O. Yamaguchi, "Face recognition based on whitening transformation of distribution of subspaces," in *Workshop on Asian Conference on Computer Vision, Subspace*, 2007, pp. 97–103.
- [592] H. Sakano and N. Mukawa, "Kernel mutual subspace method for robust facial image recognition," in *Knowledge-Based Intelligent Engineering Systems and Allied Technologies, 2000. Proceedings. Fourth International Conference on*, vol. 1, 2000, pp. 245–248 vol.1.
- [593] H. Ben-Yaacov, D. Malah, and M. Barzohar, "Recognition of 3d objects based on implicit polynomials," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 32, no. 5, pp. 954–960, May 2010. [Online]. Available: <http://dx.doi.org/10.1109/TPAMI.2009.197>
- [594] D. Smeets, T. Fabry, J. Hermans, D. Vandermeulen, and P. Suetens, "Inelastic deformation invariant modal representation for non-rigid 3d object recognition," in *Proceedings of the 6th International Conference on Articulated Motion and Deformable Objects*, ser. AMDO'10. Berlin, Heidelberg: Springer-Verlag, 2010, pp. 162–171. [Online]. Available: <http://dl.acm.org/citation.cfm?id=1875984.1876000>
- [595] Z. Lian, A. Godil, T. Fabry, T. Furuya, J. Hermans, R. Ohbuchi, C. Shu, D. Smeets, P. Suetens, D. Vandermeulen, and S. Wuhrer, "SHREC'10 Track: Non-rigid 3D Shape Retrieval," in *Eurographics Workshop on 3D Object Retrieval*, M. Daoudi and T. Schreck, Eds. The Eurographics Association, 2010.
- [596] Y. Gur and C. R. Johnson, "Generalized hardi invariants by method of tensor contraction," in *2014 IEEE 11th International Symposium on Biomedical Imaging (ISBI)*, April 2014, pp. 718–721.
- [597] L. Bloy, M. Ingahlalkar, H. Eavani, T. P. L. Roberts, R. T. Schultz, and R. Verma, *HARDI Based Pattern Classifiers for the Identification of White Matter Pathologies*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2011, pp. 234–241.
- [598] H. Skibbe and M. Reiser, "Rotation covariant image processing for biomedical applications," *Comp. Math. Methods in Medicine*, vol. 2013, pp. 931507:1–931507:19, 2013. [Online]. Available: <http://dx.doi.org/10.1155/2013/931507>
- [599] M. Kazhdan, T. Funkhouser, and S. Rusinkiewicz, "Rotation invariant spherical harmonic representation of 3d shape descriptors," in *Proceedings of the 2003 Eurographics/ACM SIGGRAPH Symposium on Geometry Processing*, ser. SGP '03. Aire-la-Ville, Switzerland, Switzerland: Eurographics Association, 2003, pp. 156–164. [Online]. Available: <http://dl.acm.org/citation.cfm?id=882370.882392>
- [600] S. Orts-Escolano, V. Morell, J. Garcia-Rodriguez, M. Cazorla, and R. B. Fisher, "Real-time 3d semi-local surface patch extraction using gpgpu," *J. Real-Time Image Process.*, vol. 10, no. 4, pp. 647–666, Dec. 2015. [Online]. Available: <http://dx.doi.org/10.1007/s11554-013-0385-7>
- [601] A. S. Mian, M. Bennamoun, and R. A. Owens, "A novel representation and feature matching algorithm for automatic pairwise registration of range images," *International Journal of Computer Vision*, vol. 66, no. 1, pp. 19–40, 2006. [Online]. Available: <http://dx.doi.org/10.1007/s11263-005-3221-0>
- [602] C. H. Loo and A. Z. Elsherbeni, "Optoelectronic 3-d object classification from 2-d images," *Journal of Lightwave Technology*, vol. 26, no. 18, pp. 3248–3255, Sept 2008.
- [603] M. S. Alam and M. A. Karim, "Fringe-adjusted joint transform correlation," *Applied Optics*, vol. 32, no. 23, pp. 4344–4350, 1993.
- [604] S. Ullman, E. Sali, and M. Vidal-Naquet, *A Fragment-Based Approach to Object Representation and Classification*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2001, pp. 85–100.
- [605] M. Alam and E. Horache, "Optoelectronic implementation of fringe-adjusted joint transform correlation," *Optics Communications*, vol. 236, pp. 59 – 67, 2004. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S003040180400269X>
- [606] C.-H. Loo and M. S. Alam, "Invariant object tracking using fringe-adjusted joint transform correlator," *Optical Engineering*, vol. 43, no. 9, pp. 2175–2183, 2004.
- [607] T. Kounalakis, N. V. Boulgouris, and G. A. Triantafyllidis, "Content-adaptive pyramid representation for 3d object classification," in *2016 IEEE International Conference on Image Processing (ICIP)*, Sept 2016, pp. 231–235.
- [608] A. Mahiddine, D. Merad, P. Drap, and J. m. Boï, "Partial 3d-object retrieval using level curves," in *2014 6th International Conference of Soft Computing and Pattern Recognition (SoCPR)*, Aug 2014, pp. 77–82.
- [609] H. J. Wolfson, "On curve matching," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 12, no. 5, pp. 483–489, May 1990.
- [610] A. S. Mian, M. Bennamoun, and R. A. Owens, "3d recognition and segmentation of objects in cluttered scenes," in *Application of Computer Vision, 2005. WACV/MOTIONS '05 Volume 1. Seventh IEEE Workshops on*, vol. 1, Jan 2005, pp. 8–13.
- [611] Y. Kitaaki, H. Okuda, H. Kage, and K. Sumi, "High speed 3-d registration using gpu," in *2008 SICE Annual Conference*, Aug 2008, pp. 3055–3059.
- [612] T. Takeguchi and S. Kaneko, "Using depth aspect images for robust and efficient search of multiple objects in optomechatronic sensing," *IEEE Transactions on Industrial Electronics*, vol. 52, no. 4, pp. 1041–1049, Aug 2005.
- [613] H. Okuda, Y. Kitaaki, M. Hashimoto, and S. Kaneko, "HM-ICP: fast 3-d registration algorithm with hierarchical and region selection approach of M-ICP," *JRM*, vol. 18, no. 6, pp. 765–771, 2006. [Online]. Available: <http://dx.doi.org/10.20965/jrm.2006.p0765>
- [614] S. Kaneko, T. Kondo, and A. Miyamoto, "Robust matching of 3d contours using iterative closest point algorithm improved by m-estimation," *Pattern Recognition*, vol. 36, no. 9, pp. 2041 – 2047, 2003, kernel and Subspace Methods for Computer Vision. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0031320303000505>
- [615] M. F. Gafar and E. E. Hemayed, "Surface area distribution descriptor for object matching," *Journal of Advanced Research*, vol. 1, no. 3, pp. 233 – 241, 2010. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S2090123210000718>
- [616] X. Xu, A. Dehghani, D. Corrigan, S. Caulfield, and D. Moloney, "Convolutional neural network for 3d object recognition using volumetric representation," in *2016 First International Workshop on Sensing, Processing and Learning for Intelligent Machines (SPLINE)*, July 2016, pp. 1–5.
- [617] D. Maturana and S. Scherer, "Voxnet: A 3d convolutional neural network for real-time object recognition," in *2015 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, Sept 2015, pp. 922–928.
- [618] X. Zhang, Y. Liu, C. Gao, and J. Liu, "An isomap-eigenanalysis-regression pose estimation algorithm of three-dimensional object," in *2008 Second International Symposium on Intelligent Information Technology Application*, vol. 3, Dec 2008, pp. 61–65.
- [619] X. Zhang, H. M. Ma, Y. S. Liu, and C. X. Gao, "An extension of locally linear embedding for pose estimation of 3d object," in *2007 International Conference on Machine Learning and Cybernetics*, vol. 3, Aug 2007, pp. 1672–1677.
- [620] A. Nelleri, U. Gopinathan, J. Joseph, and K. Singh, "Three-dimensional object recognition from digital fresnel hologram by wavelet matched filtering," *Optics Communications*, vol.

- 259, no. 2, pp. 499 – 506, 2006. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0030401805009284>
- [621] D. Kumar and N. K. Nishchal, “Three-dimensional object recognition using joint fractional fourier transform correlators with the help of digital fresnel holography,” *Optik - International Journal for Light and Electron Optics*, vol. 126, no. 20, pp. 2690 – 2695, 2015. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0030402615005409>
- [622] C. M. Do and B. Javidi, “Three-dimensional object recognition with multiview photon-counting sensing and imaging,” *IEEE Photonics Journal*, vol. 1, no. 1, pp. 9–20, June 2009.
- [623] N. Yoshikawa and Y. Ii, “Three-dimensional object recognition using multiplex complex amplitude information with support function,” in *Proceedings of the First International Conference on Innovative Computing, Information and Control - Volume 1*, ser. ICICIC '06. Washington, DC, USA: IEEE Computer Society, 2006, pp. 314–317. [Online]. Available: <http://dx.doi.org/10.1109/ICICIC.2006.175>
- [624] Y. Xiang and S. Savarese, “Object detection by 3d aspectlets and occlusion reasoning,” in *Proceedings of the 2013 IEEE International Conference on Computer Vision Workshops*, ser. ICCVW '13. Washington, DC, USA: IEEE Computer Society, 2013, pp. 530–537. [Online]. Available: <http://dx.doi.org/10.1109/ICCVW.2013.75>
- [625] *Estimating the Aspect Layout of Object Categories*, 2012.
- [626] A. N. Ravari and H. D. Taghirad, “Transformation invariant 3d object recognition based on information complexity,” in *2014 Second RSI/ISM International Conference on Robotics and Mechatronics (ICRoM)*, Oct 2014, pp. 902–907.
- [627] D. Brujic, I. Ainsworth, and M. Ristic, “Fast and accurate nurbs fitting for reverse engineering,” *The International Journal of Advanced Manufacturing Technology*, vol. 54, no. 5, pp. 691–700, 2011. [Online]. Available: <http://dx.doi.org/10.1007/s00170-010-2947-1>
- [628] L. Ming and P. Vitányi, *An introduction to Kolmogorov complexity and its applications*. Springer Heidelberg, 1997.
- [629] A. H. Tan and K. R. Godfrey, “The generation of binary and near-binary pseudorandom signals: an overview,” *IEEE Transactions on Instrumentation and Measurement*, vol. 51, no. 4, pp. 583–588, Aug 2002.
- [630] A. N. Ravari and H. D. Taghirad, “Unsupervised 3d object classification from range image data by algorithmic information theory,” in *2013 First RSI/ISM International Conference on Robotics and Mechatronics (ICRoM)*, Feb 2013, pp. 319–324.
- [631] L. A. Alexandre, “3d descriptors for object and category recognition: a comparative evaluation,” in *Workshop on Color-Depth Camera Fusion in Robotics at the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, Vilamoura, Portugal, vol. 1, no. 3, 2012, p. 7.
- [632] B. Okal and A. Nüchter, “Sliced curvature scale space for representing and recognizing 3d objects,” in *2013 16th International Conference on Advanced Robotics (ICAR)*, Nov 2013, pp. 1–7.
- [633] T. Lindeberg, “Scale-space theory,” 2001.
- [634] F. Mokhtarian and M. Bober, *Curvature Scale Space Representation: Theory, Applications, and MPEG-7 Standardization*. Norwell, MA, USA: Kluwer Academic Publishers, 2003.
- [635] H. Hoffmann, “Kernel {PCA} for novelty detection,” *Pattern Recognition*, vol. 40, no. 3, pp. 863 – 874, 2007. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0031320306003414>
- [636] D. D. Lee and H. S. Seung, “Algorithms for non-negative matrix factorization,” in *Advances in neural information processing systems*, 2001, pp. 556–562.
- [637] B.-G. Lee, Liliana, and D.-H. Shin, “Enhanced computational integral imaging system for partially occluded 3d objects using occlusion removal technique and recursive {PCA} reconstruction,” *Optics Communications*, vol. 283, no. 10, pp. 2084 – 2091, 2010. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0030401810000672>
- [638] C. Zang, K. Hashimoto, and J. Moon, “A visual tracking strategy using computer graphics and edge,” in *2011 IEEE International Conference on Robotics and Biomimetics*, Dec 2011, pp. 981–986.
- [639] T. Drummond and R. Cipolla, “Real-time visual tracking of complex structures,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 7, pp. 932–946, Jul 2002.
- [640] M. Muja, R. B. Rusu, G. Bradski, and D. G. Lowe, “Rein - a fast, robust, scalable recognition infrastructure,” in *2011 IEEE International Conference on Robotics and Automation*, May 2011, pp. 2939–2946.
- [641] M. Marques and J. Costeira, “Lamp: Linear approach for matching points,” in *2009 16th IEEE International Conference on Image Processing (ICIP)*, Nov 2009, pp. 2113–2116.
- [642] G. Sudhir, S. Banerjee, and A. Zisserman, “Finding point correspondences in motion sequences preserving affine structure,” *Computer Vision and Image Understanding*, vol. 68, no. 2, pp. 237 – 246, 1997. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S1077314297905452>
- [643] S. Salti, F. Tombari, and L. D. Stefano, “A performance evaluation of 3d keypoint detectors,” in *Proceedings of the 2011 International Conference on 3D Imaging, Modeling, Processing, Visualization and Transmission*, ser. 3DIMPVT '11. Washington, DC, USA: IEEE Computer Society, 2011, pp. 236–243. [Online]. Available: <http://dx.doi.org/10.1109/3DIMPVT.2011.37>
- [644] E. González, A. Adán, and V. Feliú, “2d shape representation and similarity measurement for 3d recognition problems: An experimental analysis,” *Pattern Recognition Letters*, vol. 33, no. 2, pp. 199 – 217, 2012. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0167865511003163>
- [645] Y. Guo, M. Bennamoun, F. Sohel, M. Lu, and J. Wan, “3d object recognition in cluttered scenes with local surface features: A survey,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 36, no. 11, pp. 2270–2287, Nov 2014.
- [646] Y. Guo, M. Bennamoun, F. Sohel, M. Lu, J. Wan, and N. M. Kwok, “A comprehensive performance evaluation of 3d local feature descriptors,” *Int. J. Comput. Vision*, vol. 116, no. 1, pp. 66–89, Jan. 2016. [Online]. Available: <http://dx.doi.org/10.1007/s11263-015-0824-y>
- [647] A. Johnson and M. Hebert, “Surface matching for object recognition in complex three-dimensional scenes,” *Image and Vision Computing*, vol. 16, pp. 635 – 651, 1998. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0262885698000742>
- [648] B. Taati and M. Greenspan, “Local shape descriptor selection for object recognition in range data,” *Computer Vision and Image Understanding*, vol. 115, no. 5, pp. 681 – 694, 2011, special issue on 3D Imaging and Modelling. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S1077314210002614>
- [649] F. Tombari, S. Salti, and L. Di Stefano, “Performance evaluation of 3d keypoint detectors,” *International Journal of Computer Vision*, vol. 102, no. 1, pp. 198–220, 2013. [Online]. Available: <http://dx.doi.org/10.1007/s11263-012-0545-4>
- [650] I. Sipiran and B. Bustos, “Harris 3d: a robust extension of the harris operator for interest point detection on 3d meshes,” *The Visual Computer*, vol. 27, no. 11, p. 963, 2011. [Online]. Available: <http://dx.doi.org/10.1007/s00371-011-0610-y>
- [651] Y. Zhong, “Intrinsic shape signatures: A shape descriptor for 3d object recognition,” in *2009 IEEE 12th International Conference on Computer Vision Workshops, ICCV Workshops*, Sept 2009, pp. 689–696.
- [652] A. Sharma, P. K. Singh, and P. Khurana, “Analytical review on object segmentation and recognition,” in *2016 6th International Conference - Cloud System and Big Data Engineering (Confluence)*, Jan 2016, pp. 524–530.
- [653] P. K. Singh, O. P. Sangwan, and A. Sharma, “A systematic review on fault based mutation testing techniques and tools for aspect-j programs,” in *2013 3rd IEEE International Advance Computing Conference (IACC)*, Feb 2013, pp. 1455–1461.
- [654] M. A. Treiber, *An Introduction to Object Recognition: Selected Algorithms for a Wide Variety of Applications*, 1st ed. Springer Publishing Company, Incorporated, 2010.
- [655] D. Hoiem and S. Savarese, *Representations and Techniques for 3D Object Recognition & Scene Interpretation*, 1st ed. Morgan & Claypool Publishers, 2011.
- [656] G. Taylor and L. Kleeman, *Visual Perception and Robotic Manipulation: 3D Object Recognition, Tracking and Hand-Eye Coordination*. Springer Publishing Company, Incorporated, 2014.

**ANEXO D – BIOARXIV ARTICLE- AUTOMATED MICROFOSSIL
IDENTIFICATION AND SEGMENTATION USING A DEEP LEARNING APPROACH**

Automated Microfossil Identification and Segmentation Using a Deep Learning Approach

Carvalho, L.E.^{1,*}, Fauth, G.³, Baecker Fauth, S.³, Krahl, G.³, Moreira, A. C.⁴,
Fernandes, C.P.⁴, von Wangenheim, A.²,

1 Graduate Program in Computer Science - Federal University of Santa Catarina - Campus João David Ferreira Lima - Trindade - Department of Informatics and Statistics - Room 320 Florianópolis/SC – CEP 88040-97.

2 Image Processing and Computer Graphics Lab - National Brazilian Institute for Digital Convergence.

3 itt Fossil – Instituto Tecnológico de Micropaleontologia, Universidade do Vale do Rio dos Sinos, Av. UNISINOS, 950, 93022-000 São Leopoldo, RS, Brazil.

4 Graduate Program in Materials Science and Engineering Federal University of Santa Catarina, Florianópolis, SC, Brazil

* icarvalho@incod.ufsc.br

Abstract

The applicability of computational analysis to paleontological images ranges from the study of the animals, plants and evolution of microorganisms to the simulation of the habitat of living beings of a given epoch. It also can be applied in several niches, such as oil exploration, where there are several factors to be analyzed in order to minimize the expenses related to the oil extraction process. One factor is the characterization of the environment to be explored. This analysis can occur in several ways: use of probes, extraction of samples for petrophysical components evaluation, the correlation with logs of other drilling wells and so on. In the samples extraction part the Computed Tomography (CT) is of importance because it preserves the sample and makes it available for several analyzes. Based on 3D images generated by CT, several analyzes and simulations can be performed and processes, currently performed manually and exhaustively, can be automated. In this work we propose and validate a method for fully automated microfossil identification and extraction. A pipeline is proposed that begins in the scanning process and ends in an identification process. For the identification a Deep Learning approach was developed, which resulted in a high rate of correct microfossil identification (98% of Intersection Over Union). The validation was performed both through an automated quantitative analysis based upon ground truths generated by specialists in the micropaleontology field and visual inspection by these specialists. We also present the first fully annotated MicroCT-acquired publicly available microfossils dataset.

Introduction

The applicability of computational image analysis to paleontological data encompasses the possibility of identifying, reconstructing and visualizing microfossils in rock samples not recovered by traditional extraction methodologies. It can also allow the taxonomic

1
2
3
4

microfossil identification even before the physical extraction from the rock sample. In addition, it is also pertinent to verify the microfossil position in a given sedimentary stratum, which can help in taxonomic inference, whereas detailed positional information is lost in the traditional preparation method [8]. Computational analysis of samples can be applied in several niches, e.g. oil exploration, habitat reconstruction and geology and paleontology research.

On the other side, in the oil exploration field, there are many factors to be taken into consideration in order to minimize oil prospecting expenses. One factor are the environmental conditions, which can be analyzed in multiple ways: use of probes, extraction of samples for petrophysical components evaluation and correlation with logs from other drilling wells.

In the area of samples extraction it is possible to perform different analyzes on a given sample. Here Computed Tomography (CT) plays a central role. More specifically, samples are analyzed with X-ray micro-tomography (MicroCT), which is a radiographic imaging technique that produces 3D images of the material's internal structure with a spatial resolution of around 1 micrometer [11]. MicroCT is of significance because it preserves the sample and makes it available for different studies. Based on MicroCT generated data volumes, various 3D data analyzes and simulations can be performed and several analysis processes can be computationally carried out and automated using state-of-the-art Computer Vision (CV) techniques. These processes are currently performed manually and in a time-consuming manner. One of these processes that can undergo automation through CV is the microfossils identification and localisation in rock samples, which is the focus of this study.

Objective and Strategy

In this work, we propose a CV workflow composed of computational methods that starts with the MicroCT scanning process of a sample and ends with the fully automated identification and localisation of individual microfossils in this sample. The main research question we try to answer is: *Is it possible to fully automatically and reliably identify microfossils in carbonatic rock samples?*

The novelty in our approach is the use of Deep Learning Convolutional Neural Network (CNN) approaches for the identification and 3D segmentation of microfossils directly in their deposition place. Our approach works directly on MicroCT data gained from carbonatic rocks, without the need of any preparation or physical extraction. For this purpose we developed an identification and segmentation strategy that employs a special category of CNN models, namely Semantic Segmentation (SS) neural networks and extends this model in order to be able to process whole 3D MicroCT sample volumes. In order to identify the best model, we extend, train, test and compare a series of different state-of-the art SS models. For the validation our approach we employ a validation strategy where we compare our results to ground truths that were manually generated by experienced micropaleontologists employing state-of-the-art automated image segmentation validation algorithms.

State of the Art

Paleontology is a well-established science and its methodological intersection with the computational field started to grow in the 1990's. In the late 1980's, most main paleontology journals still show an irregular presence of computational methods: some journal issues contained one article describing some computational method application, others presented 2 or 3 articles and very few offered a larger number of them [24]. In

the majority of journals and books the insertion of computational methods in the paleontology field still looked uneven.

In the late 1990s, however, with the widespread use of medical CT, a growth in research activities employing tomographic images occurred [24]. This boosted the development of specialized software applications such as: DRISHTI, SPIERS, SEG3D, IMAGEJ, MIMICS, VGSTUDIO MAX, AVIZO, AMIRA, Geomagic, Rhinoceros, Imaris, ITK-SNAP and TurtleSEG. These specialized tools helped change how researchers deal with specific problems in several fields, including geology and paleontology, frequently with applications to oil and gas exploration. The applicability of the set of tools and techniques that came to be called Virtual Paleontology (VP) range from animal, plant and microorganisms evolution analysis until the virtual reconstruction of a specific extinct environment [23].

On the other side, the application of the study of microfossils to the area of oil prospecting had its first appearance in 1890 in Poland [21], but it was in the USA, in 1920, with the use of microfossils to identify the age of probes extracted from drilling rigs, that a bigger advance in the development of the field of Applied Micropaleontology was attained [14].

In the last decade multiple research works contributed to improve the micropaleontology field. The latest efforts aim at the use of VP associated with CNNs in order to identify microfossils [5]. With this in mind, our research is focusing on pursuing techniques that can identify microfossils on their deposition place, *i.e.*, without the need of previous physical isolation. For this purpose we research some CV fields such as 3D segmentation applied to tomographic image and 3D object recognition, in order to apply them to microfossil identification.

In the next subsections we summarize the results of the systematic literature reviews (SLR) we performed in order to identify the state-of-the-art of the methods and procedures that potentially could be used in microfossil image studies. These reviews followed the approach originally proposed by [10] for SLRs in Computer Sciences, where first we defined a research question: *Is it possible to fully automatically and reliably identify microfossils in carbonatic rock samples?*. This broad question, in order to be more manageable, was split into 2 topics, each of which was explored in depth in a separate SLR:

- Analysis of 3D segmentation methods applied to tomographic images, which could possibly be used to segment microfossils [3];
- Analysis of methods used for 3D object recognition in a general context, aiming to evaluate which methods could be applied to the microfossils field [4].

The results of these two SLRs will be briefly summarized below. Since a detailed description would exceed the scope of this paper, we refer to the referenced SLRs for more details.

3D segmentation applied to tomographic images and 3D object recognition

An initial analysis of image processing methods employed in the fossil identification area showed difficulty in finding any works that explore microfossils. So we generalized our search for methods in other similar areas. We started by performing a systematic literature review on 3D segmentation methods applied to tomographic images [3]. Several works were analyzed which comprehended a vast group of segmentation methods. In our review, we noticed a tendency on the use of 3D segmentation methods based on models and region growing. However, its use for fossil/microfossil segmentation wasn't noticed in the literature.

We also analysed the field of 3D object recognition employing the same SLR methodology [4]. In this SLR for 3D object recognition we could identify two general pipelines. Both pipelines start with the data acquisition, which can basically vary between 3D data (MRI, CT) or 2D data (RGB and RGBD cameras), pre-processing, where methods for artifact removal, image enhancement and image simplification and data representation, wherein several authors proposed a varying amount of different object representations. Then, it comes the stage where both pipelines differ: In the first pipeline, the data representation stage is used to describe and storage the object representation chosen, which is later used for similarity calculation and object identification; In the second pipeline, the data representation is employed for training a specific recognition architecture, such as a CNN, which is afterwards used for other objects recognition. Despite having found two general approaches for 3D object recognition, we could not identify, in our review, the application of these approaches on fossil identification.

Deep learning, object recognition and paleontology

The 3D object recognition area has, in the last few years, experienced a growing boosted by the increased availability of new algorithms and models, 3D data and the popularization of a varied palette of 3D sensors. Methods developed in this area find application in a wide range of areas, from the field of robotics to the security and surveillance domain. The general tendency in this area has been the use of Deep Learning (DL) techniques.

DL is a form of machine learning that enables computers to learn from experience and understand the world in terms of a hierarchy of concepts [6]. DL employs very deep CNNs, with neural networks that sometimes consist of more than 100 layers, in contrast to the Artificial Neural Networks (ANNs) employed between the 1980's and 2000's, that typically employed only three layers. One key concept here is the Convolutional Layer (CL), a feature extraction structure, first presented in [12], that allows the hierarchical learning and representation of complex knowledge. Because DL CNNs gather knowledge from examples, there is no need for a human computer operator to formally specify all the knowledge that the computer needs. The capacity to represent a hierarchy of concepts in a network dozens of CLs deep allows a DL CNN to learn complicated concepts by building them out of simpler ones; a graph of these hierarchies would be many layers deep [6].

One work that employs DL for object recognition is the 3D Object Recognition with Deep Belief Nets approach [15], where a network of symmetrically connected neuron-like units, that performs stochastic decisions about whether to be on or off, is presented. In the Convolutional-Recursive Deep Learning for 3D Object Classification, where Socher [22] presents a model based on the combination of convolutional and recursive neural networks for the feature learning and classification in RGB-D images. Another approach is the Vision-based Robotic Grasping System Using Deep Learning for 3D Object Recognition and Pose Estimation, where Jincheng Yu [30] presents a robotic vision-based system, which can not only recognize different objects, but also estimate their pose through the use of a deep learning model. The deep learning model used is the Max-pooling Convolutional Neural Network (MPCNN). The 3D Object Recognition and Pose estimation System Using Deep Learning Method is an approach where Dong Liang [13] presents a 3D object recognition and pose estimation method using a deep learning model. Recognizing multi-view objects with occlusions using a deep architecture, where a method for efficient 3D object recognition with occlusion is presented [26]. In the Convolutional neural network for 3D object recognition based on RGB-D dataset, Jianhua Wang [25] employs a convolutional neural network model to learn features from a RGBD dataset which are given to a linear SVM to classify objects.

In the Convolutional Neural Network for 3D object recognition using volumetric representation Xiaofan Xu [29] presents an efficient 3D object volumetric representation, called Volumetric Accelerator (VOLA), which requires much less memory than a normal volumetric representation. Properly, VOLA can reduce the computational complexity of Convolutional Neural Networks (CNNs). None of these approaches tackles the problem of identification of fossils embedded in rocks or any remotely similar problem.

Material and Methods

This section describes our datasets and the CV approach we developed for fully automated microfossil identification and segmentation in carbonatic rock samples.

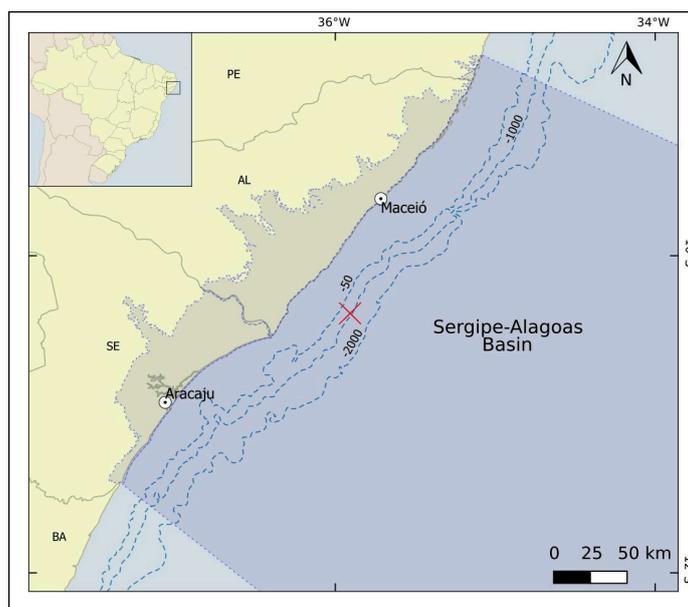


Figure 1. Sergipe Basin map with the exact rock samples extraction location marked with a red cross. The sample was collected at a depth of approximately 2,500 meters. Source: the authors.

Material

We employed two datasets: a scanned carbonatic rock sample obtained from a drilling rig probe and a set of manually isolated microfossil specimens that were afterwards obtained from this sample. The sample was collected at the Sergipe Basin Quaternary sediments (Fig. 1):

- A *carbonatic rock sample* was the material for which we developed our CV approach. The MicroCT scanner used to digitise the sample is a Versa XRM-500 (ZEISS/XRadia) with the following specifications: best resolution (pixel size) 0.7 μm , voltage 30-160 kV, power 2-10 W, CCD cameras 2048x2048 pixel, optical lenses 0.4X, 4X, 10X, 20X and 40X, a set of 12 filters for beam hardening correction, maximum sample mass capacity 15 kg and sample size limit (diameter / height) 80/300 mm. The sample acquisition parameters we employed are: Spatial resolution 1.08 mm, image size 956x1004x983, no filtering for beam correction hardening, 10X optical lens, 30 kV / 2W, angular step 0.255 (1600

projections) and exposure time 11 seconds. Figure 2 shows the rock sample and an excerpt of one slice of its digitised result.

- A set of *manually isolated microfossil specimens*, gained from the sample above, was employed in this work for illustration purposes and as a guide in order to allow us to know how the specimens in the *rock sample* would look like if cleanly segmented. These microfossils were prepared in the laboratory, following specific precautions so that there were no chemical and/or mechanical changes: (i) the sediment was first immersed in deionized water for approximately 24hrs, aiming the chemical disaggregation; (ii) then, it was washed with running water in a 63 μm sieve; (iii) next, the material was dried at 40degreeC for approximately 48 hours. After drying the samples, the main representative microfossils in the sample were selected through a magnifying glass. In this work, the microfossils specimens were stamped with the help of a multidimensional acquisition with the Zeiss Discovery V20 stereoscope (Z-stak mode in AxioVision 4.8 software). Figure 3 presents these microfossils.

The dataset containing the MicroCT data and and the manually segmented images annotated by specialists is available at:

<http://www.lapix.ufsc.br/microfossil-segmentation>

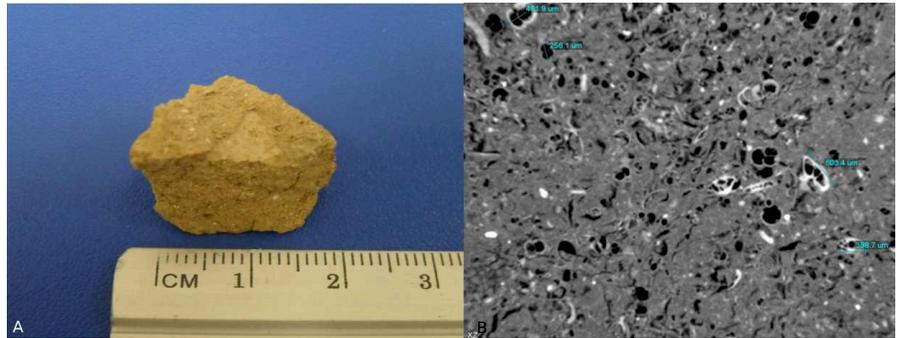


Figure 2. Analyzed rock sample (A) and one of its microtomography 2D sections image (B). Source: the authors.

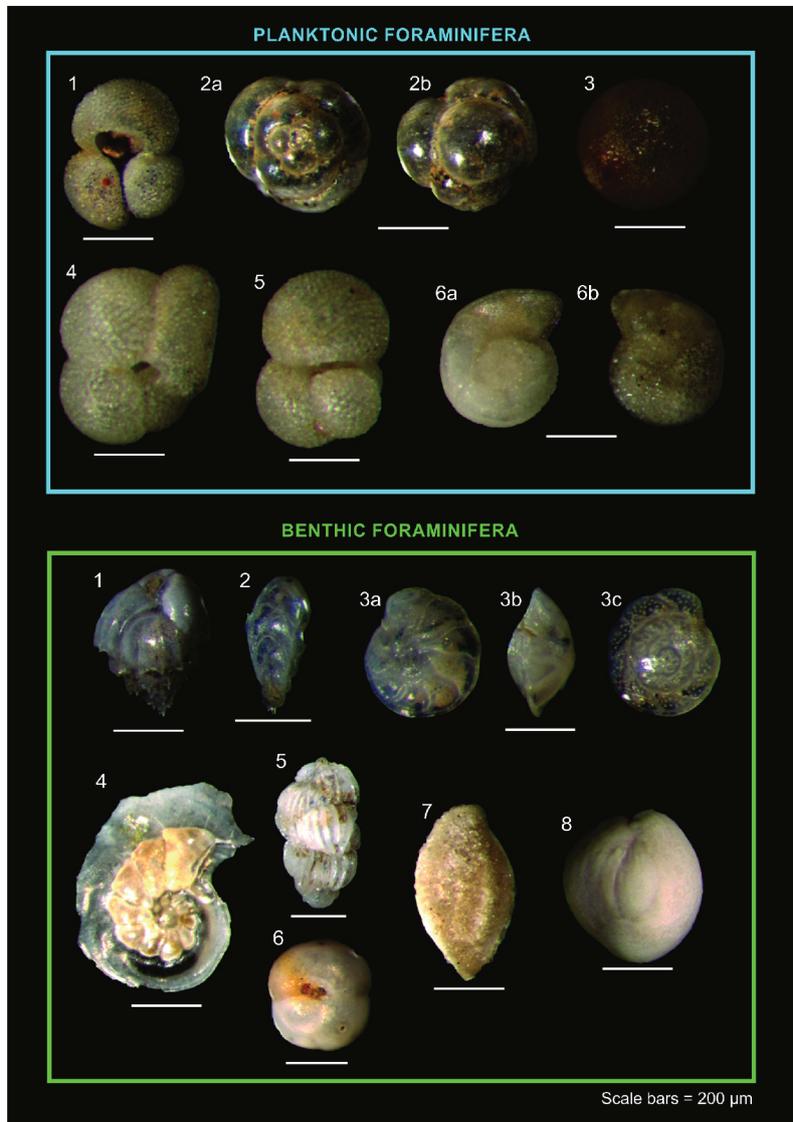


Figure 3. Analyzed foraminifera specimens. Planktonic Foraminifera: 1) *Globigerinoides ruber*; 2a-b) *Candeina nitida*; 3) *Orbulina universa*; 4) *Globigerinoides trilobus saculifera*; 5) *Globigerinoides trilobus*; 6a-b) *Globorotalia truncatulinoides*. Genus of benthic foraminifera: 1) *Bulinina*; 2a-c) *Bolivinita*; 3a-c) *Cibicidoides*; 4) *Laticarinina*; 5) *Uvigerina*; 6) *Sphaeroidina*; 7) *Siphonaperta*; 8) *Quinqueloculina*. Source: the authors.

Methods

The CV approach we present here is intended to be embedded into a broader workflow. Figure 4 presents a general overview of this workflow.

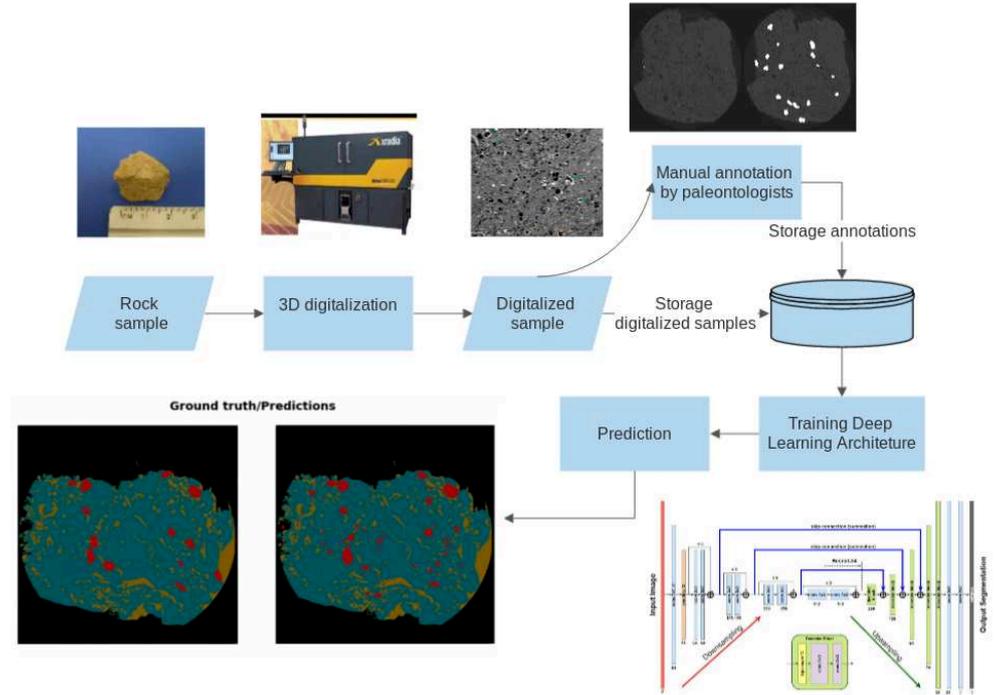


Figure 4. General workflow. Source: the authors and [18].

Non-CNN Computer Vision Methods

As part of a prospective search for CV methods for the microfossils segmentation, before we started investigating the use of CNNs, we performed a series of experiments using non-CNN, *i.e.* conventional CV methods for the segmentation of the MicroCT volume.

We analysed an extensive list of conventional CV algorithms, searching for a segmentation algorithm which, with the most appropriate input parameters, would potentially generate satisfactory results. We identified as interesting and selected the following classical segmentation algorithms: active contours [9], simple threshold and threshold with OTSU [16], all taking into account the complete tomographic volume. In order to find the best possible parameters for each segmentation algorithm, we performed a broad parameter values search running the algorithms with varied parameter sets. For the active contour algorithm, in order to find the best parameter set, we employed a genetic algorithm to search through possible input parameters. For this purpose we considered 5 input parameters: Number of steps, Sigma, Alpha, Smoothing and Theta, in a broad range of values.

The results of these conventional CV algorithms were initially analysed through *visual inspection*. For the conventional CV method that presented the best results to the visual inspection, we subsequently analysed its results also quantitatively employing the method described below.

CNN-based Segmentation Methods

In the 3D object identification and segmentation field, the most successfully and commonly used SS models in the last years have been the UNET and its variations. The UNET architecture was presented in [19], where the authors show its use for medical image segmentation. UNETs provide a general framework that can be parameterized with a specific *image classification* CNN model. The UNET then employs two slightly modified instances of this CNN, an *encoder* and a *decoder*, one for image recognition and another, employed in reverse mode, for the segment mask generation [1]: it uses the encoder to map raw inputs to feature representations and the decoder to take this feature representation as input, process it to make its decision and produce an output. As the UNET produces state of art semantic segmentation we choose it as our starting point.

In our work, we initially employed the UNET model associated with a ResNet34 [7], as our initial structure and added several state of art improvement methods. These methods were: nearest neighbour interpolation and pixel shuffling [20], Leaky Relu [28] for activation function and batchnorm [2] for batch normalization. This complete structure is available at the fastai¹ framework, which is a framework over Pytorch that contains several models, methods and state of art improvements.

Evaluation metrics

We evaluated each employed segmentation method comparing our results to the ground-truths generated by the micropaleontologists using the Intersection Over Union (IOU) [17] score, which quantifies similarity between finite sample sets, and is defined as the size of the intersection divided by the size of the union of the sample sets. The predicted labels were evaluated against a specialist generated ground truth.

Results

This section presents the results we obtained with the different algorithms and CNN models we tested.

Conventional CV algorithms

The best results under the conventional CV algorithms we obtained employing the active contours method. For this method we obtained an IOU score of 20%. The obtained active contour segmentation result is shown in figure 5.

The results we obtained indicated that conventional CV methods may not be indicated for the task of microfossil segmentation in rock samples.

CNN-based Semantic Segmentation

For our initial tests with SS CNN models, we started with the following structure: UNET associated with ResNet34 and the binary cross entropy as its loss function, a carbonatic rock sample with several microfossil specimens, scanned with the MicroCT previously described resulting in a total of 1000 slices. We employed an Intel Core i7-7700 CPU3.60GHz, 32GB memory computer and an NVIDIA GeForce GTX 1080 Ti 11GB GPU.

With this initial structure, our first experiment used only the microfossil annotation, performing a binary classification between microfossil and everything else. To improve

¹<https://www.fast.ai/>

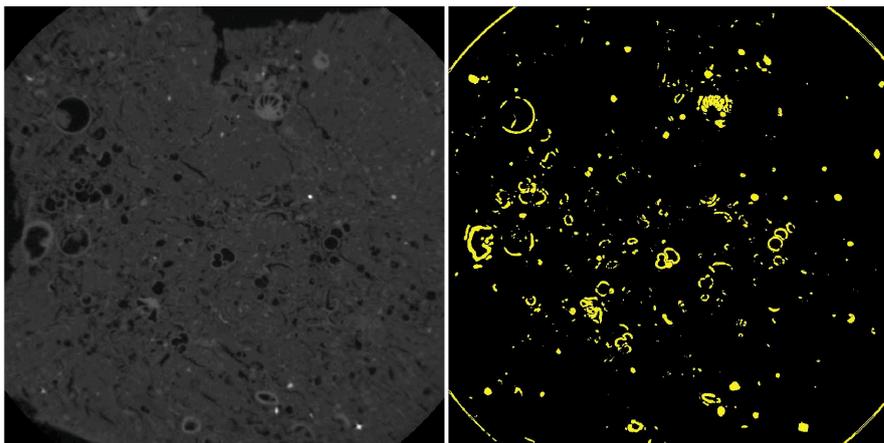


Figure 5. Best microfossil segmentation that we could obtain using 3D active contours (IOU = 20%). Source: the authors.

initial results some strategies such as data augmentation and transfer learning were applied aiming to minimize the effect of having a small database. However, the obtained IOU coefficient, used for the results evaluation, stopped in 40-45%. Trying to improve the obtained result, we increased the number of classes to four, which divided the everything else class into porous space, rock and background. With this number of classes, the obtained IOU value went from 40-45 % to 75-76 % and stopped. One problem with this approach is the data balance [31], *i.e.*, the existence, in the samples, of more annotations from the rock class in comparison with the microfossils. The figure 6 shows the result obtained after marking and training for the 4 classes setup for a selected slice.

Still using the 4-classes approach, we adjusted the hyper-parameters and applied a few performance-enhancing strategies [27], such as the *progressive input image resolution enhancing approach* (Jeremy Howards, informal communication during lecture at <https://course.fast.ai/videos/?lesson=1>) and explored data augmentation and batch size in order to obtain a 98% IOU. The microfossil GT and its resulting segmentation with this best IOU is shown in figure 7.

Our experiments resulted in an experimental environment, where we employed the UNET as base model associated with other models in the decoder part (resnet18, ResNet34, ResNet50, ResNet101), the Cross entropy as loss function and IOU for quality assessment. Table 1 shows the IOU value for each method employed and figure 8 shows the original image, its GT and the prediction results for all architectures we tested.

Method	IOU score
UNET + ResNet34	0.76
UNET + ResNet18 + hyper-parameter optimization	0.97
UNET + ResNet101 + hyper-parameter optimization	0.97
UNET + ResNet34 + hyper-parameter optimization	0.98
UNET + ResNet50 + hyper-parameter optimization	0.98

Table 1. Segmentation performance in terms of IOU value. Each method was evaluated in a set of 1000 images from annotate microfossil data.

After segmenting we took the predicted mask for resnet34 and applied to the original image. The result of this process is the easy identification of several microfossils. Figure 9 shows the mask overlap result, the identification of one microfossil specimen

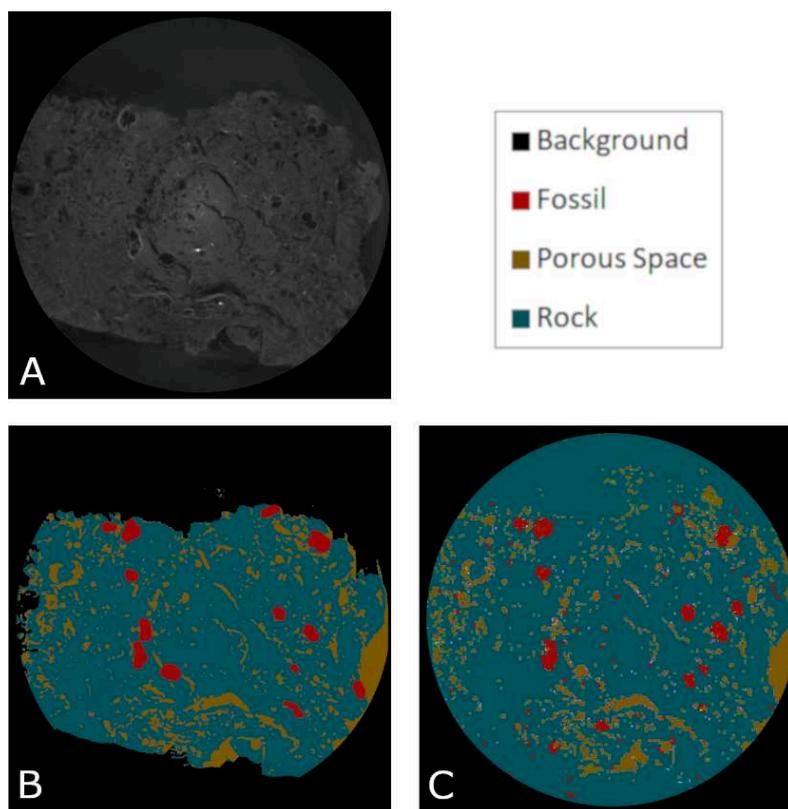


Figure 6. Obtained microfossil segmentation results with the 4-classes approach. (A) Original digitalised image. (B) Ground Truth manually generated by paleontologists. (C) UNET + ResNet34. Source: the authors.

(highlighted with the red rectangle) followed by its magnified version and the correlation of this the magnified version with other two versions of the same specimen (physically isolated and digitized with the Versa XRM-500 MicroCT and the Zeiss Discovery V20 stereoscope).

Discussion and Conclusions

In this paper we present a new nondestructive processing pipeline for the identification of microfossils in carbonatic rocks that allows for the fully automated segmentation of these fossils without the need of previous physical separation. Furthermore, we developed and validated the CV methods for this identification and segmentation. The validation was quantitatively and automatically performed against a ground truth manually generated by expert micropaleontologists.

An extremely relevant aspect of the developed pipeline for the field of paleontology, more specifically micropaleontology, resides in the nondestructive character of the method. In the micropaleontological study process an essential step is the samples preparation, aiming to separate the microfossils from the other rock and/or sediment constituents. In the traditional laboratory process, the samples are physically disaggregated (ground or milled) and subsequently performed chemical disaggregation, with addition of reagents (e.g., hydrogen peroxide and acetic acid). Both physical or chemical disaggregation can alter or even destroy microfossils characteristics. In this

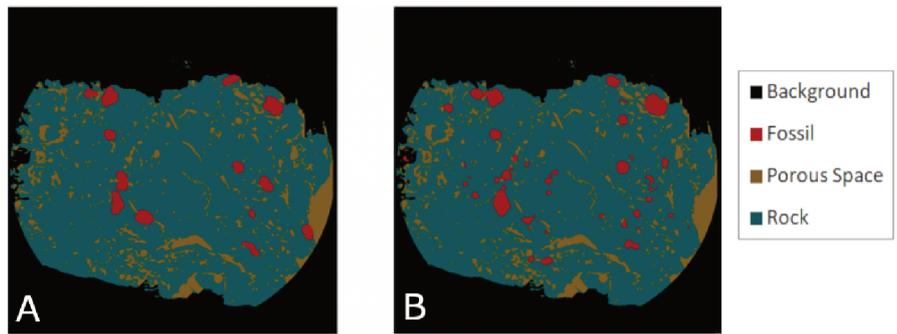


Figure 7. The ground truth (A) and the obtained microfossil segmentation result (B) with 4 classes, automated hyper-parameters search and additional data augmentation. Source: the authors.

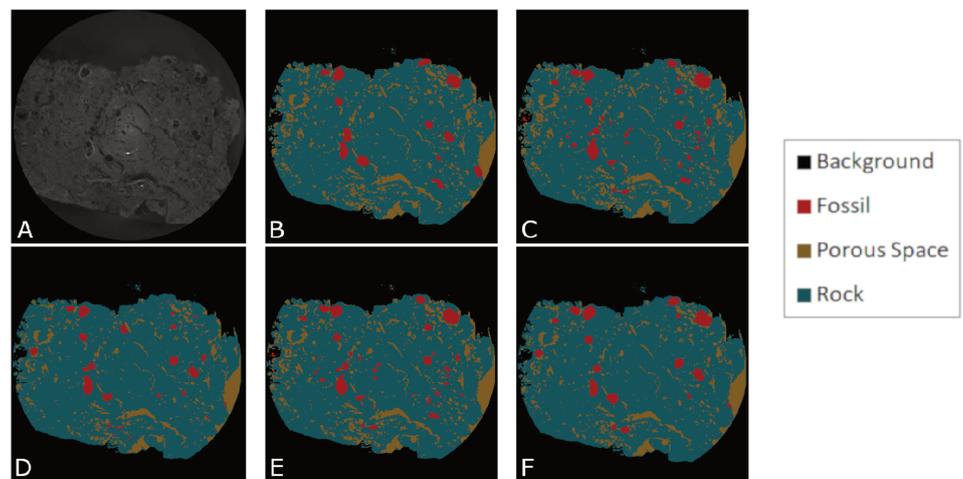


Figure 8. (A) Original digitalised image. (B) Ground Truth manually generated by paleontologists. (C) UNET + ResNet18 + hyper-parameter optimization. (D) UNET + ResNet101 + hyper-parameter optimization. (E) UNET + ResNet34 + hyper-parameter optimization. (F) UNET + ResNet50 + hyper-parameter optimization. Source: the authors.

premise, the imaging method is crucial for the morphological characteristics visualization as reliable as possible, allowing the individuals taxonomic recognition [8].

Another relevant factor that makes this method interesting is that it allows the microorganisms' preservation analysis throughout geological time, as well as aspects of fossilization, preservation and even position in which the microfossils are deposited (preserved) in the rocks. It should be emphasized that studies with the taphonomic approach are fundamental for paleoenvironmental conditions and/or diagenetic alteration processes reconstitution over geological time. Also, the use of this tool is strongly indicated in cases where it is extremely difficult to recover microfossils along specific sections and/or intervals where the material (rock) is very compact and even when it presents incipient diagenetic alteration. The microfossils identification is strategic for the exploration of petroleum due to the use in biostratigraphy, which refers to the use of microfossils from different groups to perform the temporal characterization of sedimentary rock strata, fundamental for the petroleum industry.

A few observations can be performed from the obtained results: (i) the importance of employing appropriate hyper-parameters such as learning rate, weight decay, momentum

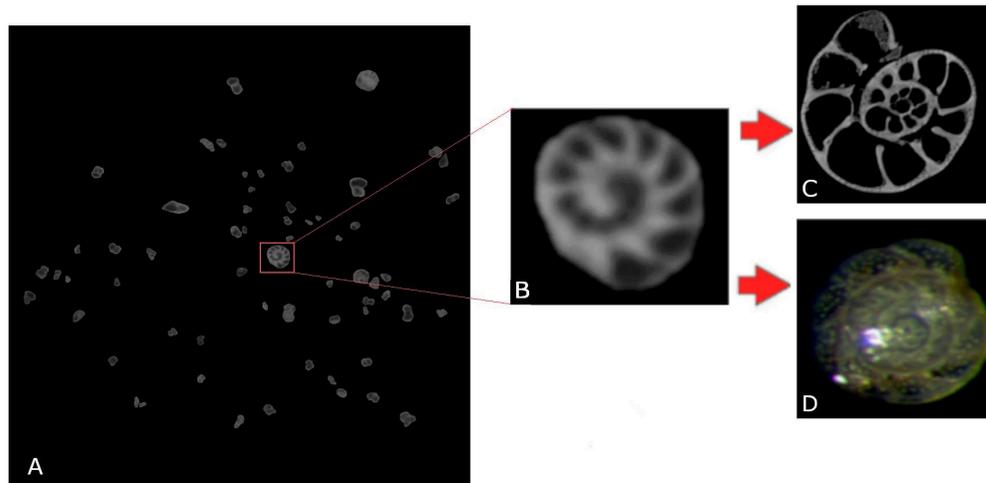


Figure 9. Result of applying the obtained segmentation mask over the digitized image. (A) contrast-enhanced 2D section image masked from the digitized MicroCT volume with one specific microfossil highlighted in red. (B) Highlighted microfossil extracted and magnified for visualisation. (C) Physically isolated microfossil digitized with The Versa XRM-500 MicroCT. (D) Cibicidoides multidimensional acquisition with the Zeiss Discovery V20 stereoscope. Source: the authors.

and batch size: with that hyper-parameters optimization we obtained an improvement of 20 %. (ii) a network architecture grown does not imply in better results. It is possible to observe that the ResNet34 shows the same results that the ResNet50 and a better result when compared with a ResNet101. However, here we have a hardware limitation: both, ResNet50 and ResNet101, couldn't run with the full image resolution on the 11 GB GeForce, even with a batch size of 1. Even so, the ResNet34 requires less execution time and hardware. (iii) Analyzing the obtained result images and comparing the visually against their Ground Truth (Figure 8), we still notice some small errors, however, we understand that this can be mitigated by adding more training samples, together with GTs from experts, to the training set when applying this pre-trained network to other, new, samples. Also, there are always new state of art improvements that could be tried aiming to reduce even more this small errors. Figure 9 shows the isolated microfossil digitalized and its correlated identification into the sample.

We understand that this process of microfossil identification without the need of physically isolate the microfossil has the potential to allow the paleontologist to analyze specific aspects of a sample such as the microfossils deposition. This is important for some applications in the oil and gas industry. It also has the potential to improve the paleontologist's work, because instead of losing time to physically isolate the microfossil he receives the microfossil already identified and can perform other analysis such as class identification and orientation.

Threats to validity

We employed a dataset that, even if it consisted of a very large quantity of images and presented a wide variety of microfossils, was gained from a sample obtained from a singular drill probe. On the other side, the samples digitisation and annotation afford a set of requirements such as: having a MicroCT working and available; the cost of the MicroCT digitisation process; a storage to keep the amount of generated data; and a paleontologist group to analyze and annotate each digitised sample. As the workflow we

suggest in this paper is new, it was not in place on any of the partners that participated in this work and to obtain more scanned and annotated samples was not possible at this point of our research.

This could jeopardize the generalizability of this work, as we have not enough data to claim that our approach will be successfully applicable to any carbonatic rock sample. On the other side, our identification and segmentation results were extremely successful and we understand that they are promising. From the authors' knowledge there is not any other publicly available carbonatic rock probe dataset, with or without specialist-annotated microfossils.

In this context, we understand our work as pioneering and pointing to a promising direction of research that can potentialize both, micropaleontological research and associated economical activities, such as oil prospection. Our publicly available fully annotated MicroCT database has also the potential to support research activities to be performed by other groups.

Conclusions

Summarizing, this work presents:

- the first fully annotated MicroCT-acquired publicly available microfossils dataset;
- a baseline for microfossil segmentation and the comparison with deep learning-based semantic segmentation and other segmentation architectures;
- a methodology for microfossil studies through MicroCT-acquired digital models;
- a tool for cases where it is extremely difficult to recover microfossils along specific sections.

With the improvement in the available hardware some future works aim to reduce even more the obtained errors by increasing the batch size and image resolution and employ more state of art deep learning improvements.

Acknowledgments

This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 001 and by PETROBRAS through the research project number 902. There are no conflicts of interest.

References

1. V. Badrinarayanan, A. Kendall, and R. Cipolla. Segnet: A deep convolutional encoder-decoder architecture for image segmentation. *IEEE transactions on pattern analysis and machine intelligence*, 39(12):2481–2495, 2017.
2. J. Bjorck, C. P. Gomes, and B. Selman. Understanding batch normalization. *CoRR*, abs/1806.02375, 2018.
3. L. E. Carvalho, A. C. Sobieranski, and A. von Wangenheim. 3d segmentation algorithms for computerized tomographic imaging: a systematic literature review. *Journal of Digital Imaging*, 31(6):799–850, Dec 2018.
4. L. E. Carvalho and A. von Wangenheim. 3d object recognition and classification: a systematic literature review. *Pattern Analysis and Applications*, Feb 2019.

-
5. Q. Ge, B. Zhong, B. Kanakiya, R. Mitra, T. Marchitto, and E. Lobaton. Coarse-to-fine foraminifera image segmentation through 3d and deep features. In *Computational Intelligence (SSCI), 2017 IEEE Symposium Series on*, pages 1–8. IEEE, 2017.
 6. I. Goodfellow, Y. Bengio, and A. Courville. *Deep Learning*. MIT Press, 2016. <http://www.deeplearningbook.org>.
 7. K. He, X. Zhang, S. Ren, and J. Sun. Deep residual learning for image recognition. *CoRR*, abs/1512.03385, 2015.
 8. S. Kachovich, J. Sheng, and J. C. Aitchison. Adding a new dimension to investigations of early radiolarian evolution. *Scientific Reports*, 9, 2019.
 9. M. Kass, A. Witkin, and D. Terzopoulos. Snakes: Active contour models. *International journal of computer vision*, 1(4):321–331, 1988.
 10. B. Kitchenham. Procedures for Performing Systematic Reviews. Technical report, Joint Technical Report TR/SE-0401, 2004.
 11. E. N. Landis and D. T. Keane. X-ray microtomography. *Materials characterization*, 61(12):1305–1316, 2010.
 12. Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner. Gradient-based learning applied to document recognition. In *Proceedings of the IEEE*, pages 2278–2324, 1998.
 13. D. Liang, K. Weng, C. Wang, G. Liang, H. Chen, and X. Wu. A 3d object recognition and pose estimation system using deep learning method. In *2014 4th IEEE International Conference on Information Science and Technology*, pages 401–404, April 2014.
 14. E. Molina and ed. *Micropaleontología (2ª edición)*. Prensas Universitarias de Zaragoza. Colección Textos Docentes, 2004.
 15. V. Nair and G. E. Hinton. 3d object recognition with deep belief nets. In *Proceedings of the 22Nd International Conference on Neural Information Processing Systems*, NIPS’09, pages 1339–1347, USA, 2009. Curran Associates Inc.
 16. N. Otsu. A threshold selection method from gray-level histograms. *IEEE transactions on systems, man, and cybernetics*, 9(1):62–66, 1979.
 17. M. A. Rahman and Y. Wang. Optimizing intersection-over-union in deep neural networks for image segmentation. In *International symposium on visual computing*, pages 234–244. Springer, 2016.
 18. A. Rakhlin, A. Davydow, and S. Nikolenko. Land cover classification from satellite imagery with u-net and lovász-softmax loss. In *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pages 257–2574. IEEE, 2018.
 19. O. Ronneberger, P. Fischer, and T. Brox. U-net: Convolutional networks for biomedical image segmentation. *CoRR*, abs/1505.04597, 2015.
 20. W. Shi, J. Caballero, F. Huszár, J. Totz, A. P. Aitken, R. Bishop, D. Rueckert, and Z. Wang. Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1874–1883, 2016.

-
21. A. Singh. Micropaleontology in petroleum exploration. In *7th International Conference and Exposition of Petroleum Geophysics*, pages 14–16, 2008. 426
427
 22. R. Socher, B. Huval, B. Bhat, C. D. Manning, and A. Y. Ng. Convolutional-recursive deep learning for 3d object classification. In *Proceedings of the 25th International Conference on Neural Information Processing Systems*, NIPS'12, pages 656–664, USA, 2012. Curran Associates Inc. 428
429
430
431
 23. M. Sutton, I. Rahman, and R. Garwood. *Techniques for Virtual Palaeontology*. New Analytical Methods in Earth and Environmental Science. Wiley, 1 edition, 2014. 432
433
434
 24. J. C. Tipper. Computer applications in paleontology: Balance in the late 1980s? *Computers & Geosciences*, 17(8):1091 – 1098, 1991. 435
436
 25. J. Wang, J. Lu, W. Chen, and X. Wu. Convolutional neural network for 3d object recognition based on rgb-d dataset. In *2015 IEEE 10th Conference on Industrial Electronics and Applications (ICIEA)*, pages 34–39, June 2015. 437
438
439
 26. Y. Xia, L. Zhang, W. Xu, Z. Shan, and Y. Liu. Recognizing multi-view objects with occlusions using a deep architecture. *Information Sciences*, 320:333 – 345, 2015. 440
441
442
 27. J. Xie, T. He, Z. Zhang, H. Zhang, Z. Zhang, and M. Li. Bag of tricks for image classification with convolutional neural networks. *arXiv preprint arXiv:1812.01187*, 2018. 443
444
445
 28. B. Xu, N. Wang, T. Chen, and M. Li. Empirical evaluation of rectified activations in convolutional network. *CoRR*, abs/1505.00853, 2015. 446
447
 29. X. Xu, A. Dehghani, D. Corrigan, S. Caulfield, and D. Moloney. Convolutional neural network for 3d object recognition using volumetric representation. In *2016 First International Workshop on Sensing, Processing and Learning for Intelligent Machines (SPLINE)*, pages 1–5, July 2016. 448
449
450
451
 30. J. Yu, K. Weng, G. Liang, and G. Xie. A vision-based robotic grasping system using deep learning for 3d object recognition and pose estimation. In *IEEE International Conference on Robotics and Biomimetics, ROBIO 2013, Shenzhen, China, December 12-14, 2013*, pages 1175–1180, 2013. 452
453
454
455
 31. W. Zhu, Y. Huang, H. Tang, Z. Qian, N. Du, W. Fan, and X. Xie. Anatomynet: Deep 3d squeeze-and-excitation u-nets for fast and fully automated whole-volume anatomical segmentation. *CoRR*, abs/1808.05238, 2018. 456
457
458