



UNIVERSIDADE FEDERAL DE SANTA CATARINA
CENTRO TECNOLÓGICO
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA ELÉTRICA

Lennon José Gazola dos Santos

**Ferramenta para registro e extração de atributos cinemáticos e morfológicos de vídeos
para detecção de padrões de atividades domésticas em humanos**

FLORIANÓPOLIS

2020

Lennon José Gazola dos Santos

**Ferramenta para registro e extração de atributos cinemáticos e morfológicos de vídeos
para detecção de padrões de atividades domésticas em humanos**

Dissertação submetida ao Programa de Pós-Graduação
de Engenharia Elétrica da Universidade Federal de
Santa Catarina para a obtenção do título de mestre em
engenharia elétrica

Orientadora: Prof^ª. Dra. Daniela Ota Hisayasu Suzuki

Coorientador: Prof. Dr. José Marino Neto

Florianópolis

2020

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

Gazola dos Santos, Lennon José
Ferramenta para registro e extração de atributos
cinemáticos e morfológicos de vídeos para detecção de padrões
de atividades domésticas em humanos / Lennon José Gazola
dos Santos ; orientadora, Daniela Ota Hisayasu Suzuki,
coorientador, José Marino Neto, 2021.
118 p.

Dissertação (mestrado) - Universidade Federal de Santa
Catarina, Centro Tecnológico, Programa de Pós-Graduação em
Engenharia Elétrica, Florianópolis, 2021.

Inclui referências.

1. Engenharia Elétrica. I. Suzuki, Daniela Ota
Hisayasu. II. Marino Neto, José. III. Universidade Federal
de Santa Catarina. Programa de Pós-Graduação em Engenharia
Elétrica. IV. Título.

Lennon José Gazola dos Santos

**Ferramenta para registro e extração de atributos cinemáticos e morfológicos de vídeos
para detecção de padrões de atividades domésticas em humanos**

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

Prof. Renato Garcia Ojeda, Dr.
Universidade Federal de Santa Catarina

Prof. Jefferson Luiz Brum Marques, Ph.D
Universidade Federal de Santa Catarina

Prof. Rafael Krummenauer, Dr.
Universidade Estadual de Maringá

Prof. Cesar Ramos Rodrigues, Dr. (suplente)
Universidade Federal de Santa Catarina

Certificamos que esta é a **versão original e final** do trabalho de conclusão que foi julgado
adequado para obtenção do título de Mestre em Engenharia Elétrica

Prof. Telles Brunelli Lazzarin, Dr.
Coordenador do Programa

Prof^a. Daniela Ota Hisayasu Suzuki, Dra.
Orientadora

Prof. José Marino Neto, Dr.
Coorientador

Florianópolis, 22 de Dezembro de 2020.

Este trabalho é dedicado à minha mãe, que sempre me apoiou e nunca mediu esforços para me fazer chegar até aqui.

AGRADECIMENTOS

Agradeço primeiramente a Deus pelo dom da vida, por nunca me deixar desistir e pela oportunidade de concluir mais um dos desafios no caminho que escolhi trilhar. Ele sempre esteve comigo, nos momentos bons e ruins, me guiou e me permitiu chegar até aqui, sem Ele eu não seria nada.

Agradeço a minha família, a minha irmã e meu padrasto pelo apoio e por sempre estarem ao meu lado nesses anos, nos dias felizes e tristes. A minha mãe que sempre me apoiou, em todas as minhas escolhas, me ensinando o caminho do bem e que o esforço de hoje será a recompensa do amanhã, sem ela eu não chegaria até aqui.

Agradeço aos meus amigos do mestrado, que se tornaram minha família em Florianópolis, Eduardo, Gian e Suzana, obrigado pela amizade sincera, por todos os conhecimentos trocados e principalmente pela cumplicidade, passamos por momentos difíceis juntos e também comemoramos juntos o sucesso. Eu admiro cada um de vocês de forma particular, vocês fazem parte da minha história. Nessa reta final, agradeço em especial o apoio que eu Suzana tivemos um com o outro, fez toda a diferença e me deu forças para concluir este trabalho, obrigado Suh.

Agradeço a todos os meus professores do mestrado, em especial ao professor Jefferson e a professora Daniela, pelos conhecimentos compartilhados e por indicarem o caminho para encontrar ainda mais conhecimento.

Agradeço ao meu orientador José Marino Neto por conduzir o andamento deste trabalho, por me orientar e compartilhar uma infinidade de ensinamentos, que levarei não só para a carreira acadêmica, mas também para a vida.

Agradeço ao Instituto de Engenharia Biomédica por permitir a realização deste projeto, fornecendo a estrutura necessária.

Agradeço a Universidade Federal de Santa Catarina pela oportunidade e pela qualidade na estrutura de ensino.

Agradeço a CAPS por financiar os materiais da pesquisa.

Agradeço a todos os que fizeram parte e contribuíram de alguma forma para a conclusão deste trabalho.

RESUMO

O estudo das características retiradas de imagens tem sido de grande importância para a criação de sistemas de identificação e predição de padrões, com diversas aplicações na área da saúde. O Instituto de Engenharia Biomédica da Universidade Federal de Santa Catarina busca o desenvolvimento de uma plataforma de saúde ubíqua voltado para o monitoramento de idosos e pessoas com deficiências. Como parte desta plataforma, o presente trabalho teve como objetivo geral propor e implementar uma ferramenta fundamentada em plataformas de prototipagem eletrônica e de programação em código aberto, para extração de atributos cinemáticos e morfológicos de humanos em ambientes domésticos, baseados em algoritmos de visão computacional. Para tanto, foi implementado um sistema para aquisição e armazenamento de amostras de vídeos, obtidos de humanos em ambiente doméstico (cozinha). O hardware do sistema de aquisição de vídeos contou com um Raspberry Pi e uma câmera USB, instalados em um laboratório (simulador de ambiente doméstico cozinha). A ferramenta de extração de atributos cinemáticos e morfológicos foi implementada utilizando programação de código aberto, com a criação de uma interface para o usuário. Para realizar a aquisição de amostras de vídeos foi desenvolvido um projeto piloto, submetido ao comitê de ética humano, visando formalizar os procedimentos diante de todos os padrões éticos da pesquisa. O hardware de aquisição de amostras foi testado, avaliado e validado com 7 voluntários executando situações de atividade humana real em um ambiente doméstico do tipo cozinha (laboratório simulador), obtendo 29 amostras de vídeo, que foram subdivididas em 116 amostras de atividades domésticas específicas. A ferramenta consiste em três interfaces diferentes: uma para visualização e corte das amostras, que permite ao usuário subdividir as amostras de forma prática; uma para detectar o rosto da pessoa no vídeo e aplicar uma proteção de imagem, que consiste em um borrão na face (*blurring*); e uma interface para extrair os atributos cinemáticos e morfológicos do humano detectado em cena. Foram realizados testes para validação do algoritmo de cada atributo. Foi possível extrair mais de 20 atributos diferentes, utilizando filtros e técnicas de morfologia matemática para reduzir possíveis ruídos. Todos os atributos selecionados pelo usuário são extraídos e salvos em um arquivo no formato .CSV. Foi construído um banco de amostras para testes de conceito, utilizando 15 dos vídeos obtidos, realizando cortes e separando em cinco atividades diferentes: abrir\fechar geladeira, cozinhar, comer e lavar-louça. Para cada vídeo de cada atividade foram extraídos todos os atributos disponíveis na ferramenta, e comparados os resultados entre amostras de um mesmo voluntário utilizando quatro classes no método *k-means*, obtendo resultados que prenunciam que esses dados são promissores para serem utilizados como entrada em redes neurais artificiais classificadoras.

Palavras-chave: Reconhecimento de atividades, extração de características, plataforma de saúde ubíqua, análise de movimento de imagem, detecção humana.

ABSTRACT

The study of the characteristics taken from images has been of great importance for the creation of systems for the identification and prediction of patterns, with several applications in the health area. The Institute of Biomedical Engineering at the Federal University of Santa Catarina seeks to develop a ubiquitous health platform aimed at monitoring the elderly and people with disabilities. As part of this platform, the present work had as general objective to propose and implement a tool, based on electronic prototyping and open source programming platforms, for extracting kinematic and morphological attributes from humans in domestic environments, based on computer vision algorithms. Therefore, a system for the acquisition and storage of video samples, obtained from humans in a domestic environment (kitchen), was implemented. The hardware for the video acquisition system included a Raspberry Pi and a USB camera, installed in a laboratory (kitchen home environment simulator). The tool for extracting kinematic and morphological attributes was implemented using open source programming, with the creation of an interface for the user. To perform the acquisition of video samples, a pilot project was developed, submitted to the human ethics committee, aiming to formalize the procedures before all the ethical standards of the research. The sample acquisition hardware was tested, evaluated and validated with 7 volunteers performing situations of real human activity in a domestic kitchen-type environment (simulator laboratory), obtaining 29 video samples, which were subdivided into 116 samples of specific domestic activities. The tool consists of three different interfaces: one for viewing and cutting the samples, which allows the user to subdivide the samples in a practical way; one to detect the person's face in the video and apply image protection, which consists of a blurring on the face; and an interface to extract the kinematic and morphological attributes of the human detected on the scene. Tests were performed to validate the algorithm of each attribute. It was possible to extract more than 20 different attributes, using filters and mathematical morphology techniques to reduce possible noise. All attributes selected by the user are extracted and saved to a file in .CSV format. A database sample was built for concept tests, using 15 of the videos obtained, making cuts and separating them into five different activities: opening/closing the refrigerator, cooking, eating and washing dishes. For each video of each activity, all the attributes available in the tool were extracted, and the results were compared between the same volunteer, using four classes in the k-means method, obtaining results that foresee that these data are promising to be used as input into artificial neural networks. classifiers

Keywords: *Activity recognition, feature extraction, ubiquitous health platform, image movement analysis, human detection.*

LISTA DE FIGURAS

Figura 1 - Desenho geral da plataforma em desenvolvimento	17
Figura 2 - Representação da conectividade de um pixel.....	23
Figura 3 - Etapas do processamento digital de imagens	25
Figura 4 - Exemplo de valores numéricos para pixels	27
Figura 5 - Representação numérica de pixels em RGB	27
Figura 6 - Pré-processamento com filtro Gaussiano.....	28
Figura 7 - Representação ideal de uma borda em degrau	29
Figura 8 - Cronologia de atividades.....	36
Figura 9 - Componentes do Raspberry Pi 3 Model B.....	39
Figura 10 - Protótipo do Módulo Multissensorial	40
Figura 11 - LifeCam HD-3000	40
Figura 12 - Diagrama de fluxo da Interface de Aquisição de Dados.....	42
Figura 13 - Diagrama de fluxo da Interface de Cortar e Salvar Vídeos.....	44
Figura 14 - Diagrama de fluxo da Interface de Detecção e <i>Blurring</i> Facial	46
Figura 15 - Representação da coleta e divisão de amostras no Registro.	51
Figura 16 - Planta baixa do laboratório com equipamento de coleta de dados.	54
Figura 17 - Campo de visão da câmera utilizada.....	55
Figura 18 - Procedimento de proteção de privacidade facial dos participantes gravados em vídeo.....	56
Figura 19 - Raspberry Pi com a Case feita em impressora 3D.....	58
Figura 20 - Tela de configuração e aquisição de dados	59
Figura 21 - Ajuste do parâmetro global <i>Acquisition Rate</i>	60
Figura 22- Ajuste do parâmetro <i>Audio Sampling Frequency</i>	60
Figura 23 - Controle do parâmetro <i>Sensors</i>	61
Figura 24 - Ajuste do parâmetro Câmera Index	61
Figura 25 - Ajuste do parâmetro <i>Resolution</i>	62
Figura 26 - Ajuste do parâmetro <i>Video Format</i>	62
Figura 27 - Botões para aplicar mudanças, iniciar ou finalizar uma aquisição	63
Figura 28 - Tela principal da VSST	63
Figura 29 - Tela principal da FBDT	66

Figura 30- Frame com faces detectadas.....	68
Figura 31 - Tela principal da KMFET	69
Figura 32 - Opções de métodos de extração de fundo	71
Figura 33 – Opções de atributos para serem extraídos do objeto de interesse.....	71
Figura 34 - Opções de técnicas de morfologia matemática e os seus respectivos <i>Kernels</i>	75
Figura 35 - Dados no arquivo.CSV	76
Figura 36 - Extração de fundo com detecção do contorno do objeto	78
Figura 37 – Representação da subdivisão de vídeos do roteiro principal em quatro atividades	81
Figura 38 - Organização do diretório das subamostras.....	81
Figura 39 – Frame com o resultado da subtração de fundo utilizando KNN para a atividade "cozinhar"	82
Figura 40 - Imagem de referência utilizada para o método SSIM.....	83
Figura 41 – Frame com o resultado da subtração de fundo utilizando SSIM para a atividade "cozinhar"	83
Figura 42 - Correção de ruídos utilizando técnicas de morfologia matemática	85
Figura 43 - Resultado da remoção de ruídos utilizando a técnica "fechamento", para a atividade “comer”.	85
Figura 44 - Resultado da remoção de ruídos utilizando a técnica "fechamento" para a atividade “lavar louça”.	86
Figura 45 - Resultado da remoção de ruídos utilizando a técnica "fechamento" para a atividade “abrir\fechar geladeira”.	87
Figura 46 - Resultado da remoção de ruídos utilizando a técnica "fechamento" para a atividade “abrir\fechar geladeira”.	87
Figura 47 - Organização de arquivos de teste no diretório	88
Figura 48 - Gráfico de dispersão de dados (<i>Scatter-plot</i>) dos resultados do método <i>K-means</i>	90
Figura 49 - Gráfico de dispersão de dados (<i>Scatter-plot</i>) dos resultados do método <i>K-means</i>	91

LISTA DE TABELAS

Tabela 1 - Resultado dos testes de cortes de vídeos	65
Tabela 2 - Resultado dos testes de funcionamento do algoritmo de detecção de faces	68
Tabela 3 - Resultados do teste completo de extração de <i>features</i>	77
Tabela 4 - Fatores determinantes da iluminação adequada	78
Tabela 5 - Parametrização utilizada para extrair os atributos dos 20 vídeos de atividades	88

LISTA DE ABREVIATURAS E SIGLAS

ABNT Associação Brasileira de Normas Técnicas
AVI Audio Video Interleave
CCD Charge-coupled device
CEPHS Comitê de Ética em Pesquisa com Seres Humanos
CMOS Complementary Metal Oxide Semiconductor
FBDT Face Blur and Detection Tool
IEB-UFSC Instituto de Engenharia Biomédica
IOT Internet Of Things
JPEG Joint Photographic Experts Group
KMFET Kinematic and Morphological Feature Extract Tool
KNN K-nearest neighbours
MPEG Moving Picture Experts Group
PNG Portable Network Graphics
RAM Random Access Memory
RPi Raspberry Pi
SSIM Structural Similarity Index
TCLE Termo de Consentimento Livre e Esclarecido
TIFF Tagged Image File Format
UCL Unidade Concentradora Local
UFSC Universidade Federal de Santa Catarina
USM Unidade de Sensores Múltiplos
VSST Video Sample Split tool
WMV Windows Media Video

SUMÁRIO

1	INTRODUÇÃO	15
1.1	CONTEXTUALIZAÇÃO E JUSTIFICATIVA	15
1.2	OBJETIVOS	18
1.2.1	Objetivo Geral.....	18
1.2.2	Objetivos Específicos.....	18
2	FUNDAMENTAÇÃO TEÓRICA	20
2.1	Saúde Ubíqua e o monitoramento contínuo de saúde	20
2.2	Domótica e sistemas de supervisão de indivíduos em situações domésticas	21
2.3	Softwares e ferramentas de extração de atributos	22
2.4	Processamento digital de imagem	23
2.5	Etapas do processamento de imagem.....	24
2.5.1	Aquisição e digitalização de uma imagem	25
2.5.2	Pré-processamento	28
2.5.3	Segmentação	28
2.5.4	Detecção de bordas.....	29
2.5.5	Limiarização	29
2.5.6	Método de subtração de fundo.....	30
2.5.7	Pós processamento.....	30
2.5.8	Extração de atributos de objetos em cena	32
2.5.9	Descritores cinemáticos e morfológicos	32
2.5.10	Reconhecimento de padrões.....	33
2.6	Processamento de imagens aplicados na domótica e saúde	33
2.7	Ética e proteção de imagens de humanos	34
3	MATERIAL E MÉTODOS	35
3.1	Cronologia de atividades desenvolvidas	35

3.2	Plataforma de Saúde Ubíqua.....	36
3.3	Módulo de aquisição e processamento de dados	36
3.4	Hardware do Módulo de Aquisição de dados.....	37
3.4.1	Unidade concentradora Local.....	37
3.4.2	Módulo Multissensorial.....	39
3.4.3	Módulo de Câmera	40
3.5	Interface para Aquisição de Dados.....	40
3.6	Interfície para cortar e salvar vídeos.....	42
3.7	Interfície para detecção e <i>blurring</i> facial	44
3.8	Interface de Extração de Características Cinemáticas e Morfológicas	46
3.9	Coleta de dados	49
3.9.1	Protocolo para coleta de dados	49
3.9.2	Submissão do projeto ao comitê de ética humano (CEPH)	51
3.9.3	Laboratório de Testes.....	53
3.9.4	Proteção de privacidade das amostras coletadas	55
4	RESULTADOS.....	57
4.1	Instalação do <i>hardware</i> do sistema de aquisição de dados	57
4.2	Software de aquisição de áudio e vídeo	58
4.2.1	<i>Video Sample Split tool - VSST</i>.....	63
4.2.1.1	Overview da interface.....	63
4.2.1.2	Algoritmos utilizados	64
4.2.1.3	Testes de funcionamento	65
4.2.2	<i>Face Blur and Detection Tool - FBDT</i>.....	66
4.2.2.1	Overview da interface.....	66
4.2.2.2	Algoritmos utilizados	67
4.2.2.3	Testes de funcionamento	67
4.2.3	<i>Kinematic and Morphological Feature Extract Tool - KMFET</i>	69

4.2.3.1	Overview da interface.....	69
4.2.3.2	Algoritmos utilizados	76
4.2.3.3	Testes de funcionamento	77
4.2.4	Execução da Coleta de Dados	78
4.2.5	Preparação dos vídeos para construção do banco de dados	80
4.2.6	Utilização dos dados para encontrar padrões em atividades domésticas	80
4.2.7	Covid-19.....	91
5	DISCUSSÃO.....	93
6	CONCLUSÃO	97
	REFERÊNCIAS	99
	ANEXO A – Questionário utilizado para a coleta de dados.....	107
	ANEXO B – Termo de Consentimento Livre e Esclarecido	109
	ANEXO C – Parecer Consubstanciado do CEP	115

1 INTRODUÇÃO

Neste capítulo são apresentados a contextualização, justificativa e os objetivos gerais e específicos.

1.1 CONTEXTUALIZAÇÃO E JUSTIFICATIVA

O estudo das características retiradas de imagens e vídeos tem sido de extrema importância para a criação de dispositivos e sistemas de identificação e predição de situações, que podem colocar em risco a vida do ser humano. A possibilidade de acidentes domésticos com idosos que vivem sozinhos é uma preocupação relevante, especialmente em vista do contínuo aumento de expectativa de vida autônoma das populações humanas (e.g. LUBITZ, et.al, 2003). Assim, as possibilidades de criação de ambientes inteligentes têm sido examinadas por diversos autores (e.g (ABTOY, et.al, 2020),(BOZAN, BERGER, 2019)), visando melhorar a acessibilidade e usabilidade de projetos de monitoramento de idosos em tempo real, de forma discreta e não invasiva.

Um ambiente inteligente pode ser caracterizado como aquele que é capaz de obter e aplicar conhecimentos sobre o meio ambiente e sobre seus habitantes, com a intenção de melhorar a experiência naquele ambiente. O seu desenvolvimento e construção envolve uma variedade de disciplinas como inteligência artificial, robótica, sensores, Internet das Coisas (IOT - *Internet Of Things*), entre outras (COOK, et.al, 2007) (YOUNGBLOOD, et.al, 2005). Existem diferentes tipos de ambientes inteligentes; entre eles podemos citar as casas inteligentes, saúde inteligente, cidades inteligentes e indústrias inteligentes (Indústria 4.0), que claramente possuem objetivos de inteligência diferentes, voltados tanto para o pessoal (bem-estar e saúde) quanto para o comercial (negócios). As casas, por exemplo, são ambientes propícios para hospedar uma tecnologia inteligente, pois já existe um grande número de dispositivos tecnológicos sendo utilizados diariamente, mesmo que operando de forma individual; além disso a casa é um ambiente controlado, ou seja, o proprietário pode investir na automatização e promover a manutenção de forma adequada.

Os serviços de assistência, executados em uma casa inteligente, fornecem suporte ao usuário em suas atividades diárias, sendo elas domésticas (ex. acender lâmpadas, abrir e fechar janelas, ligar eletrodomésticos), relacionadas à saúde (ex. sensores de gás e incêndio) ou segurança (ex. sistemas de alarme). A base desses serviços são os mecanismos de automação residencial, que proporcionam o monitoramento de um bloco de construção completo (sala, cozinha, banheiro e etc) (GOMEZ, et.al, 2019).

Além do uso de sensores para captar dados ambientais (como som, temperatura, umidade e pressão), pode ser necessário a utilização de outros tipos de sensores comumente usados para obter os sinais vitais de uma pessoa (como pulso, temperatura, frequência respiratória e pressão arterial), realizando um acompanhamento médico mais específico, como por exemplo o caso de idosos com doenças cardíacas. A coleta desses sinais de forma constante e em tempo real pode ser um incômodo para um paciente, podendo atrapalhar sua liberdade de locomoção e conforto. Uma forma de suavizar esse desconforto é o uso de dispositivos vestíveis, também conhecidos como *wearables*, que utilizam os mesmos sensores de forma miniaturizada e comunicação sem fio realizando a transmissão de dados para algum terminal próximo do paciente para o eventual registro das informações captadas, seja em um ambiente doméstico ou hospitalar (AZODO, et.al, 2020) (WU, et.al, 2008) (TAPIA, et.al, 2007). A combinação de ambientes inteligentes e *wearables* pode resultar em ferramentas úteis como um dispositivo que detecta a queda de uma pessoa e o seu posicionamento no ambiente (CHEN, et.al, 2006).

Também é possível monitorar as interações de uma pessoa com o ambiente utilizando sistemas baseados em vídeo, de forma não invasiva, suavizando o desconforto de um monitoramento. Com apenas uma câmera por cômodo e aplicando técnicas de visão computacional nas imagens extraídas, pode-se detectar a presença de um indivíduo, obter dados como movimentação, posicionamento e até reconhecer algum comportamento, como a postura, por exemplo. Essas informações reunidas podem identificar diversas situações cotidianas e principalmente situações que representam algum tipo de risco (CUCCHIARA, et.al, 2004) (PANINI, CUCCHIARA, 2003).

Ao falar sobre o monitoramento, temos que abordar o conceito de saúde ubíqua, isto é, a coleta e análise de dados de dados relevantes para a saúde a todo tempo e em todos os lugares, que permitam, de forma contínua, análises, previsões e tomadas de decisões pelos agentes de saúde, visando uma mudança de paradigma em direção ao "gerenciamento do bem-estar", que incorpora a saúde social, física e mental (BROW, et.al, 2007).

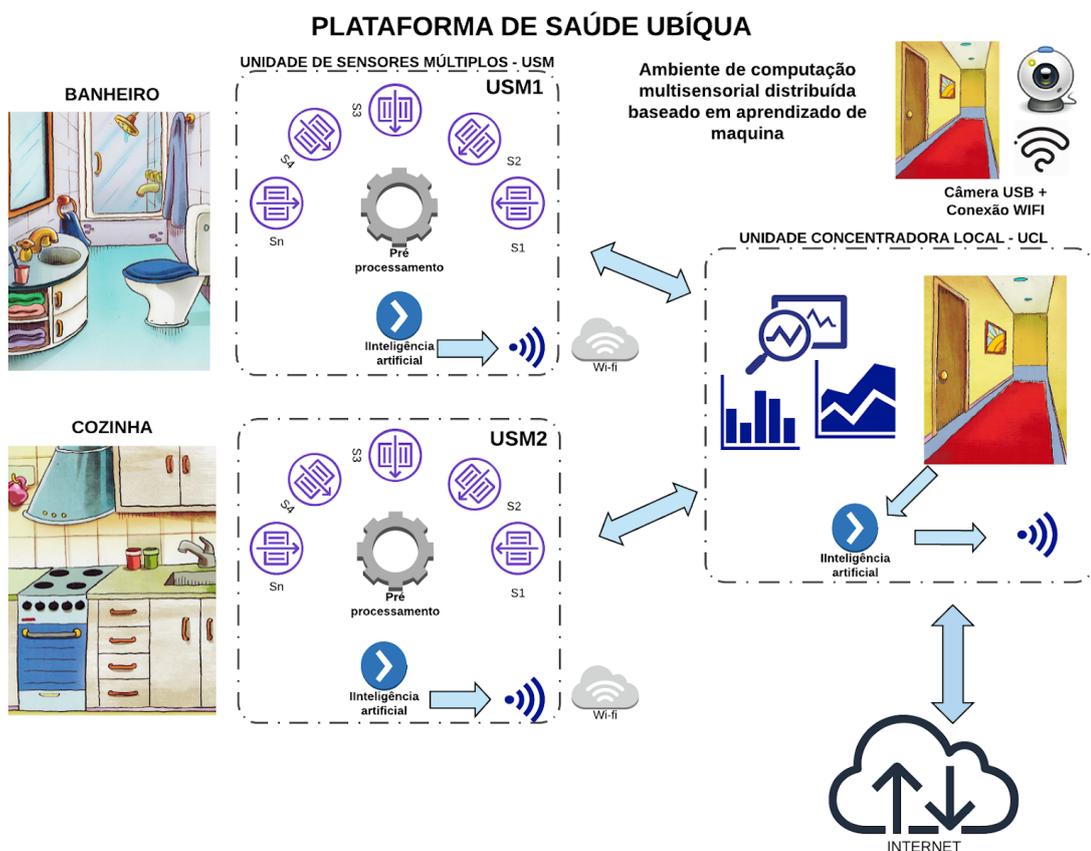
Uma plataforma de Saúde Ubíqua pode utilizar sensores ambientais múltiplos, agrupados e de baixo consumo de energia, coletando dados em tempo real, processando e enviando para uma unidade concentradora. Com o intuito de examinar e testar a viabilidade de obter dados, que são relevantes sobre uma situação doméstica, o Instituto de Engenharia Biomédica da UFSC (IEB-UFSC) está desenvolvendo uma plataforma de saúde ubíqua

voltada, entre outros aspectos, para a supervisão de atividades domésticas relevantes para a saúde e o bem-estar de indivíduos idosos ou com necessidades especiais.

Os sensores devem coletar dados como som, umidade, luminosidade e temperatura, e as características cinemáticas e morfológicas obtidas através da imagem (vídeo). Todos os dados devem ser processados com o auxílio de algoritmos de aprendizado de máquina, realizando diagnósticos de eventos domésticos, enquanto as informações devem ser distribuídas aos cuidadores, contando com um programa de código aberto.

Na Figura 1 contém um desenho geral da plataforma, onde diversas unidades de sensores múltiplos (USM), contendo sensores de iluminação, temperatura, umidade e microfones (S1, S2, etc.) enviam dados pré-processados in situ para uma Unidade Concentradora Local (UCL). A UCL comporta a recepção de imagens, ferramentas de análise e tomada de decisão por algoritmos de aprendizagem de máquina (IA). A distribuição de informações acontece via internet.

Figura 1 - Desenho geral da plataforma em desenvolvimento



Fonte: Laboratório de Bioengenharia IEB - UFSC (2019)

Como parte dessa Plataforma de Saúde Ubíqua, o presente trabalho buscou desenvolver uma ferramenta capaz de coletar características cinemáticas e morfológicas de indivíduos em situação doméstica, usando somente dados extraídos continuamente de vídeo. Esta ferramenta deverá ser articulada com outros componentes da plataforma (*e.g.*, destinados a extrair informações sonoras e de outras variáveis ambientais, a criar bancos de dados sobre as atividades típicas em um certo ambiente doméstico e processar estes dados usando algoritmos de aprendizado de máquina) (SIRIDAKIS, 2020). Estes sistemas de sensoriamento (os USMs, ver Figura 1), assim como os demais componentes da plataforma, deverão ser desenvolvidos utilizando plataformas de código aberto (tanto de hardware quanto de software). Esta opção de desenvolvimento justifica-se pelo fato de que outras pessoas possam utilizar, redistribuir e modificar, evitando sua descontinuação. Uma plataforma de código aberto pode contribuir com diversos projetos no meio acadêmico, principalmente aqueles relacionados à saúde.

1.2 OBJETIVOS

Nas seções 1.2.1 e 1.2.2 estão descritos o objetivo geral e os objetivos específicos.

1.2.1 Objetivo Geral

O presente trabalho tem como objetivo geral propor e implementar uma ferramenta para registro e extração de atributos cinemáticos e morfológicos utilizados para detecção de padrões de atividades domésticas em humanos, fundamentado em plataformas de prototipagem eletrônica e de programação em código aberto.

1.2.2 Objetivos Específicos

Os objetivos específicos são fundamentais para alcançar o objetivo geral e estão divididos em três partes:

- 1) Implementar, usando hardware e software de código aberto um sistema para aquisição e armazenamento de amostras de imagem obtidas de humanos realizando atividades em ambiente doméstico (cozinha).
- 2) Implementar no sistema desenvolvido no objetivo 1, usando software de código aberto, um conjunto de algoritmos de extração de descritores morfológicos e cinemáticos a partir das imagens obtidas, de relevância para a

detecção de padrões comportamentais e de localização na situação doméstica acima indicada.

- 3) **Testar, avaliar e validar** estas ferramentas em situações de atividade humana real em ambiente de cozinha.

2 FUNDAMENTAÇÃO TEÓRICA

Neste capítulo são apresentados e definidos diversos conceitos importantes para a contextualização deste trabalho. O conceito de saúde ubíqua e domótica são apresentados para enfatizar a aplicação da visão computacional e o impacto no monitoramento da saúde de um indivíduo. São abordados o processamento digital de imagem e as técnicas utilizadas para realizar etapas como: aquisição de imagem, pré-processamento, transformações morfológicas na imagem, pós processamento, extração de atributos e detecção de objetos em cena. O conceito de ética e proteção de imagem também é abordado para evidenciar a importância de realizar um projeto dentro das normas éticas, para a preservação da privacidade de um voluntário humano.

2.1 Saúde Ubíqua e o monitoramento contínuo de saúde

A saúde ubíqua é um paradigma de acesso onipresente a serviços de saúde. Sua preocupação está na manutenção e melhoramento de serviços de saúde implementados através da disponibilização de dispositivos inteligentes no ambiente (CASTAÑEDA, et.al, 2011). Uma operação de saúde ubíqua deve se infiltrar de forma sutil nas atividades do dia-a-dia, obtendo medidas relevantes às condições de saúde através de dispositivos de monitoramento, comunicar dados através de uma rede e utilizar a informação adquirida para prover serviços de saúde, o que requer a incorporação de diferentes tecnologias (SOHN, HAHN, LEE, 2006).

O uso de tecnologias computacionais como a computação ubíqua contribui para melhorar o cuidado e a atenção em domicílios. Através de sensores utilizados no ambiente domiciliar, o paciente pode ser monitorado de forma contínua e em qualquer cômodo da casa. Em associação ao conceito de *Smart Homes*, plataformas de Saúde Ubíqua possibilitam o monitoramento e intervenção em ambientes domésticos, além de dados essenciais para a tomada de decisões informadas nas ações de saúde individual e coletiva (DAWADI, COOK, 2016).

Idosos e pessoas com deficiências que dependem de cuidadores para sobreviver, e necessitam de monitoramento constante, podem ter sua saúde mental afetada pelo fato de se sentir um fardo para a família e para a sociedade. A saúde ubíqua é vista como uma oportunidade de fornecer a devida independência, privacidade e um alcance de cuidados que vai além ao da saúde tradicional, realizando monitoramentos constantes e até diagnósticos através de sinais comportamentais como os fisiológicos (BARDRAM, MATIC, 2020).

2.2 Domótica e sistemas de supervisão de indivíduos em situações domésticas

O termo domótica, do francês *Domotique*, é usado na robótica e definido como a junção de mecanismos automáticos de um espaço, visando simplificar o cotidiano das pessoas em diversos aspectos como segurança e conforto. Automatizar um espaço doméstico de forma abrangente, controlando pontos como a iluminação, dispositivos de segurança, de comunicação, de lazer e de conforto. Atualmente, existem sistemas domóticos utilizados para a supervisão de pessoas em suas residências, voltados inclusive para o cuidado da saúde. Esse tipo de sistema pode proporcionar qualidade de vida para pessoas com deficiências e idosos, que precisam morar sozinhos ou dependem de cuidadores, melhorando aspectos como conforto, autoestima e estado psicológico (PONCE, et.al, 2019) (LOPEZ, et.al 2016) (GHAZAL, AL-KHATIB, 2015).

Um sistema composto pela combinação de câmeras de vídeo e sensores ambientais, distribuídos pelos cômodos de uma casa, pode ser usado para identificar diversos comportamentos em atividades domésticas diárias (*e.g* usar a cozinha na preparação de uma refeição). Com a análise de vídeo é possível detectar a localização de uma pessoa no ambiente, e identificar a postura naquele exato momento comparando com modelos pré-definidos (ZOUBA, et.al, 2010).

Definir um modelo de contexto orientado a eventos permite analisar comportamentos comuns de um ser humano em um certo ambiente (como estar na cozinha, sentado e comendo algo) e diferenciar de comportamentos emergenciais (como uma queda ou desmaio), que no caso de idosos pode representar fatalidade se houver demora no atendimento. É possível combinar diferentes modelos, como um modelo humano de eventos (*e.g.* cozinhar, assistir televisão, beber água) e um modelo humano de estados (*e.g.* sentado, andando, em pé), sincronizar e definir uma tomada de decisão sobre a situação (*e.g.* a pessoa está em pé bebendo água; a pessoa está sentada assistindo televisão) (CAO, et.al, 2009).

Existem normas técnicas que abordam sistemas de ambiente assistidos, como a norma alemã DIN SPEC 77002:2013-04, que especifica os requisitos necessários para esse tipo de serviço em uma casa ou uma área residencial. Existem requisitos gerais para que o sistema instalado aborde todas as normas de forma segura e com qualidade, como o desempenho do sistema (DIN SPEC, 2013)

2.3 Softwares e ferramentas de extração de atributos

A extração de características de uma imagem vem sendo facilitada com o auxílio de algoritmos e ferramentas de código aberto, sendo aplicada para extração de diversos tipos (*e.g.* objetos, animais, humanos, ambientes). Por outro lado, a quantidade de ferramentas que concentram esses algoritmos em uma interface não é tão escassa quanto os próprios algoritmos, dessa forma, o usuário precisa implementar com o código fonte para poder utilizar e obter os resultados esperados. Uma ferramenta com interface autoexplicativa pode tornar a utilização desses algoritmos mais prática e possibilitar o uso para pessoas de diferentes áreas.

Existem estudos de ferramentas computacionais para a extração de características cinemáticas e morfológicas através de imagem, que foram desenvolvidas por alunos de pós-graduação do laboratório de Bioengenharia da UFSC. O EthoWatcher® (distribuído no endereço eletrônico www.ethowatcher.ufsc.br) é uma ferramenta computacional distribuída de forma gratuita para apoiar a etografia detalhada, o *video-tracking* e a extração de atributos cinemáticos de arquivos de vídeos de animais de laboratório (roedores e moscas).

É composto por dois módulos, um de etografia que permite ao usuário construir e usar um catálogo comportamental para registrar eventos comportamentais continuamente ou quadro a quadro e outro módulo de rastreamento, que permite realizar a segmentação controlada do animal a partir de um *background* do arquivo de vídeo, utilizando os atributos para calcular distância percorrida, orientação, comprimento e área do animal (JUNIOR, et.al, 2012).

Como houve uma grande adesão da comunidade científica na utilização do EthoWatcher®, foi desenvolvido um novo programa de código aberto transferindo as mesmas funcionalidades do *software* existente, mas possibilitando a aquisição de imagem, vídeo, rastreamento de descritores cinemáticos e morfológicos, além da exibição dos resultados em uma interface gráfica (MARCOLAN, COLUSSI, 2018).

Em 2018 também houve mais uma contribuição nesse novo programa, possibilitando a detecção e rastreamento de múltiplos objetos e múltiplos segmentos de um mesmo objeto, o que resolveu o problema de perda de rastreamento em caso de sobreposição quando havia detecção de mais de um objeto na imagem (COLUSSI, MARCOLAN, 2019).

Existem ferramentas voltadas para análise de comportamento humano (*eg.* TrackLab da Noldus® (NOLDUS, 2020), D-Lab Video da Ergoneers (D-LAB, 2020), SemnsLab da Bitbrain® (SENNSLAB, 2020), Eye tracking: Screen-based da Imotions® (EYE

TRACKING, 2020)), que possuem interface moderna e diversas funcionalidades (*e.g.* coleta, análise, visualização gráfica dos dados e análise de comportamentos), com aplicações em diversas áreas como shoppings, escolas e hospitais, porém são soluções não gratuitas, e conseqüentemente não possuem código aberto.

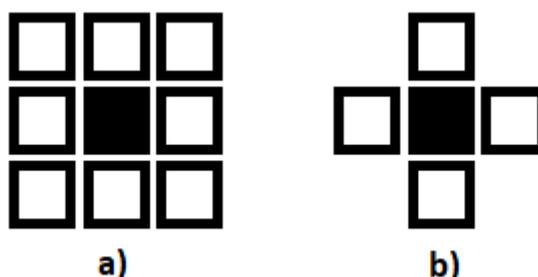
2.4 Processamento digital de imagem

A área de processamento digital de imagem diz respeito ao processamento de imagens digitais, realizadas por um computador digital. Podemos definir uma imagem como uma função bidimensional ($f(x,y)$), onde x e y são coordenadas espaciais e a intensidade ou nível de cinza da imagem é a amplitude da função em qualquer par de coordenadas x e y . Quando a intensidade e os valores de x e y possuem quantidades discretas e finitas, denominamos imagem digital, que é composta por um número finito de elementos, chamados de pixels (GONZALES, WOODS, 2010) (JÄHNE, 2002).

Os pixels são elementos básicos em uma imagem e podem ser representados em forma retangular ou quadrada (mais comuns). A organização de uma imagem pode ser realizada em forma de matriz de pixels com simetria quadrada, dessa forma um pixel passa a ter quatro vizinhos de borda e quatro vizinhos na diagonal (ESQUEF, et.al, 2003), conforme a Figura 2. É importante conhecer a matriz de pixels e sua estrutura, pois com ela podemos realizar diversas transformações na imagem (*e.g.* transformações morfológicas erosão, dilatação, abertura e fechamento).

Na Figura 2 podemos ver a representação da conectividade do pixel central (quadrado preenchido) com os vizinhos (quadrados vazios). a) Conectividade do pixel central com 8 vizinhos, onde 4 são vizinhos de borda e 4 são vizinhos de diagonal. b) Conectividade do pixel central com os 4 pixels de borda.

Figura 2 - Representação da conectividade de um pixel



Fonte: Adaptado de ESQUEF, et.al, 2003.

O processamento digital de imagens tem sido fundamental em diversas aplicações da visão computacional para o reconhecimento de padrões em atividades humanas, pois a área de aplicação é extremamente vasta (eg. Reconhecimento de padrões em atividades domésticas (JALAL, et.al, 2015) (JALAL, KAMAL, KIM, 2015), sensores de quedas para idosos (GUNALE, MUKHERJI, 2018), Sistemas de vigilâncias, monitoramento de saúde humana (PRATI, et.al, 2019) e outros).

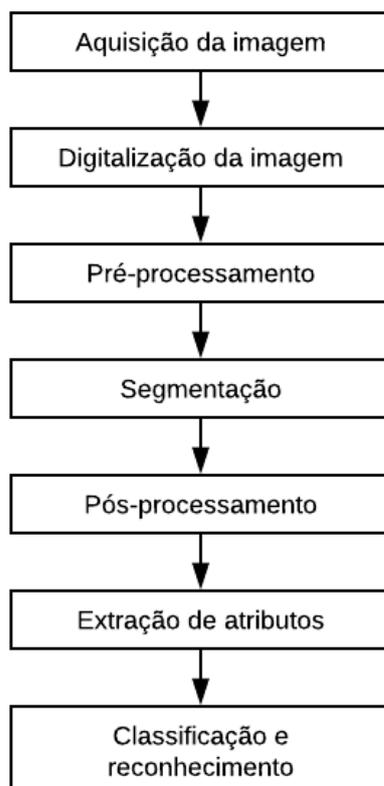
O principal desafio no reconhecimento de padrões em atividades humanas é o pré-processamento de dados, na extração de características que sejam adequadas para a atividade realizada e a classificação dessas características para cada atividade. O processo de extração de atributos continua sendo a chave para qualquer sistema de reconhecimento de atividades humanas e diversos métodos vêm sendo utilizados para representações, sombras e extração de pontos de interesse em um espaço (*e.g* transformadas, análise de componentes principais e análise de movimentos) (XU, et.al, 2012).

Existem uma série de atributos extraídos de imagens que representam ou caracterizam um padrão, que são extraídos de forma comum por diversos autores, como os atributos de distâncias locais, que são as distâncias entre pontos de uma silhueta humana (*e.g*. pontos locais de membros superiores e inferiores, como mãos, cabeça e pés, bem como a distância entre eles); características invariantes (*e.g*. Invariante a escala, rotação e translação) gradientes de orientação (*e.g*. centroide, centro de massa) e atributos de silhueta (*e.g* área do objeto), que são utilizados para a identificação de padrões (JALAL, et.al, 2019) (KAMAL, JALAL, 2016) (KAMAL, et.al, 2018) (BIBI, et.al, 2018).

2.5 Etapas do processamento de imagem

O processamento digital de imagens é composto por um conjunto de passos (Figura 3), que tem como o primeiro passo a captura (obtenção) da imagem, que é realizada utilizando um sistema de aquisição com um dispositivo físico (no nosso caso uma câmera digital). Após a obtenção por um processo de digitalização, a imagem pode ser representada em duas ou mais dimensões (como imagens 2D e 3D). Em seguida acontecem as etapas de pré-processamento, segmentação e pós processamento, que devem ser bem executadas pois interferem diretamente na qualidade da extração de atributos. Por último a classificação e reconhecimento de padrões.

Figura 3 - Etapas do processamento digital de imagens



Fonte: Autor

2.5.1 Aquisição e digitalização de uma imagem

A primeira etapa do processamento digital de imagens é a aquisição da imagem, que pode ser obtida através de scanners, câmeras digitais, satélites, entre outros dispositivos. O dispositivo de importância para o presente trabalho é a câmera digital, que funciona utilizando os sensores CCDs (*Charge-coupled device*). A gravação da imagem é realizada em três passos. Inicialmente ocorre a conversão da exposição à luz em carga elétrica em alguns pontos do sensor (pixels); em seguida as cargas são transferidas dentro do fotodiodo de silício e por fim transformadas em uma tensão para serem descarregadas (GONZALES, WOODS, 2010) (CHITYALA, PUDIPEDDI, 2020).

É necessário conhecer a representação matemática de uma imagem para facilitar a aplicação de técnicas de transformações de imagem. Uma imagem monocromática pode ser representada por uma função bidimensional para a intensidade luminosa $f(x,y)$, onde qualquer valor em f representa um nível de cinza (ou brilho) da imagem naquele ponto e as

coordenadas x e y representam a largura e altura. A função de intensidade da luz pode ser representada como:

$$f(x, y) = i(x, y) \cdot r(x, y) \quad f(x, y) = i(x, y) \cdot r(x, y)$$

Onde:

i representa a função de iluminação do ambiente, $0 < i(x, y) < \infty$

r representa a função de refletância dos objetos, $0 \leq r(x, y) \leq 1$.

Para representar uma imagem em forma digital é necessário adotar escalas para as coordenadas espaciais x e y e também para a intensidade da luz, transformando de um espaço contínuo para um espaço discreto, ou seja, a imagem deve ser convertida em números para que o computador possa efetuar a leitura. A primeira etapa consiste na medição dos níveis de cinza na imagem resultando em uma matriz de elementos (pixels), que determinam a resolução espacial da imagem conforme a área, em potências de 2 (e.g. 64x64, 256x256, 128x128, 512x512) (GONZALES, WOODS, 2010) (DE QUEIROZ, GOMES, 2006) (CHITYALA, PUDIPEDDI, 2020). Cada pixel na imagem terá um valor numérico inteiro correspondente a um valor de cinza, que pode variar de 0 a 255 para uma imagem em tons de cinza. Na Figura 4 podemos ver a representação numérica de cada pixel em uma imagem em escalas de cinza. (A) Imagem original. (B) Imagem com a representação numérica de cada pixel, onde a cor preta é representada pelo 0 e a cor branca pelo valor 255.

Figura 4 - Exemplo de valores numéricos para pixels

	A	B	C	D	E	F	G	H	I	J
1										
2										
3										
4										
5										
6										
7										
8										
9										
10										
				(A)						
1	0	30	60	90	120	150	180	210	230	255
2	0	30	60	90	120	150	180	210	230	255
3	0	30	60	90	120	150	180	210	230	255
4	0	30	60	90	120	150	180	210	230	255
5	0	30	60	90	120	150	180	210	230	255
6	0	30	60	90	120	150	180	210	230	255
7	0	30	60	90	120	150	180	210	230	255
8	0	30	60	90	120	150	180	210	230	255
9	0	30	60	90	120	150	180	210	230	255
10	0	30	60	90	120	150	180	210	230	255
				(B)						

Fonte: Autor

Para imagens coloridas a representação em cada ponto x e y é definida pela luminância, matiz e saturação, utilizando três matrizes de duas dimensões, conhecidas como RGB (*Red, Green, Blue*), pela combinação das três cores primárias: Azul, Verde e Vermelho (cada matriz representa uma cor), conforme a Figura 5.

Figura 5 - Representação numérica de pixels em RGB

0	30	60	90	120	150	180	210	230	255
0	30	60	90	120	150	180	210	230	255
0	30	60	90	120	150	180	210	230	255

Fonte: Autor

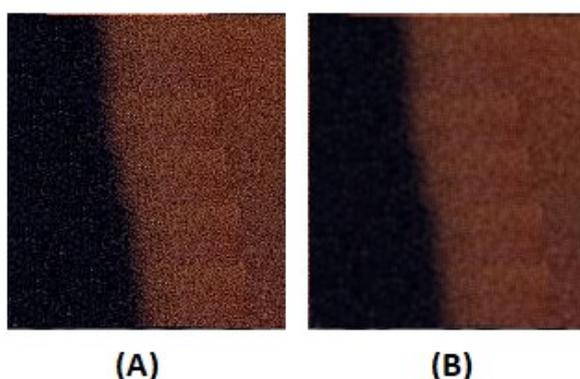
É importante ressaltar que imagens podem ocupar um grande espaço na memória dos dispositivos, por isso técnicas de compactação de imagens podem ser empregadas: com perdas, onde pequenos detalhes são eliminados de forma que não afete a imagem visualmente (*e.g.* JPEG) e sem perdas (*e.g.* PNG, TIFF). Para aplicações onde o dispositivo não possui um armazenamento em abundância ou onde a qualidade máxima da imagem não é relevante, podemos optar por formatos como o JPEG (*Joint Photographics Experts Group*), que possui

uma compactação de 16 bits e consegue manter arquivos com tamanhos pequenos e uma boa qualidade (DE QUEIROZ, GOMES, 2006) (CHITYALA, PUDIPEDDI, 2020).

2.5.2 Pré-processamento

O pré-processamento é onde acontece a remoção de ruídos, filtragem e correções em problemas como iluminação e distorções, com o objetivo de melhorar a qualidade da imagem. Essas técnicas operam em domínio espacial e no domínio da frequência, sendo que as técnicas no domínio espacial se baseiam em filtros para manipular o plano da imagem enquanto o domínio da frequência se baseia em filtros sobre o espectro da imagem. Na Figura 6 é possível ver um exemplo pré-processamento, utilizando uma imagem ruidosa (A) e aplicando um filtro Gaussiano com Kernel (3,3) obtendo como resultado uma imagem suavizada (B). (GONZALES, WOODS, 2010) (ACHARYA, RAY, 2005) (ESQUEF, et.al, 2003).

Figura 6 - Pré-processamento com filtro Gaussiano



Fonte: Autor

As principais técnicas utilizadas nesta etapa são melhorias no brilho da imagem, melhorias no contraste, correção de iluminação, realce de bordas, correção de fundo, entre outras (GONZALES, WOODS, 2010).

2.5.3 Segmentação

O próximo passo consiste na segmentação da imagem, que nada mais é que a separação da imagem em grupos de pixels de interesse, como por exemplo a separação do que é considerado fundo e o que é considerado objeto em primeiro plano, o que depende do ponto de vista de quem está analisando. No caso do presente trabalho, o grupo de pixels que constituem o humano em cena poderiam representar o “objeto”, e todo o resto considerado como fundo. Para a resolução de problemas onde o objetivo é identificar um objeto ou regiões

de interesse, o processo de segmentação faz toda a diferença para obter o sucesso, pois um erro nesta etapa reflete em todas as etapas posteriores (GRUOSSO, et.al, 2020) (BADRINARAYANAN, et.al, 2017).

A imagem de saída da segmentação pode ser representada por pixels pretos (que são referentes ao fundo da imagem ou objetos que não são do grupo de interesse) e pixels brancos que representam os objetos de interesse, que serão analisados. Os algoritmos de segmentação se baseiam em descontinuidade e similaridade, sendo que na descontinuidade podemos citar as mudanças na intensidade como as bordas. Na similaridade podemos citar a divisão da imagem em regiões que apresentam semelhança em critérios predefinidos (*e.g.* limiarização) (GONZALES, WOODS, 2010).

2.5.4 Detecção de bordas

A detecção de bordas é uma técnica utilizada para segmentar imagens de acordo com as variações repentinas de intensidade e os modelos de borda são classificados conforme o perfil de intensidade que representa (*e.g.* borda em degrau, borda em rampa e em forma de telhado). A técnica pode ser dividida em três passos que seriam a redução de ruído da imagem através da suavização; detecção de todos os pontos potenciais em fazer parte da borda e o último passo seria localizar os pontos pertencentes à borda. A Figura 7 apresenta uma demonstração de uma borda degrau de acordo com seu perfil de intensidade.

Figura 7 - Representação ideal de uma borda em degrau



Fonte: Adaptado de GONZALES, WOODS, 2010.

Entre os diversos métodos de extração de borda, podemos destacar o método de Canny, Sobel, Gradiente limiarizado e o algoritmo de Marr-Hildreth.

2.5.5 Limiarização

O método de limiarização é um dos processos de segmentação de imagens que consiste em escolher um ou mais valores (limiares) para separar os pixels em diferentes

grupos na imagem, comparando o valor de cada um dos pixels com o valor de referência. A imagem pode ser segmentada em: um grupo de pixels com níveis acima do limiar (atribui valor máximo) e um grupo de pixels com níveis abaixo do limiar (atribui valor mínimo). A limiarização pode ser aplicada em uma imagem binária e também em imagens com níveis de cinza (GONZALES, WOODS, 2010) (CHITYALA, PUDIPEDDI, 2020).

2.5.6 Método de subtração de fundo

O método de extração de fundo consiste em separar o fundo de uma imagem do objeto de interesse, ou separar um objeto de interesse de outros objetos que não representam importância para uma análise. É necessário obter uma imagem de entrada como referência, não contendo o objeto de interesse que será analisado. Em seguida a imagem contendo o objeto de interesse é subtraída da imagem de entrada. Após a subtração, tudo que for diferente da imagem de entrada será considerado o conjunto de pixels resultantes. Esse método é amplamente utilizado em vídeos para separação de objetos em movimento em um cenário (*e.g.* detecção em câmeras de segurança (SHARMA, 2020), detecção automática de pessoas, detecção de objetos em movimento) (GONZALES, WOODS, 2010) (DE QUEIROZ, GOMES, 2006) (CHITYALA, PUDIPEDDI, 2020).

Além disso o método de subtração de fundo é muito utilizado para separar pessoas de um cenário, obtendo somente os pixels referentes ao humano que serão úteis para detecção de padrões humanos (ABD ELAZIZ, LU, 2019) (PICCARDI, 2004).

2.5.7 Pós processamento

O pós-processamento é onde acontece a correção dos principais defeitos obtidos pela segmentação, utilizando filtros morfológicos e técnicas de morfologia matemática (*e.g.* erosão, dilatação, abertura, fechamento) (GONZALES, WOODS, 2010) (ACHARYA, RAY, 2005). É importante ter o conhecimento de todos os passos no processamento digital de imagens, para evitar ruídos externos no processo de obtenção e utilizar corretamente as ferramentas para pré e pós processamento da imagem.

A morfologia matemática lida com formas e estruturas geométricas na imagem, é definida como o estudo e análise de imagens utilizando operadores não lineares. Serve como ferramenta para descrever e representar componentes de imagens, podendo ser utilizada tanto para o pré-processamento quanto para o pós processamento de imagens. Existem algoritmos

para realizar essas transformações morfológicas, utilizando a biblioteca *openCV - Open Source Computer Vision Library* (BRADSKI, 2000) (OPEN CV, 2015), para cada operação é necessário realizar uma entrada com a imagem original e uma matriz de entrada contendo o elemento estruturante, conhecido como máscara (Kernel). Quatro técnicas de morfologia matemática (erosão, dilatação, abertura e fechamento) que são relevantes para aplicações de extração de características cinemáticas e morfológicas serão explicadas a seguir.

A erosão consiste em desgastar os limites do objeto em primeiro plano, matematicamente podemos representar com a seguinte equação, sendo um conjunto A e um conjunto B : $A \ominus B = \{x : B + x \subseteq A\} = \{x : Bx \subseteq A\}$, onde $A \ominus B$ é a erosão do conjunto A pelo conjunto B .

A dilatação é definida como o oposto da erosão, onde os limites do objeto em primeiro plano são aumentados. Pode ser representada pela equação $A \oplus B = [(A)' \oplus B']'$ onde A é uma imagem e B é um Kernel (ou elemento estruturante). É comum utilizar primeiro uma transformação de erosão e logo em seguida uma transformação de dilatação para remover ruídos de uma imagem, pois a erosão desgasta e remove os ruídos, mas conseqüentemente desgasta também o objeto em primeiro plano, nesse caso a dilatação promove o aumento nos limites do objeto retornando ao seu tamanho original antes da erosão.

Como a erosão e a dilatação corrigem ruídos na imagem, mas não mantém os pixels pertencentes a um objeto de interesse do mesmo tamanho, existe uma alternativa, utilizando a iteratividade, que é corrigir os ruídos da imagem mantendo o seu tamanho original, que são as transformações chamadas de abertura e fechamento (opening e closing). Com o auxílio dessas transformações, o usuário pode utilizar o que for útil de acordo com o tipo de imagem que está sendo analisada. No caso da análise de atributos de um humano, é interessante manter os seus contornos para preservar características que possam representar algum padrão, nesse caso, uma alternativa seria a utilização de uma abertura, que suaviza o contorno e elimina pequenos ruídos.

Uma abertura pode ser definida como uma erosão seguida de uma dilatação. Podemos definir a abertura de uma imagem A por um elemento estruturante B (Kernel) como $A \circ B = (A \ominus B) \oplus B$. O fechamento pode ser definido como uma dilatação seguida de uma erosão, utilizando a mesma máscara (Kernel). Pode ser definido como $A \bullet B = (A \oplus B) \ominus B$ e diferente da abertura, o fechamento pode ser utilizado para suavizar as bordas internas da

imagem, fechando pequenos buracos e conectando componentes. O fechamento pode ser útil quando a área do objeto de interesse contém “buracos” que precisam ser fechados.

2.5.8 Extração de atributos de objetos em cena

É onde acontece a extração dos atributos característicos de uma imagem, ou seja, a descrição dos grupos de pixels através de valores (quantitativos). Essas medidas são feitas utilizando a imagem em tons de cinza ou segmentadas e podem ser divididas em medidas de região, que são características específicas de cada objeto (*e.g.* formato, tamanho, posicionamento) e medidas de campo, que são características específicas do campo (*e.g.* área total de objetos, quantidade de objetos na imagem) (ALSMADI, 2020) (KABBAL, et.al, 2019). Esses atributos podem ser classificados em cinemáticos e morfológicos podem ser utilizados para o reconhecimento de um padrão de um objeto, por exemplo.

2.5.9 Descritores cinemáticos e morfológicos

Um objeto possui forma, textura, cor, dimensões e propriedades que podem ser medidas. Chamamos esse tipo de mensuração de descritor, que é um índice que representa um fenômeno visual e pode ser obtido antes ou depois do processo de segmentação, buscando extrair informações importantes, que podem ser utilizadas para procedimentos de reconhecimento de padrões como a morfologia. Um descritor morfológico descreve diferentes aspectos de um objeto de interesse como textura, contorno, forma ou região. No processamento de imagem de humanos podemos destacar características morfológicas como o contorno, área e forma, que podem sofrer variações ao longo do tempo.

Quando um objeto possui movimento, de um quadro para outro por exemplo, descrevemos as características relacionadas a esse movimento com a cinemática. Os descritores cinemáticos representam características do movimento de objetos, como a velocidade média, aceleração e distância percorrida.

Os descritores cinemáticos e morfológicos representam informações importantes sobre um objeto (ou situação), discriminando orientação, tamanho, rotação, velocidade entre outros parâmetros, que podem ser usados como dados de entrada de um modelo classificador. Nesse contexto existem diferentes conjuntos de atributos que servem para casos específicos como objetos, humanos, insetos, ambiente entre outros, tanto em 2D quanto em 3D (BOULAY, et.al, 2006) (DE MIGUEL, et.al, 2017).

2.5.10 Reconhecimento de padrões

O reconhecimento de padrões é a etapa final do processamento digital de uma imagem e consiste em atribuir automaticamente um padrão para alguma classe. Um padrão é considerado um conjunto de descritores, e um ou mais conjuntos que compartilham propriedades semelhantes são considerados uma classe de padrões (GONZALES, WOODS, 2010). A classificação de padrões permite a aplicação em diversas áreas e representa uma grande evolução, pois a máquina é capaz de reconhecer automaticamente objetos, comportamentos de animais (WANG, 2019), comportamentos humanos (ASLAN, et.al, 2020), expressões de humanos (NIMMAKAYALA, et.al, 2020), quedas de idosos (DE MIGUEL, et.al, 2017) e diversas outras áreas incluindo a saúde.

2.6 Processamento de imagens aplicados na domótica e saúde

O processamento digital de imagens apresenta uma grande contribuição para o desenvolvimento de soluções para a saúde das pessoas, principalmente dos idosos. Atualmente soluções utilizando câmeras tem contribuído para o cuidado de idosos e pessoas com deficiências que necessitam de monitoramento constante, porém, em muitos casos o processamento de imagem requer uma combinação com outras tecnologias (*e.g.* sensores e *hardwares*) que possibilitam a implementação da Internet das Coisas (*IoT – internet of things*) (NIMMAKAYALA, et.al, 2020).

Uma queda, por exemplo, pode ser fatal para uma pessoa caso a ajuda não chegue em tempo suficiente, então a criação de ferramentas que podem monitorar se uma pessoa sofreu uma queda e comunicar alguma autoridade (*e.g.* polícia, bombeiro) ou a própria família, pode ser de grande ajuda para idosos que moram sozinhos e não possuem cuidadores. Esse monitoramento constante de pacientes é um desafio que os médicos e cuidadores enfrentam, tanto nos hospitais quanto em domicílios, e tecnologias utilizando sensores e câmeras, conectados na nuvem para enviar e receber dados em tempo real, pode ser uma solução para este desafio (CHEN, et.al, 2006) (DE MIGUEL, et.al, 2017) (BELSHAW, et.al, 2011) (GUNALE, MUKHERJI, 2018) (ALONSO, et.al, 2019).

2.7 Ética e proteção de imagens de humanos

O reconhecimento de imagens utilizando humanos é um assunto delicado quando se trata da ética humana e proteção da privacidade. Existem muitos riscos associados ao gravar imagens que identifiquem rostos ou o desenvolvimento de sistemas de reconhecimento facial, principalmente se isso for feito sem o consentimento da pessoa. Uma vez que um vídeo ou foto de um rosto fica disponível na internet sem a proteção de imagem devida, corre o risco de ser utilizado para fins antiéticos, bem como os bancos de dados com atributos extraídos dessas imagens (*e.g* estudos com rostos sem proteção\borrão facial (WANG, et.al, 2019))

É necessário planejar a forma de obtenção de dados envolvendo humanos e a ética envolvida no estudo, juntamente com um comitê de ética humano, para seguir todas as normas e leis necessárias e preservar a privacidade do voluntário. Registrar todos os detalhes e documentar a coleta de amostras através de um termo de consentimento, com assinatura de ambas as partes também é um processo que garante a segurança, tanto do pesquisador quanto do voluntário, deixando claro todos os direitos e deveres.

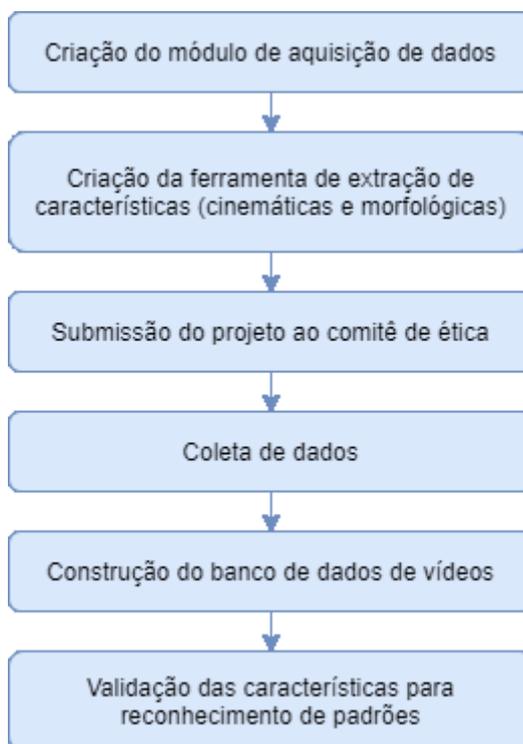
3 MATERIAL E MÉTODOS

Neste capítulo são descritas todas as ferramentas e procedimentos utilizados para alcançar os objetivos propostos. A execução do protocolo experimental foi realizada pela equipe pesquisadora do Projeto Plataforma de Saúde Ubíqua, do laboratório de Bioengenharia IEB –UFSC, sendo que todos os procedimentos adotados foram aprovados pelo Comitê de Ética em Pesquisa com Seres Humanos (CEPHS). De acordo com o documento de aprovação, nomeado como Parecer Consubstanciado do CEP, o projeto foi aprovado em 12 de novembro de 2019 com o número do parecer 3.700.866 (Anexo C).

3.1 Cronologia de atividades desenvolvidas

A metodologia utilizada no presente trabalho foi planejada e desenvolvida em uma série de etapas, em ordem cronológica, pois haviam passos que só puderam ser desenvolvidos após a conclusão de outros (*e.g* coleta de dados só aconteceu após a submissão e aprovação do projeto pelo comitê de ética humano). Podemos destacar como etapas principais a criação do módulo de aquisição de dados, a criação da ferramenta de extração de características (cinemáticas e morfológicas), submissão do projeto ao comitê de ética, coleta de dados, construção do banco de dados e validação da ferramenta para reconhecimento de padrões. Todas as etapas serão explicadas a seguir. Além disso, foi dedicado um capítulo, na seção 4 - Resultados, descrevendo os impactos da pandemia e Coronavírus (COVID-19) neste trabalho. Uma representação da cronologia de atividades, em forma de diagrama de blocos, pode ser vista na Figura 8.

Figura 8 - Cronologia de atividades



Fonte: Autor.

3.2 Plataforma de Saúde Ubíqua

A ferramenta de extração de características cinemáticas e morfológicas, desenvolvida no presente trabalho, faz parte da Plataforma de Saúde Ubíqua (Descrita no Capítulo 1), projeto que vem sendo desenvolvido pelo laboratório de domótica do IEB-UFSC de forma colaborativa, envolvendo alunos da pós-graduação e professores. A plataforma busca o seu desenvolvimento com a utilização de *hardware* livre e *software* de código aberto (*open source*), sendo composta por um módulo de aquisição e processamento de dados (multissensorial e câmera) e uma unidade concentradora local (responsável pelo recebimento, armazenamento, fusão de dados e reconhecimento de atividades).

3.3 Módulo de aquisição e processamento de dados

O módulo de aquisição e processamento de dados é composto por duas unidades: a unidade multissensorial e a unidade de câmera. Essas unidades foram construídas para funcionar de forma síncrona durante a aquisição de dados, mantendo a mesma quantidade de pacotes de dados (multissensoriais) e *frames* (vídeo), permitindo a fusão de dados

posteriormente. A interface utilizada para controlar a aquisição de dados é a mesma para as duas unidades e funciona em conjunto com os sensores e câmera, já o pré-processamento dos dados acontece de forma separada, utilizando uma interface específica para cada unidade.

A unidade multissensorial foi desenvolvida, testada e validada por um dos mestrandos participantes do projeto, em forma de dissertação (SIRIDAKIS, 2020). Como a aquisição de dados teria que funcionar em sincronia para as duas unidades, o desenvolvimento da interface de aquisição de dados foi realizado em conjunto com este projeto (desenvolvimento colaborativo) utilizando plataformas de versionamento online (Git). Esta unidade será comentada brevemente no presente trabalho. A unidade multissensorial utilizada para aquisição, pré-processamento e transmissão dos dados adquiridos pelos sensores, é composta por um conjunto de sensores concentrados em uma plataforma microcontrolada, utilizada para adquirir dados ambientais e antropogênicos.

A unidade de câmera utilizada para aquisição de dados em formato de vídeo, é composta por uma câmera capaz de adquirir vídeos. Esta unidade faz parte do tema do presente trabalho, e será explicada de forma detalhada nos próximos capítulos.

3.4 Hardware do Módulo de Aquisição de dados

O *Hardware* construído para a aquisição de dados é composto por uma câmera e um módulo contendo todos os sensores da unidade multissensorial agrupados, que devem ser conectados na unidade concentradora local. A estrutura foi planejada para ocupar o mínimo de espaço possível em um ambiente, ou seja, que não chame atenção visual das pessoas no ambiente, parte da premissa de que uma pessoa não se sente à vontade sendo monitorada constantemente, e o fato de ter um sistema pouco perceptível de monitoramento poderia reduzir esse desconforto.

3.4.1 Unidade concentradora Local

A unidade concentradora local é composta por um microcomputador (ou computador de placa única) Raspberry Pi (RPi) conectado via cabo (flat ou USB) com uma câmera. O RPi é uma plataforma compacta e facilmente montável em espaços reduzidos, levando em conta as suas especificações técnicas e o preço baixo em relação a computadores com especificações semelhantes, sua utilização acaba se tornando um diferencial para projetos.

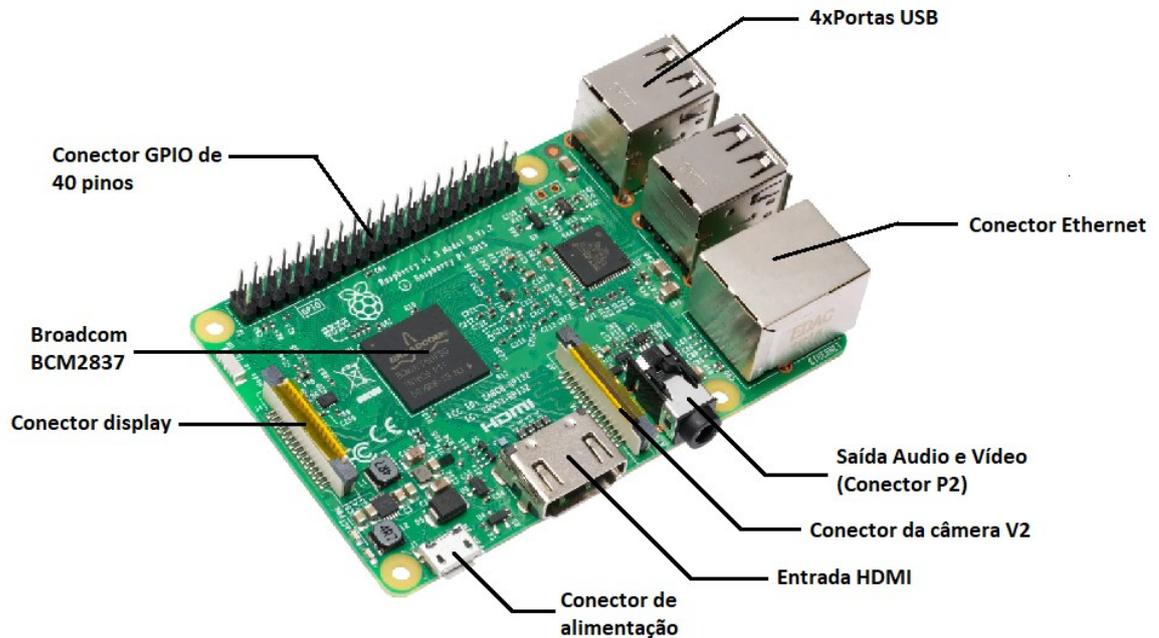
O RPi é extremamente versátil, podendo funcionar como um computador convencional (com navegador, dispositivos de mídia, conexão com mouse e teclado),

aceitando sistemas operacionais como Linux, Ubuntu, Windows, entre outros. As versões mais recentes trazem memória RAM de 4GB, o que permite uma alta produtividade com diversos softwares. Como já existem projetos e aplicações IOT (*Internet Of Things*) utilizando Raspberry Pi, e visto que ele atende as necessidades do nosso projeto como unidade concentradora local (para realizar a conexão com outros dispositivos e funcionar como uma unidade de processamento) optou-se pelo Raspberry Pi 3 Model B. No site oficial da placa podem ser encontradas diversas aplicação com projetos de IOT (RASPBerry PI, 2020).

O processador do Raspberry Pi 3 model B é um Broadcom BCM837, Quad Core com 1.2GHz, baseado no Cortex-A53 (ARMDEVELOPER, 2020), suportando até 64 bits de instruções e memória RAM de 1GB, suficiente para realização de diversas tarefas de forma simultânea, que é pré-requisito neste projeto, como atuar como uma unidade de processamento de áudio e vídeo em conjunto com diversos sensores e câmeras.

Uma das principais vantagens em utilizar este modelo de RPi é a comunicação sem fio (Bluetooth e Wifi), logo é possível estar conectado à internet sem precisar de cabos de rede. Para utilizar em apenas um cômodo de uma casa (como é o caso do nosso laboratório que possui cerca de 15m²), o RPi model B seria suficiente em relação ao alcance de internet sem fio e poderíamos utilizá-lo em qualquer local do cômodo. O RPi também oferece entradas USB e HDMI, que facilitam a conexão com câmeras (USB), teclados, mouses, além de telas HD que podem rodar vídeos em 1080p. Informações retiradas no site oficial Raspberry® (RASPBerryPI, 2020). Na Figura 9 podemos ver uma imagem da placa com a localização dos seus principais componentes.

Figura 9 - Componentes do Raspberry Pi 3 Model B

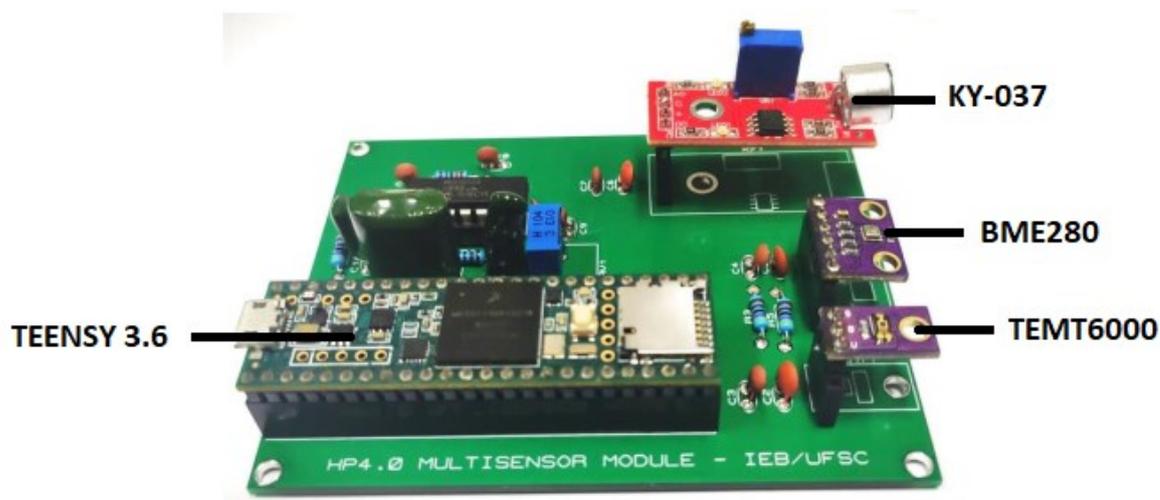


Fonte: Adaptado de RASPBERRYPI, 2020.

3.4.2 Módulo Multissensorial

O Módulo Multissensorial (SIRIDAKIS, 2020) é composto por uma plataforma de prototipagem eletrônica e *hardware* livre chamado Teensy 3.6, utilizado para amostrar, pré-processar e se comunicar com os sensores. Foram utilizados três sensores: o sensor sonoro KY-037, o sensor de iluminância TEMA6000 e o sensor de umidade e temperatura BME80 (SIRIDAKIS, 2020). Todos os componentes do módulo multissensorial foram conectados através de uma placa de circuito impresso (PCI). O protótipo do módulo multissensorial, com o componente de controle Teensy 3.6 e sensores conectados, pode ser visto na Figura 10, além disso na PCI foram utilizados alguns componentes para filtrar ruídos, como capacitores.

Figura 10 - Protótipo do Módulo Multissensorial



Fonte: Adaptado de (SIRIDAKIS, 2020).

3.4.3 Módulo de Câmera

O Módulo de Câmera deve utilizar uma câmera USB, conectada diretamente com a unidade concentradora local. A câmera USB que foi utilizada é da marca Microsoft®, modelo HD-3000, que de acordo com o site do fabricante possui sensor de imagem CMOS, resolução máxima 1280x720 pixels, captura máxima de vídeo 1280x720 pixels, taxa de frames igual a 30 fps (frames por segundo) e foco fixo de 0.3 a 1.5 metros (LIFECAM HD, 2020). Na Figura 11 podemos ver uma foto retirada do site do fabricante do modelo utilizado.

Figura 11 - LifeCam HD-3000



Fonte: LIFECAM HD, 2020.

3.5 Interface para Aquisição de Dados

Como a aquisição de dados pode ser um processo controlado (*e.g.* definição de parâmetros específicos, tempo específico, formato específico), para dados multissensoriais e

vídeos, a construção de uma interface para realizar o controle total dos módulos e dos parâmetros utilizados, proporciona segurança e conforto de usabilidade ao usuário do sistema de aquisição. Com isso o usuário não precisará programar ou alterar códigos fontes para fazer uma implementação, tudo será concentrado em uma única interface, permitindo que usuários de diversas áreas de conhecimento possam utilizar.

A interface verifica quais módulos estão conectados e o seu endereço de conexão (*e.g.* a porta USB utilizada), além disso permite ao usuário controlar qual dos módulos será utilizado. No caso do Módulo Multissensorial é possível escolher quais dos sensores irão coletar os dados e qual a taxa de aquisição (*e.g.* frequência de aquisição do sensor de áudio). No caso do Módulo de Câmera, permite escolher qual câmera será utilizada e qual a taxa de aquisição (frames por segundo).

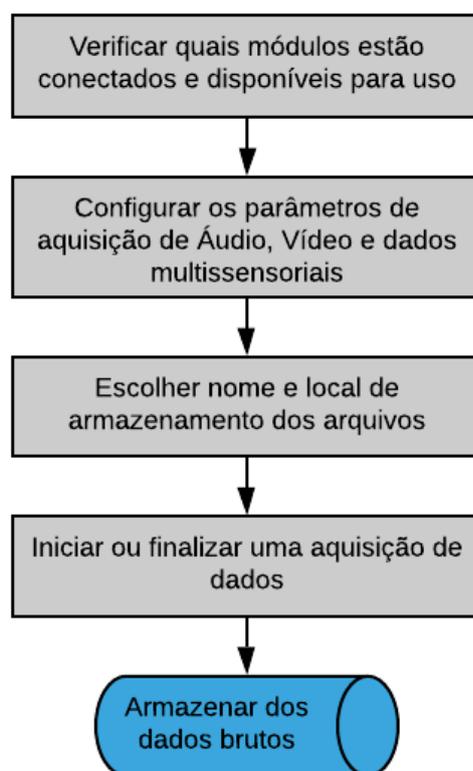
O arquivo de saída, onde serão gravados os dados multissensoriais, será no formato .CSV, que possibilita a exportação e importação em uma linguagem que diversos aplicativos podem realizar a leitura. O arquivo de saída de vídeo deve ser escolhido pelo usuário e deve ser nos formatos .AVI, .MPEG e .WMV, que já possuem seus formatos de compressão (compactação) pré-definidos no algoritmo, impossibilitando modificações do usuário para evitar possíveis erros.

O usuário deve escolher o nome do arquivo de saída e o local (pasta de arquivos da unidade concentradora) onde devem ser salvos os arquivos. O nome e o local devem ser comuns para todos os tipos de arquivos de dados (multissensoriais, áudio e vídeo), para facilitar a identificação dos arquivos de uma respectiva amostra coletada. A interface de aquisição de dados permite controlar quando iniciar e finalizar uma coleta.

Todos os frames de texto utilizados na interface foram ser escritos na língua inglesa, pois é uma das línguas mais faladas em todo o mundo, o que permite que pessoas da comunidade acadêmica mundial possam utilizar a interface com uma linguagem global. A linguagem de programação utilizada para o desenvolvimento da interface foi o Python e para auxiliar no design da interface a biblioteca PyQt5, que é uma ligação (*binding*) do python para o *framework* multiplataforma chamado Qt (criado em C++ na versão original).

A organização do fluxo da interface de Aquisição de dados pode ser vista na Figura 12.

Figura 12 - Diagrama de fluxo da Interface de Aquisição de Dados



Fonte: Autor

3.6 Interface para cortar e salvar vídeos

Quando se trata de análise de amostras de vídeo, principalmente para detecção de objetos ou reconhecimento de padrões, é comum dividir uma amostra de vídeo original em pequenas amostras, contendo somente os frames referentes ao objeto ou atividade de interesse. É importante analisar e extrair dados de forma isolada de cada objeto ou atividade de interesse, no caso de uma atividade específica humana, podemos analisar somente os quadros de interesse da realização da atividade e tudo que não fizer parte disso pode ser removido e separado em outra amostra (*e.g.* separar uma amostra de uma pessoa cozinhando de uma amostra de não realização de atividade). Informações que pertencem somente ao objeto ou atividade de interesse podem facilitar o treinamento de redes neurais (LAN, et.al, 2017) (CAO, et.al, 2013).

Como o presente trabalho prevê a coleta de dados de vídeos de humanos realizando atividades diversas em um ambiente, a criação de uma interface para dividir as amostras

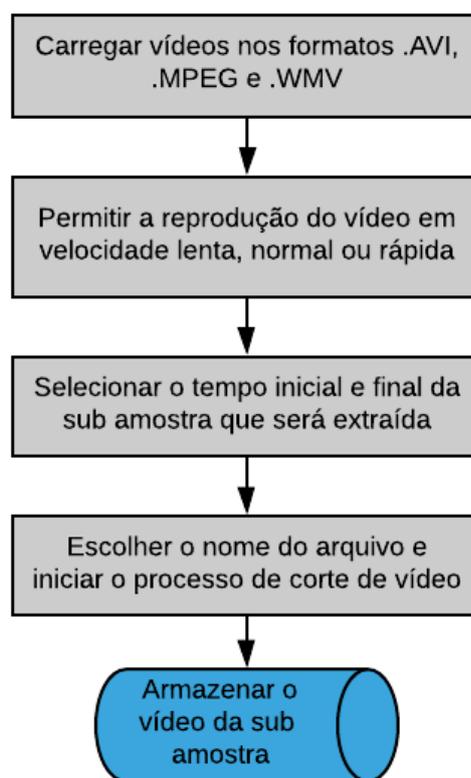
coletadas em sub amostras, facilitaria ao usuário a divisão de uma amostra de vídeo em várias sub amostras.

A interface de cortar e salvar vídeos permite o carregamento de um vídeo nos formatos .AVI, .MPEG, .WMV, que são os mesmos formatos disponíveis para salvar vídeos utilizando a interface de aquisição de dados. Após o carregamento do vídeo o usuário deve ser capaz de reproduzir o vídeo, com velocidade de reprodução lenta, normal ou rápida. A reprodução em diferentes velocidades pode auxiliar o usuário na hora de escolher o tempo exato de corte do vídeo para a sub amostra.

É possível escolher o tempo inicial e final (em horas, minutos e segundos) da amostra de vídeo a ser extraída do vídeo original e o nome do arquivo a ser salvo (o diretório onde será salvo a sub amostra será o mesmo do vídeo original). A interface possui também, um botão para iniciar o corte e salvamento do vídeo, apresentando uma mensagem ao concluir a tarefa. O vídeo contendo a amostra original não sofrerá alterações e não será descartado após a extração da sub amostra.

Todos os frames de texto utilizados na interface foram escritos na língua inglesa e a linguagem de programação utilizada para o desenvolvimento da interface foi o Python. Para auxiliar no design da interface será utilizada a biblioteca PyQt5. A organização do fluxo da Interface de Cortar e Salvar Vídeos pode ser visto na Figura 13.

Figura 13 - Diagrama de fluxo da Interface de Cortar e Salvar Vídeos



Fonte: autor.

3.7 Interface para detecção e *blurring* facial

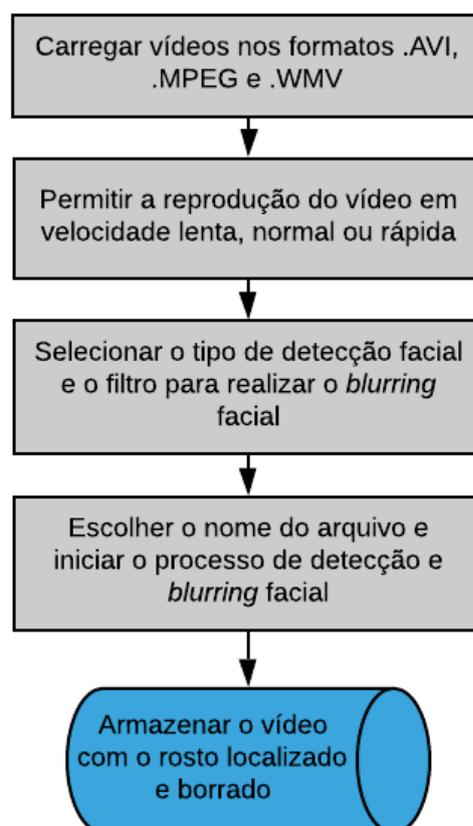
A utilização de amostras de imagens ou vídeos envolvendo seres humanos, necessitam de um cuidado especial ao proteger a privacidade da pessoa. Uma das premissas para o desenvolvimento do presente trabalho é realizar a pesquisa de forma transparente e ética, e para isso, é necessário seguir à risca alguns procedimentos que garantem a proteção da imagem. Um desses procedimentos é proteger o rosto de todos os humanos que aparecerem em amostras de vídeo, de forma que não seja possível realizar a sua identificação.

Para realizar a proteção de rostos será utilizada uma técnica conhecida como *blurring facial* (turvação facial ou borrão facial), que consiste em borrar todos os pixels que constituem um rosto, utilizando filtros de tratamento para esta finalidade (*e.g gaussian blur, Bilateral Filtering*). Para identificar os pixels que constituem um rosto, é necessário a utilização de algum algoritmo para realizar a detecção facial de humanos (CAO, et.al, 2013) (PAVITHRA, et.al, 2020) (SARWAR, RINNER, CAVALLARO, 2019).

A interface construída permite o carregamento de um vídeo nos formatos .AVI, .MPEG, .WMV, que são os mesmos formatos disponíveis para salvar vídeos utilizando a interface de aquisição de dados. Após o carregamento do vídeo o usuário deverá ser capaz de reproduzir o vídeo, com velocidade de reprodução lenta, normal ou rápida. A interface contém um menu para que o usuário escolha o tipo de detecção facial que deseja utilizar para localizar um rosto na imagem e também escolher o tipo de filtro que irá realizar o borrão facial.

A interface possui um campo específico para digitar e definir o nome do arquivo de vídeo que será salvo ao final do procedimento (contendo o rosto da pessoa localizado e borrado), um botão para iniciar a identificação facial e o processo de *blurring*, apresentando uma mensagem ao concluir a tarefa. O vídeo contendo a amostra original não sofrerá alterações e não será descartado.

Todos os frames de texto utilizados na interface foram escritos na língua inglesa e a linguagem de programação utilizada para o desenvolvimento da interface foi o Python. Para auxiliar no design da interface foi utilizada a biblioteca PyQt5. A organização do fluxo da Interface de Detecção e *Blurring* Facial pode ser visto na Figura 14.

Figura 14 - Diagrama de fluxo da Interface de Detecção e *Blurring* Facial

Fonte: autor.

3.8 Interface de Extração de Características Cinemáticas e Morfológicas

A Interface de Extração de Características Cinemáticas e Morfológicas extrai a partir de vídeos, características de um objeto de interesse encontrado na imagem, no caso do presente trabalho o objeto de interesse será uma pessoa. A interface construída permite o carregamento de um vídeo nos formatos .AVI, .MPEG, .WMV, que são os mesmos formatos disponíveis para salvar vídeos utilizando a interface de aquisição de dados. Após o carregamento do vídeo o usuário deverá ser capaz de reproduzir o vídeo, com velocidade de reprodução lenta, normal ou rápida.

Para identificar um objeto de interesse a partir de vídeos, deverá ser utilizado algum método de extração de fundo, que consista em subtrair uma imagem estática da imagem atual. O método de extração de fundo é uma técnica comumente utilizada para criar uma máscara utilizando uma câmera estática, ou seja, uma imagem binária contendo todos os pixels referentes ao objeto que está em movimento. O processo consiste em subtrair o quadro atual

por um modelo de fundo estático, que não contenha o objeto a ser analisado. O resultado dessa subtração é uma imagem que contém a área do objeto em pixels com 1-bit e o fundo da cena (parte estática) com 0-bit.

Além do modelo estático, existem outros modelos utilizados como: o Modelo Estatístico, onde os objetos de primeiro plano são modelados utilizando estimativas em relação ao plano de fundo; o modelo probabilístico conhecido como Modelo de Mistura Gaussiana (Gaussian Mixture Model), onde o fundo estimado é sensível para a utilização em ambientes com altas velocidades de movimentação, nesse caso os parâmetros não podem ser estimados com flexibilidade para ambientes do mundo real que são cheio de ruídos; o modelo avançado ViBE que utiliza uma substituição aleatória e perde qualidade em ambientes com pouca movimentação; e modelos de aprendizado de rede (Learning Network Model) que necessita de alto custo computacional e memória para realizar o aprendizado.

Para realizar a extração de fundo a partir de um vídeo, a utilização de uma imagem recente como fundo estático evita a criação de ruídos por alterações bruscas na luminosidade ou por objetos que foram modificados no cenário. Em um ambiente controlado (como no nosso caso que será uma cozinha) existem diversos objetos que podem ser alterados de posição durante atividades de uma pessoa, logo, o ideal é buscar métodos que forneçam um fundo estático atualizado (imagem modelo). A interface fornece um menu para que seja escolhido o método de subtração de fundo a ser utilizado (caso seja implementado mais de um).

Como explicado anteriormente (Capítulo 2), ao realizar uma técnica de extração de fundo, a imagem resultante poderá conter os pixels do objeto de interesse e também diversos pixels que são considerados ruídos. A interface fornece ao usuário, métodos de morfologia matemática para modificar os pixels de interesse e também os pixels ruidosos.

A morfologia matemática lida com formas e estruturas geométricas na imagem, é definida como o estudo e análise de imagens utilizando operadores não lineares. Serve como ferramenta para descrever e representar componentes de imagens, podendo ser utilizada tanto para o pré-processamento quanto para o pós-processamento de imagens. Uma operação binária funciona examinando uma vizinhança ao redor de um ponto central, de forma que essa vizinhança seja definida como elemento estruturante.

As principais operações de morfologia matemática aplicado a imagens são: Erosão, Dilatação, Abertura e Fechamento. Para cada operação é necessário realizar uma entrada com

a imagem original e uma entrada contendo o elemento estruturante, conhecido como máscara (Kernel), que é uma matriz. A interface permite ao usuário escolher quais operações de morfologia matemática e qual a respectiva máscara utilizada em cada operação.

A interface oferece um menu onde é possível escolher o método de detecção de bordas (*e.g* Canny ou Sobel) e tipo de contornos que será utilizado, pois diversas características morfológicas são definidas a partir do contorno (*e.g* momentos da imagem, pontos extremos superiores e inferiores). O algoritmo utilizado deverá localizar o maior contorno dentro do quadro, pois o presente trabalho não utiliza casos contendo mais de um objeto de interesse em cena, ou seja, no vídeo haverá somente uma pessoa em cena. Localizar o maior contorno e considerá-lo como objeto de interesse deverá resolver problemas como ruídos ou conjunto de pixels referentes a objetos movimentados em cena que não são interessantes para a análise. A interface também contém um menu em formato de *checkbox* para que o usuário escolha as características cinemáticas e morfológicas que serão extraídas do objeto de interesse (pessoa).

As características que estarão disponíveis para serem extraídas da imagem são atributos comuns utilizados em diversas aplicações de visão computacional envolvendo detecção de humanos e reconhecimento de atividades (conforme explicado no capítulo 2). A interface deverá permitir ao usuário escolher quais, entre as seguintes características, serão extraídas da imagem: Quantidade total de pixels no quadro (pixels); Quantidade total de pixels do objeto de interesse (pixels); Razão do objeto/imagem (Divisão do número de pixels do objeto pelo número de pixels na imagem, convertidos em taxa percentual (%)); Centro em X - Cálculo do centro do eixo X do objeto de interesse; Centro em Y - Cálculo do centro do eixo Y do objeto de interesse; Excentricidade - Divisão do maior eixo pelo menor eixo; Maior ponto de contorno à direita - Cálculo do ponto de contorno que está mais à direita do objeto; Maior ponto de contorno à esquerda - Cálculo do ponto de contorno que está mais à esquerda do objeto; Maior ponto de contorno superior - Cálculo do ponto de contorno que está mais à cima do objeto; Maior ponto de contorno inferior - Cálculo do ponto de contorno que está mais à baixo do objeto; Centro em X; Centro em Y; Momentos centrais - Momentos centrais são invariantes ao movimento; 7 Momentos de Hu - Momentos HU são invariantes às transformações de imagem, sendo que os primeiros 6 deles são invariantes à movimento, escala, rotação e reflexo; o sétimo não é invariante ao reflexo

3.9 Coleta de dados

A coleta de dados foi uma das etapas mais importantes de todo o trabalho e o planejamento desta etapa fez toda a diferença na hora de sua execução. As amostras foram obtidas através de vídeos, áudios e dados multissensoriais de voluntários realizando atividades domésticas, conforme um roteiro. Foi necessário a criação de um documento, para registrar todas etapas planejadas para a coleta, que foi submetido ao comitê de ética humano. A coleta de dados só foi iniciada após a aprovação do CEPH. As etapas relacionadas à coleta de dados e seu planejamento serão explicadas a seguir.

3.9.1 Protocolo para coleta de dados

A coleta de dados consiste em duas etapas. Na primeira etapa o participante responde a um questionário contendo questões cadastrais e clínicas (ANEXO A). Dentre as questões cadastrais estão nome, idade, nascimento, sexo e telefone para contato. Dentre as questões clínicas estão perguntas relacionadas à saúde do paciente: se o mesmo possui alguma doença ou problema de coordenação motora; se está fazendo uso de algum medicamento controlado; se possui alguma lesão que possa afetar o movimento corporal ou coordenação motora; se possui algum problema de concentração ou se faz uso de algum medicamento que altere a mesma. Cada questionário possui um campo chamado ID, que será preenchido pelo laboratório como forma de identificação codificada do participante.

Na segunda etapa o voluntário é submetido ao roteiro de testes, contendo todas as ações que devem ser realizadas. Os eventos que foram registrados seguiram a seguinte sequência:

1 - *O participante entra no ambiente pela porta e se dirige para a geladeira; abre a geladeira e retira dois objetos (uma garrafa plástica contendo água e um pote de plástico contendo esferas plásticas); deposita os dois objetos sobre a mesa e fecha a porta da geladeira.*

2 – *O participante recolhe os dois objetos da mesa e se dirige até o balcão (que contém um fogareiro com uma panela (em cima), uma caixa de fósforos sem a lixa de ignição e um pote com diversas colheres); deposita água e esferas plásticas dentro da panela; simula a ignição de um fósforo com a caixa e o direciona ao fogareiro simulando o ato de acender o fogão. Em seguida, escolhe uma colher, pega e realiza movimentos circulares (10 a 15 movimentos)*

dentro da panela, movimentando as esferas na água; em seguida, simula o “desligamento” do fogão.

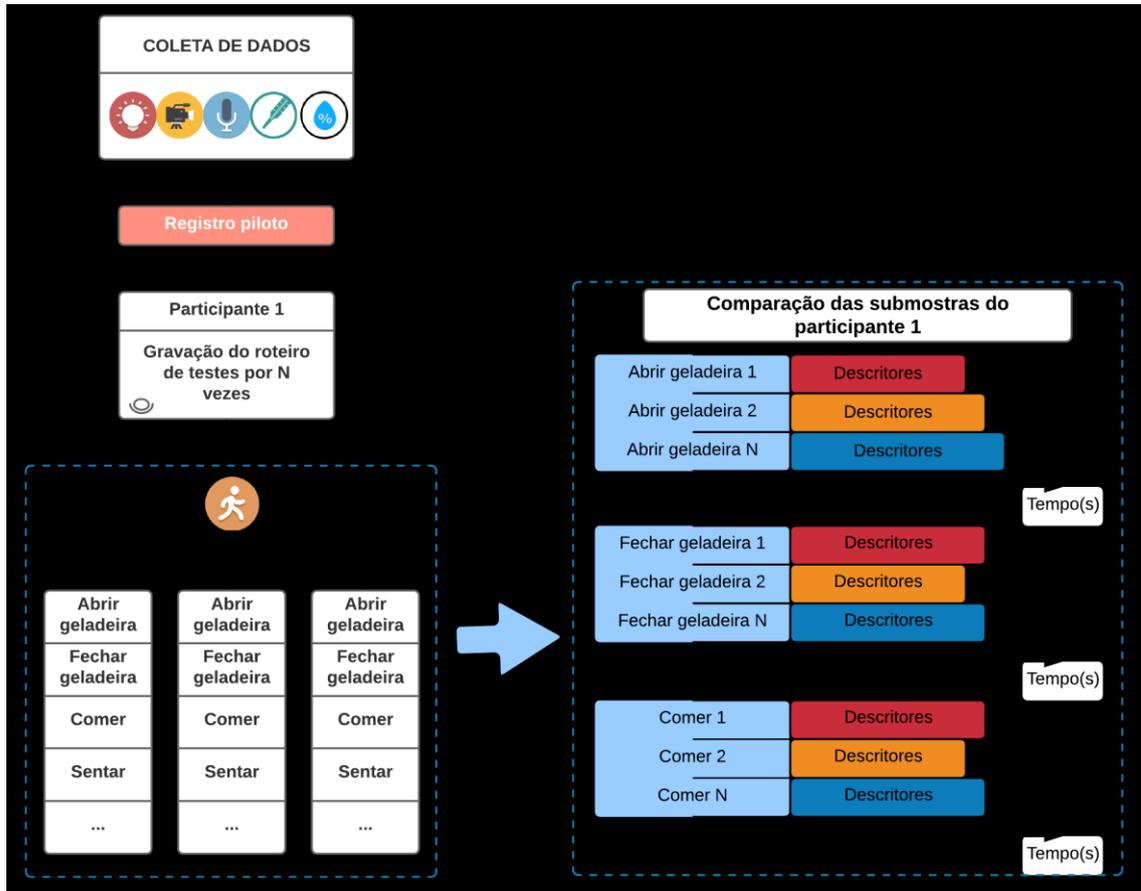
3 – O participante se dirige ao balcão da pia, escolhe um prato e o leva até a mesa; se dirige ao fogareiro, recolhe a panela e deposita a água e as esferas dentro do prato que está sobre a mesa; deixa a panela sobre a mesa e coloca a colher dentro do prato.

4 – O participante se senta na cadeira e utiliza a colher para simular o ato de “comer”, movimentando a colher, do prato em direção próxima da boca, por cerca de 10 vezes.

5 – O participante se levanta, recolhe a panela, o prato e a colher e os deposita dentro da pia (no balcão da pia contém: um pote plástico de cor branca, uma embalagem de detergente (com água no lugar de detergente), uma esponja e escorredor de louças); deposita a água e as esferas dentro do pote plástico de cor branca; simula o ato de “lavar louça”: abre a torneira, coloca o “detergente” na esponja, esfrega o utensílio e enxagua em água corrente; deposita os utensílios lavados no escorredor de louças; seca as mãos com um pano seco ao lado da pia; se retira do cenário pela porta. Fim do roteiro de atividades.

A coleta de amostras de atividades foi realizada como um conjunto de ações comuns (típicas) na cozinha, que os sujeitos deveriam realizar após entrar no ambiente de testes. Estas ações foram efetuadas em uma sequência e com objetos pré-determinados. As variações entre as diferentes repetições das atividades pelo mesmo sujeito (intra-sujeito), bem como aquelas inerentes à forma com que cada sujeito realiza o conjunto de atividades, de acordo com seus hábitos (entre-sujeitos), acrescenta variabilidade aos dados obtidos, favorecendo o treinamento de um futuro classificador. A Figura 15 ilustra esse processo:

Figura 15 - Representação da coleta e divisão de amostras no Registro.



Fonte: Autor

3.9.2 Submissão do projeto ao comitê de ética humano (CEPH)

Para a realizar a coleta de amostras, foi submetido um projeto ao Comitê de Ética em Pesquisas com Seres Humanos, detalhando todos os procedimentos, riscos e objetivos do projeto no Termo de Consentimento Livre e Esclarecido, com aprovação realizada no dia 12 de novembro de 2019.

Para que fique claro a importância da organização e atenção às diretrizes éticas e morais em relação à proteção de privacidade dos voluntários, é necessário explicar a cronologia das atividades que foram executadas desde o início do projeto para obter um projeto modelo, ou seja, que atenda aos requisitos mínimos para fazer pesquisa com seres humanos. Inicialmente houveram diversas reuniões com colegas de laboratório e coordenadores do projeto para planejar as ações relacionadas à ética em pesquisa, antes da execução do trabalho. Uma disciplina sobre ética em pesquisa também foi oferecida pela pós-graduação e todos os participantes do projeto cursaram, dessa forma já obtiveram

conhecimento e fontes (materiais didáticos) para entender a importância e atenção aos detalhes éticos durante um projeto. Como o projeto principal, denominado “Plataforma de Saúde Ubíqua voltada à Assistência domiciliar de idosos: detecção multissensorial de eventos em ambientes”, tem objetivos voltados ao cuidado de pessoas, incluindo análise de atividades, para a coleta de dados essas atividades teriam que ser executadas indispensavelmente por seres humanos. Para coletar dados de pessoas de forma invasiva e não invasiva, que envolvam imagem (como é o nosso caso), é necessário ter uma autorização do órgão responsável por estes assuntos, que é o Comitê de Ética em Pesquisas com Seres Humanos (CEPSC), um órgão colegiado interdisciplinar, consultivo, deliberativo e educativo. Existe uma sede dentro da própria Universidade Federal de Santa Catarina, logo denominamos CEPSC-UFSC. É importante ressaltar que por mais que o comitê esteja vinculado à universidade, a tomada de decisões ocorre de forma independente, de forma que protejam os participantes de uma pesquisa em sua integridade e dignidade, dentro dos padrões éticos.

Para que o CEPSC-UFSC analise seu projeto, é necessário realizar a submissão online, através de uma plataforma chamada “Plataforma Brasil”, onde todos os participantes do projeto, incluindo os coordenadores, devem estar regularmente cadastrados. Além disso existem uma série de questões relacionadas aos detalhes do projeto que devem ser respondidas, inclusive alguns documentos devem ser submetidos, como a liberação do local de pesquisa, autorização da universidade para execução do projeto, entre outros. A equipe responsável pela Plataforma de Saúde Ubíqua realizou todos esses passos de forma minuciosa. Um documento chamado Termo de Consentimento Livre e Esclarecido (TCLE) foi produzido, contendo todas os direitos e deveres do participante da pesquisa (ele funciona como uma autorização, onde o participante assina antes de iniciar a coleta de dados, o documento TCLE está disponível na seção ANEXO A). Com o projeto aprovado pelo comitê de ética, iniciamos a coleta de dados.

Conforme descrito nos documentos submetidos, o espaço para coleta de dados foi preparado previamente, respeitando todas as normas para proteção do participante, realizando os testes necessários antes de coletar cada amostra (o laboratório de teste é explicado com detalhes na seção Laboratório de testes). Foi construído um *checklist* (que está disponível na seção ANEXOS) contendo todos os itens que devem ser preparados e testados antes de iniciar cada coleta, o que é de extrema importância para que o procedimento não seja interrompido por falhas mecânicas ou por falta de itens necessários. Cada participante assinou e recebeu

uma via impressa do TCLE e outra via ficou com o laboratório, ambas assinadas pelos responsáveis do projeto. Esse documento é de extrema importância para o laboratório para firmar que as participações são de forma voluntária, transparente e de acordo com todas as normas previamente exigidas.

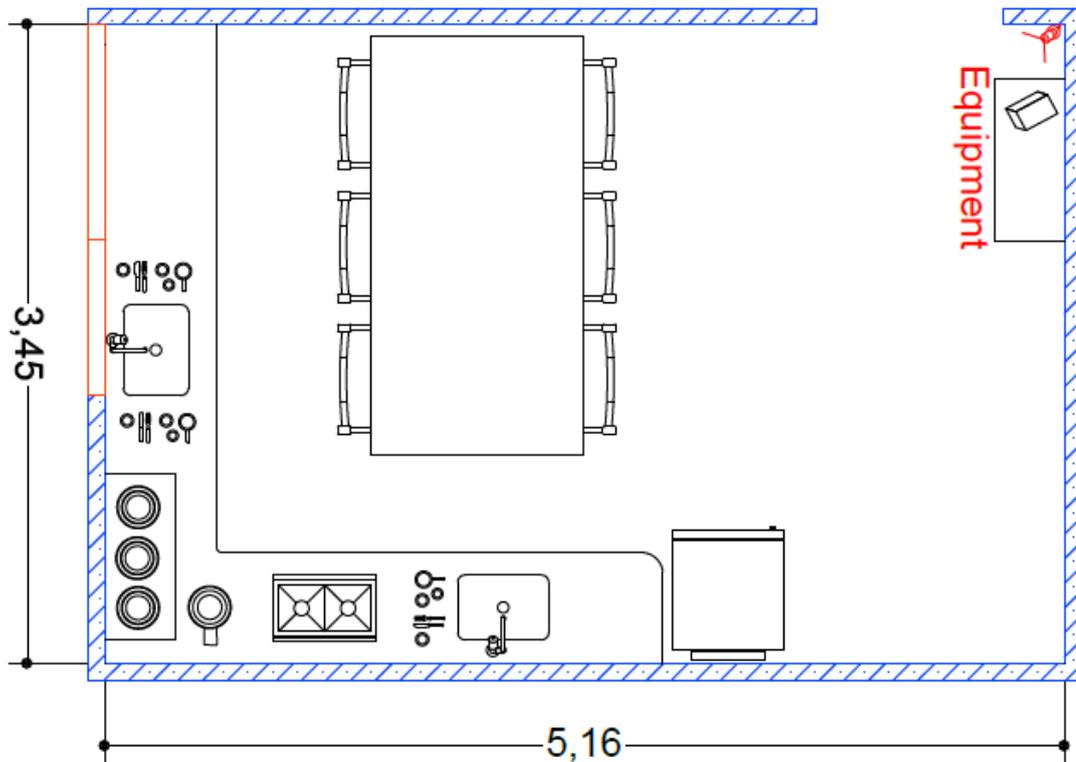
3.9.3 Laboratório de Testes

O laboratório utilizado para a realização dos testes e coleta de dados foi o laboratório de bioengenharia, no Instituto de Engenharia Biomédica da UFSC. Foram realizadas diversas modificações para transformar em um ambiente doméstico do tipo cozinha. As modificações serão explicadas brevemente a seguir. A posição da altura da câmera foi de 2,4 metros do chão, com possibilidade de ajuste de altura através de uma canaleta de PVC (Policloreto de Polivinila), improvisada como um trilho de movimentação.

O cômodo do laboratório tem formato retangular (5,16m por 3,45m), então escolheu-se um dos quatro cantos para fixar o módulo de câmera, justamente onde uma parede forma 90 graus com a outra. Direcionando a área de visão da câmera para a cozinha montada é possível obter uma amplitude de visão de forma abrangente para toda a área de testes. É importante ressaltar que o ângulo de abertura da câmera não conseguiu varrer todos os três cantos do cômodo, mesmo estando instalado em um dos quatro cantos, logo o espaço não varrido foi deixado vazio, não contendo nenhum eletrodoméstico ou item de importância e inclusive não fazendo parte da área de circulação de pessoas.

O cenário utilizado foi o ambiente doméstico cozinha, montado com alguns eletrodomésticos (geladeira, fogão, cafeteira e micro-ondas), móveis (mesa, cadeira, prateleiras) e diversos utensílios (panelas, pratos, talheres e copos). A necessidade da montagem do cenário passivelmente fiel ao real se deu devido a interação do indivíduo com o ambiente, especificado no roteiro de testes que será explicado posteriormente. Na Figura 16 podemos ver um esboço da planta baixa do laboratório, com os eletrodomésticos, móveis e a câmera em suas respectivas posições.

Figura 16 - Planta baixa do laboratório com equipamento de coleta de dados.



Fonte: Autor

Algumas estruturas como tomadas, interruptores, janelas, balcão, pias e lâmpadas não foram modificadas.

A câmera utilizada (LifeCam HD-3000) possui um campo de visão diagonal de 68,5 graus e foi posicionada buscando uma visão do ambiente de testes, enquadrando todos os eletrodomésticos, móveis e objetos utilizados durante o roteiro da coleta de dados. A instalação do equipamento foi feita ao lado da porta de entrada (como pode ser visto na Figura 17).

Optamos por uma visão mais abrangente, de um local alto, de forma a observar uma pessoa na imagem lateralmente, de frente ou de costas. A escolha desse tipo de visão tem alguns custos, como por exemplo a dificuldade em definir uma profundidade (se um objeto está perto ou distante da câmera) ou definir alguma métrica de referência (como o tamanho do campo de visão em metros ou centímetros), porém garante a visão de todos os membros da pessoa como pernas, pés, braços e mãos, inclusive diferenciar situações onde uma pessoa pode estar somente agachada no chão e não sentada em uma cadeira, por exemplo, que são informações que uma vista superior (câmera sobre o ambiente de testes) não conseguiria obter com facilidade.

Figura 17 - Campo de visão da câmera utilizada



Fonte: autor

3.9.4 Proteção de privacidade das amostras coletadas

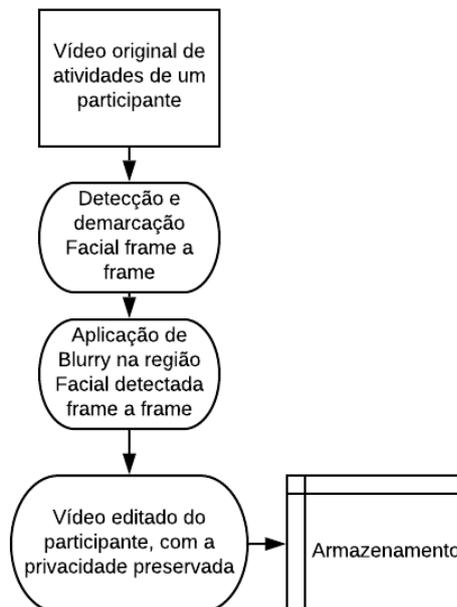
A primeira parte do protocolo do presente estudo envolve o preenchimento do questionário (dados cadastrais e clínicos referentes a coordenação motora (ANEXO A)). A segunda parte do protocolo consiste na coleta de vídeos pelo sistema de registro (câmera). Os pesquisadores e os professores responsáveis foram os únicos a terem acesso aos dados coletados e tomaram os devidos cuidados para manter o sigilo da identidade do participante, identificando o participante apenas por meio de códigos (cada participante possui um número de referência). Os dados coletados são armazenados em uma unidade de disco rígido utilizado unicamente para este fim, e cujo acesso foi permitido apenas aos pesquisadores e professores responsáveis.

Todas as imagens gravadas em formato de vídeo que foram utilizadas para testes, passaram por um processo de proteção de privacidade, no qual foi aplicado um borrão no rosto do participante, de forma que não seja possível realizar o reconhecimento facial. Os vídeos originais foram descartados, e os vídeos processados para impedir reconhecimento

facial foram identificados apenas por números. Se alguma imagem das gravações for utilizada em trabalhos científicos, a proteção de privacidade será garantida ocultando o rosto do participante.

A proteção de privacidade do vídeo dos participantes será realizada sempre que a amostra for utilizada, através do reconhecimento facial utilizando a interface desenvolvida, sempre que a face do participante for identificada em um frame, um campo será demarcado ao redor dessa região e um borrão será adicionado sobre esse campo identificado, resultando em um vídeo final com um rosto totalmente borrado frame a frame. Caso ainda existam frames em que a interface não conseguir identificar o rosto, poderão ser utilizados softwares livres disponíveis no mercado para esta finalidade. A identidade facial do sujeito fica totalmente protegida e o arquivo de vídeo protegido será salvo para posteriormente ser analisado. Na Figura 18 podemos visualizar os procedimentos para a proteção de privacidade.

Figura 18 - Procedimento de proteção de privacidade facial dos participantes gravados em vídeo



Fonte: autor

4 RESULTADOS

Neste capítulo serão apresentados todos os resultados obtidos na montagem do sistema de coleta de dados (*hardware* e interface), no desenvolvimento das interfaces de: cortar e salvar vídeos, proteção de imagem e de extração de características; testes realizados em cada etapa e os resultados da coleta de dados. Os resultados serão descritos de acordo com a ordem proposta no material e métodos, seguindo a cronologia de atividades.

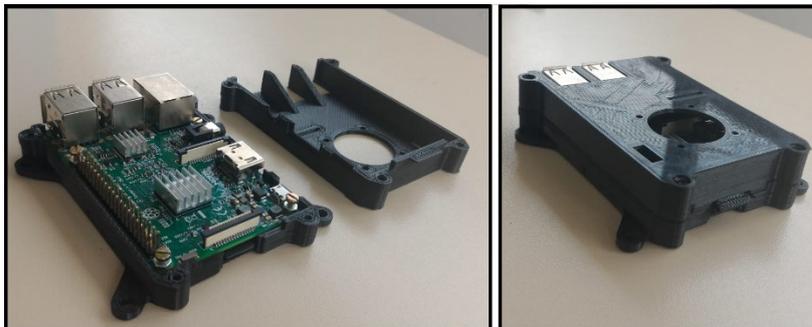
4.1 Instalação do *hardware* do sistema de aquisição de dados

O *hardware* da unidade concentradora local planejado para ser utilizado na plataforma de saúde ubíqua foi um Raspberry Pi, porém, para a primeira fase do projeto onde a aquisição de dados estava em desenvolvimento, foi utilizado um computador (Notebook Acer E1 571-6854, processador Intel Core i5, 6GB de memória RAM e 500GB de HD) controle de interface e armazenamento das amostras de áudio e vídeo coletadas. Dessa forma o módulo multissensorial e a câmera utilizada foram conectadas diretamente ao computador via USB e controladas através da interface criada para aquisição de dados.

A mudança da unidade concentradora local para realizar a coleta de dados foi necessária para agilizar o processo de aquisição de dados com cada voluntário, visto que cada roteiro teve uma duração mínima de 4 minutos, e para cada voluntário houve uma repetição de 5 vezes esse mesmo roteiro.

Para garantir a proteção da placa do Raspberry Pi, contra quedas, sujeiras ou qualquer tipo de danificação, foi construída uma case na impressora 3D, que pode ser vista na Figura 19. A case foi desenhada utilizando a ferramenta online e gratuita chamada Tinkercad, que pertence à Autodesk (TINKERCAD, 2018), deixando espaço para a instalação de um ventilador para resfriar a placa, caso necessário; o material utilizado para a impressão foi o Ácido Polilático, mais conhecido como PLA, de cor preta. A impressão foi realizada no Laboratório de Prototipagem do IEB - Instituto de Engenharia Biomédica da UFSC, localizado no primeiro andar do prédio do Instituto, com uma impressora 3D. Além disso, com a case construída foi possível fixar o Raspberry na parede com parafusos.

Figura 19 - Raspberry Pi com a Case feita em impressora 3D



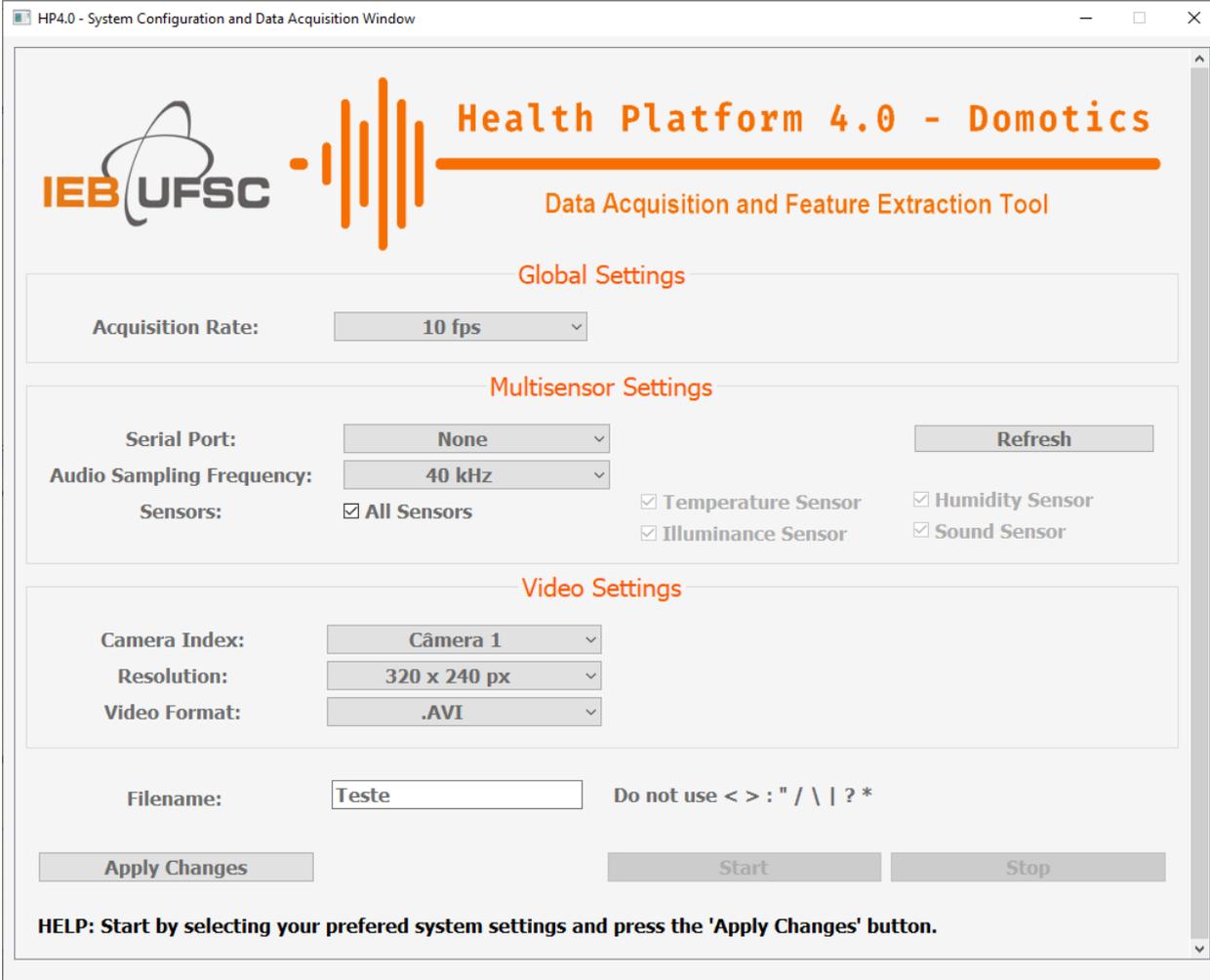
Fonte: autor

4.2 Software de aquisição de áudio e vídeo

Para facilitar a interação do usuário com o sistema de aquisição e extração de dados, foi construída uma interface utilizando a ferramenta PyQT, em Python. A interface pertence à Plataforma de Saúde Ubíqua, onde o sistema de coleta de vídeo funciona em sincronia com a coleta de áudio e dados sensoriais (temperatura, umidade e luminosidade). A interface foi construída em conjunto com o projeto de SIRIDAKIS, que diz respeito ao módulo multissensorial. No presente trabalho será detalhado somente a utilização da interface em relação à coleta e extração de dados de vídeo, descrevendo sobre os multissensores brevemente.

É importante ressaltar que, em consenso com os participantes do projeto, a ferramenta como um todo (Aquisição e extração) foi intitulada como “*Health Platform 4.0 – Domotics, Data Acquisition and Feature Extraction Tool*”. A primeira interface a ser explicada será aquela utilizada para configurar parâmetros e realizar a coleta de dados, tanto de vídeo, quanto de áudio e sensores, intitulada como “*System Configuration and Data Acquisition Window*”. Foi utilizado o idioma em inglês, seguindo o padrão de linguagem utilizado nos *softwares* anteriormente desenvolvidos pelo laboratório. A janela foi dividida em três seções principais: *Global Settings*, *Multisensor Settings* e *Video Settings*. Na Figura 20 podemos ver uma imagem da tela principal:

Figura 20 - Tela de configuração e aquisição de dados



HP4.0 - System Configuration and Data Acquisition Window

IEB UFSC - Health Platform 4.0 - Domotics
Data Acquisition and Feature Extraction Tool

Global Settings

Acquisition Rate: 10 fps

Multisensor Settings

Serial Port: None Refresh

Audio Sampling Frequency: 40 kHz

Sensors: All Sensors Temperature Sensor Humidity Sensor
 Illuminance Sensor Sound Sensor

Video Settings

Camera Index: Câmera 1

Resolution: 320 x 240 px

Video Format: .AVI

Filename: Teste Do not use < > : " / \ | ? *

Apply Changes Start Stop

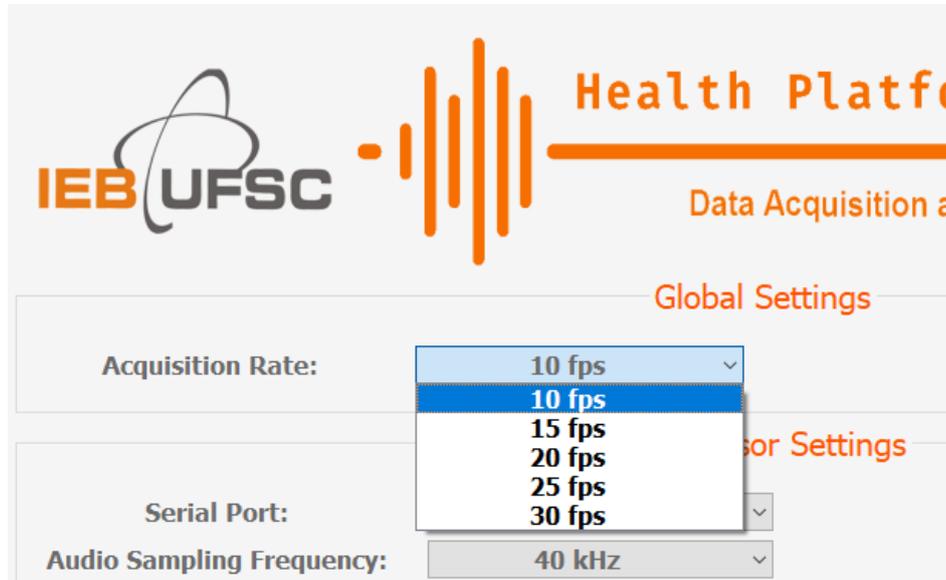
HELP: Start by selecting your preferred system settings and press the 'Apply Changes' button.

Fonte: Elaborada pelo autor

A primeira seção *Global Settings* aborda os parâmetros globais do software, tanto da câmera quanto do módulo multissensorial. Nesta primeira versão implementada, a interface possui apenas o parâmetro *Acquisition rate* (taxa de aquisição), que diz respeito a frequência de amostragem da câmera em quadros por segundo e com os sensores ambientais em amostras por segundo. A unidade de medida é em fps (frames por segundo), ajustável entre os valores fixos: 10, 15, 20, 25, 30.

Foi necessário determinar valores fixos para garantir a sincronia na coleta dos pacotes de dados de áudio, vídeo e sensores, onde 1 frame equivale a 1 quadro sonoro e 1 amostra de cada um dos sensores ambientais no mesmo intervalo de tempo. Sem essa equivalência, um vídeo gravado poderia ter o tempo diferente do áudio coletado da mesma ação. O ajuste dos valores de *Acquisition rate* são realizados por meio de um botão do tipo *dropdown*, que por *default* recebe o valor inicial de 10fps, e podem ser vistos na Figura 21

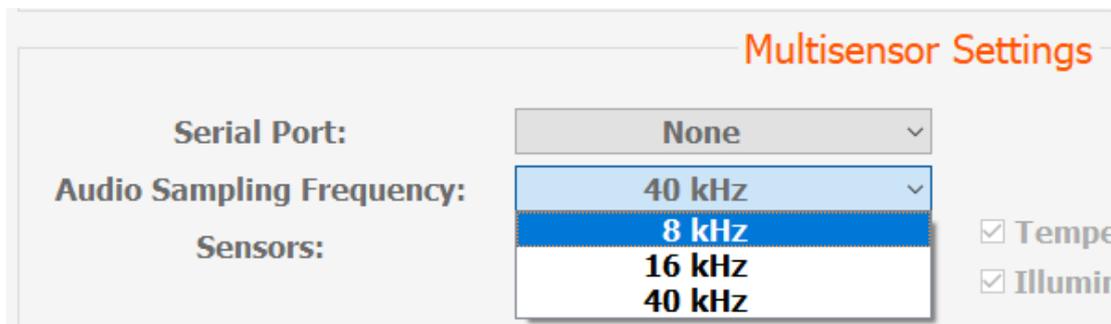
Figura 21 - Ajuste do parâmetro global *Acquisition Rate*



Fonte: Autor

A segunda seção *Multisensor Settings* diz respeito aos parâmetros multisensoriais (áudio, temperatura, umidade e luminosidade). O item *Serial Port* identifica em quais portas USB os sensores e microfones estão conectados (ao iniciar a interface ocorre um teste automático para verificar quais portas possuem conexão disponível) e as mesmas ficam disponíveis para escolha. No item *Audio Sampling Frequency* é possível alterar o valor de frequência das amostras de áudio apenas entre os valores fixos 8.000, 16.000 e 40.000 na unidade de medida Hertz (Figura 22). Estes valores foram escolhidos e pré-definidos de acordo com os testes realizados na unidade multisensorial (SIRIDAKIS, 2020).

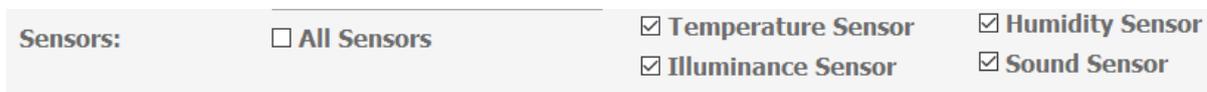
Figura 22- Ajuste do parâmetro *Audio Sampling Frequency*



No item *Sensors* é possível selecionar (em forma de *checkboxlist*) quais sensores serão utilizados para coletar dados, onde é possível escolher todos, apenas um, ou qualquer

combinação que o usuário necessitar. O botão chamado *All Sensors* funciona em forma de *checklist* para a seleção de todos os sensores disponíveis (Figura 23).

Figura 23 - Controle do parâmetro *Sensors*

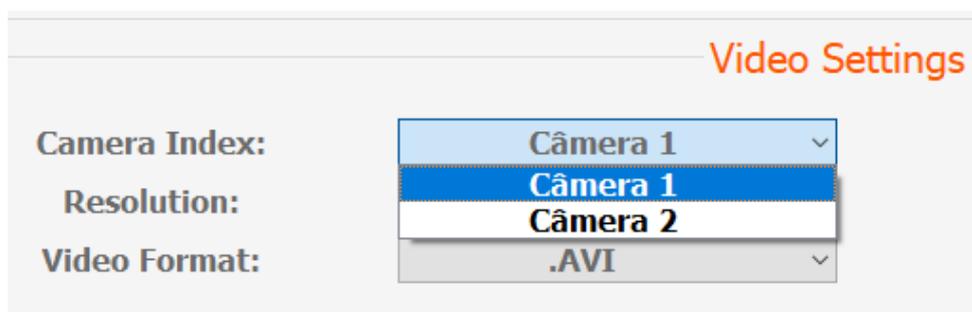


Fonte: autor.

Existe um botão ainda nesta seção que é chamado de *Refresh*, que funciona como um “atualizador”, ou seja, ao clicá-lo, a interface faz a busca de novas conexões via USB, útil no caso de um usuário conectar os sensores com a interface já em execução. Esse botão também funciona para a seção *Video Settings* (que será explicada abaixo), buscando se existem novas câmeras conectadas ao dispositivo e disponibilizando à escolha.

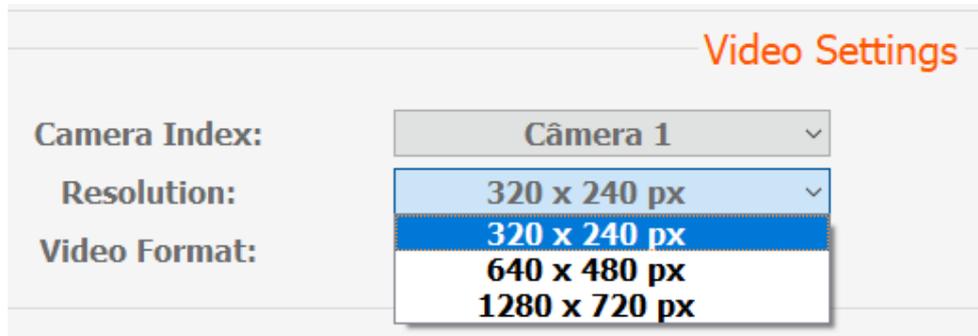
A seção *Video Settings* diz respeito aos parâmetros de configuração da câmera, como resolução e formato de vídeo que será salvo. No parâmetro *Camera Index* é possível escolher qual câmera será utilizada, no caso de haver mais de uma conectada (Este item é atualizado quando o botão *Refresh* é pressionado). O algoritmo implementado faz uma busca para identificar todas as câmeras conectadas no dispositivo e lista no botão *dropdown* do parâmetro *Câmera index*. Na Figura 24 temos um exemplo de funcionamento da interface, onde foram identificadas duas câmeras no computador do usuário.

Figura 24 - Ajuste do parâmetro Câmera Index



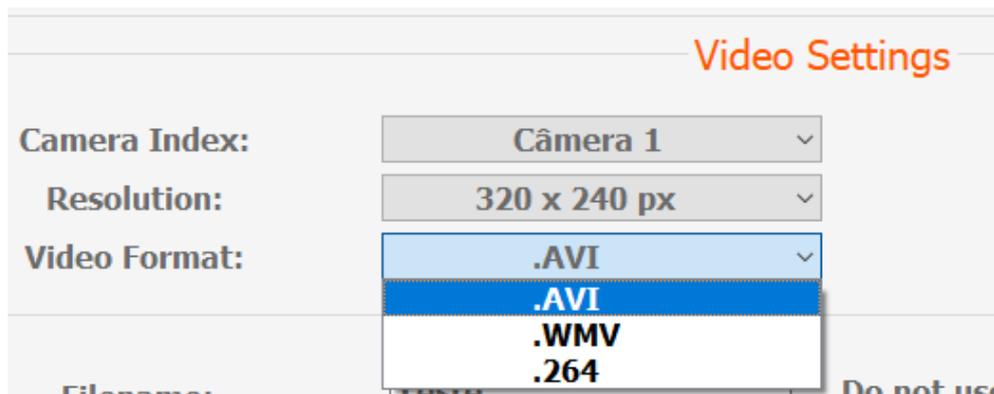
Fonte: autor.

No item *Resolution* é possível escolher qual resolução será utilizada, onde as opções de resolução variam conforme o tipo/modelo de câmera. O algoritmo implementado na interface realiza testes automáticos para verificar quais resoluções as câmeras conectadas suportam, e disponibiliza ao usuário estes valores para escolha. Na Figura 25 temos um exemplo de funcionamento da interface, onde foram identificadas quais as resoluções disponíveis para a “Câmera 1” encontrada.

Figura 25 - Ajuste do parâmetro *Resolution*

Fonte: autor.

O botão *Video Format* serve para escolher em qual formato o vídeo será salvo (encapsulado), onde o usuário pode escolher entre os formatos: .AVI (Audio Video Interleave), .WMV (Windows Media Video) e H.264, conforme a Figura 26.

Figura 26 - Ajuste do parâmetro *Video Format*

Fonte: autor.

Uma vez que todos os parâmetros forem modificados, é necessário pressionar o botão *Apply Changes* para aplicar as mudanças. No campo *Filename* é possível nomear (com letras e números) os arquivos que serão salvos (o nome *Default* é "Teste"). Conforme a observação na janela, não é possível usar caracteres do tipo < > : " / \ | ? * para nomear os arquivos, pois podem ocorrer erros na hora de salvar as amostras. Ao pressionar o botão *Start*, o usuário deverá selecionar o diretório onde as amostras ficarão salvas e clicar em "ok", assim a coleta de dados terá início e só terminará quando o botão *Stop* for pressionado. Conforme pode ser visto na Figura 27, existe uma mensagem de ajuda ao usuário, instruindo-o a selecionar os parâmetros e apertar o botão *Apply Change*.

Figura 27 - Botões para aplicar mudanças, iniciar ou finalizar uma aquisição



Fonte: autor.

4.2.1 Video Sample Split tool - VSST

4.2.1.1 Overview da interface

A ferramenta de corte de amostras de vídeo (*Video Sample Split tool - VSST*) permite ao usuário carregar vídeos nos formatos .AVI, .WMV e MPEG, visualizar em uma nova janela e cortar o vídeo em subamostras, selecionando o tempo inicial, final e salvando no mesmo formato do vídeo original.

O intuito desta ferramenta é auxiliar o usuário na hora de extrair pequenas amostras (samples) de uma amostra original. No caso do presente trabalho, uma amostra de vídeo original, contendo um roteiro completo de um voluntário, possui em média 3 a 6 minutos. Com a ferramenta podemos dividir em pequenas amostras, com tempos menores e somente com atividades de interesse, ou seja, cortamos tudo que não faz parte da ação. Na Figura 28 podemos ver a interface que o usuário utiliza para realizar este procedimento.

Figura 28 - Tela principal da VSST



Fonte: autor.

O primeiro passo é carregar um vídeo utilizando o botão “Open Video”, que pode ser nos formatos .AVI, .WMV e MPEG. Existe um controle multimídia de vídeo, com botões de play, pause\stop, avançar e retroceder, inclusive um controle de velocidade de reprodução, que pode ser 0.5 (metade da velocidade normal), 1 (velocidade normal) ou 2x (dobro da velocidade normal). Existe também o botão “Fullscreen” para reproduzir o vídeo em tela cheia (ao ativar o modo tela cheia, não é possível controlar o tempo de vídeo, somente assistir). Ao pressionar o botão “play”, é possível assistir ao vídeo, o que facilita na escolha do tempo inicial e final para executar o corte do vídeo.

É necessário informar o tempo inicial e final para realizar o corte do vídeo, e isso pode ser feito de duas formas: digitando manualmente dentro dos campos corretos (minutos e segundos); ou clicando nos botões “*Set initial time*” e “*Set end time*”, que o tempo atual do *slide* do player do vídeo, será copiado para os respectivos campos. Escolhido o tempo inicial e o tempo final, existe a opção de pressionar o botão “Start”. Feito isso, o vídeo original não sofrerá nenhuma edição e será mantido no mesmo diretório de origem, local onde o corte de vídeo será salvo. O formato de vídeo da nova amostra será o mesmo do arquivo original (*e.g.* se o vídeo original for .AVI, a nova amostra também terá formato .AVI).

4.2.1.2 Algoritmos utilizados

O módulo utilizado no algoritmo dessa ferramenta de corte foi o MoviePy, que é um módulo Python de edição de vídeo para operações básicas como cortes, composições, processamentos e criação de efeitos. É um módulo de código aberto, originalmente escrito por Zulko e atualmente encontra-se hospedado na plataforma GitHub (MOVIEPY, 2020).

No código foi importado da biblioteca principal a função `ffmpeg_extract_subclip`, que permite extrair partes de um vídeo. Os requisitos para essa função são: o vídeo original carregado, os tempos inicial e final desejados (*e.g.* para cortar um vídeo de 1min 20seg até 2min e 19s, o tempo inicial será 1:20 e o tempo final será 2:19), e o nome que será dado ao novo vídeo cortado.

4.2.1.3 Testes de funcionamento

Para testar o funcionamento do algoritmo utilizado na interface, foi carregado um vídeo no formato .AVI, de 640x480 (Largura x Altura do quadro), com 30fps (frames por segundo) e duração total de 7 minutos e 30 segundos, que foi utilizado para extrair diversos cortes de vídeos, com o intuito de verificar se os parâmetros do vídeo original seriam mantidos, além do tempo correto.

O primeiro teste foi para verificar se o algoritmo permitia cortes do menor tempo possível, ou seja, de 0 minutos e 0 segundos, até o valor 0 minutos e 50 segundos. O segundo teste foi para verificar se cortes entre minutos retornariam tempos corretos, então foi realizado um corte entre 1 minuto e 34 segundos até 2 minutos e 50 segundos. O terceiro teste foi para verificar se o algoritmo permitia cortes até o tempo máximo do vídeo, então foi realizado um corte entre 6 minutos e 20 segundos e 7 minutos e 30 segundos.

O quarto teste foi realizando utilizando tempo final acima do tempo máximo do vídeo original. Os resultados dos testes podem ser vistos na TABELA 1.

Tabela 1 - Resultado dos testes de cortes de vídeos

TESTE	TEMPO (Min:Seg)			LARGURA	ALTURA	FPS	FORMATO
	INICIAL	FINAL	RESULTADO				
TESTE 1	00:00	00:50	00:49	640	480	10	.AVI
TESTE 2	01:34	02:50	01:15	640	480	10	.AVI
TESTE 3	06:20	07:30	01:09	640	480	10	.AVI
TESTE 4	06:00	07:50	01:29	640	480	10	.AVI

Fonte: autor

Como pode ser visto na TABELA 1, a largura, altura, quantidade de FPS e formato, foram mantidos de acordo com o vídeo original. Os tempos de corte inicial e final funcionaram para todos os testes, executando cortes com a duração correta. É importante notar que o algoritmo considera o tempo inicial como parte da duração total, como podemos ver nos Testes 1, 2 e 3, então para obter, por exemplo, um vídeo final com duração total de 00:50, teria que usar tempo inicial 00:00 e tempo final 00:51. No teste 4, ao utilizar um tempo final acima da duração máxima do vídeo original, o algoritmo utilizou o tempo máximo do vídeo original como tempo final de corte (e.g. usando 6:00 como tempo inicial e 7:50 como

tempo final, o esperado seria um vídeo com 1:49 de duração, porém, o tempo final foi 7:30, pois esta é a duração total do vídeo)

4.2.2 Face Blur and Detection Tool - FBDT

4.2.2.1 Overview da interface

A ferramenta de detecção e desfoque de rosto (*Face Blur and Detection Tool – FBDT*) possui um detector de faces que utiliza Haar Cascades da biblioteca Open CV. Na Figura 29 podemos ver a interface que o usuário utiliza esse procedimento.

Figura 29 - Tela principal da FBDT



Fonte: autor.

O primeiro passo é carregar um vídeo utilizando o botão “Open Video”, que pode ser nos formatos .AVI, .WMV e MPEG. Existe um controle multimídia de vídeo, com botões de play, pause\stop, avançar e retroceder, inclusive um controle de velocidade de reprodução, que pode ser 0.5 (metade da velocidade normal), 1 (velocidade normal) ou 2x (dobro da velocidade normal). Existe também o botão “Fullscreen” para reproduzir o vídeo em tela cheia (ao ativar o modo tela cheia, não é possível controlar o tempo de vídeo, somente assistir).

Após carregar o vídeo é preciso escolher para o método *Cascade Classifier*, o *scaleFactor*, que é o fator de escala de como será dimensionada a imagem de entrada antes da detecção que varia de 0 a 100%, porém a entrada é numérica (e.g. 1.01 equivale a 1%, 1.1

equivale a 10% e 1.6 equivale a 60%). Também é preciso escolher o valor do *minNeighbors*, que determina a robustez da detecção, onde os valores devem ser inteiros e o menor valor aceito é 1 (valores próximos de 1 podem detectar rosto menores, porém pode aumentar os falsos positivos; valores acima de 6 ou mais podem exigir uma confiança maior para detectar os rostos, então algum rosto que estiver obstruído pode não ser detectado). Após parametrizar basta apertar o botão “Start” para iniciar o processo de detecção e borrão facial.

O filtro utilizado para realizar o borrão é um filtro Gaussiano, que já possui valores fixos com um kernel de (23,23) e um valor de sigmaX e sigmaY igual a 30 (Esses valores fixos garantem o borrão de forma que o rosto não seja identificado). Ao clicar no botão “Start” a detecção facial começa e assim que um rosto é detectado em um frame do vídeo, é aplicado o filtro gaussiano em todos os pixels pertencentes ao rosto, borrando o rosto encontrado. Esse tipo de detecção é eficaz, porém não é tão robusta quanto aquelas que utilizam redes neurais (e.g. MTCNN - *Multi-Task Cascaded Convolutional Neural Network*), então podem haver frames que a face não é detectada, e nesse caso, poderá ser necessária uma análise e borrão manual utilizando outros softwares de edição de vídeo.

4.2.2.2 Algoritmos utilizados

Os algoritmos utilizados foram todos de bibliotecas de código aberto com as suas respectivas licenças de uso incluídas no cabeçalho do código. O algoritmo utilizado para detecção foi o *Cascade Classifier*, que é um método de detecção de objetos usando classificadores em cascata, proposto por Viola e Jones (VIOLA, JONES, 2001). É baseado em aprendizado de máquina com uma função em cascata, treinada com imagens positivas e negativas. A biblioteca do openCV oferece um modelo pré treinado para detecção de rostos, em um arquivo no formato .XML, que foi utilizado como base no algoritmo dessa interface.

Além disso, quando ocorre a detecção e um rosto é identificado, um quadrado é desenhado limitando toda a região do rosto, pois é somente dentro dessa região que o filtro gaussiano será aplicado. As bordas desse quadrado são incolores.

4.2.2.3 Testes de funcionamento

Os testes foram realizados utilizando primeiramente apenas um frame com dois rostos para serem detectados, em seguida apenas um frame com 13 rostos para serem detectados e por último uma subamostra de vídeo da coleta de dados, contendo somente a

atividade comer, onde o voluntário está a todo momento com o rosto de frente para a câmera. O vídeo da subamostra utilizado tinha duração de 00:10 segundos, com um 10 fps, totalizando 50 frames e apenas um rosto a ser detectado por frame. Os parâmetros para detecção *scaleFactor* e *minNeighbors* foram respectivamente 1.01 e 7. Foi implementado um contador de frames, para contar quantos rostos fossem detectados em um frame. O resultado dos testes pode ser visto na TABELA 2.

Tabela 2 - Resultado dos testes de funcionamento do algoritmo de detecção de faces

TESTE	FRAMES	<i>scaleFactor</i>	<i>minNeighbors</i>	FACES	
				EXISTENTES	DETECTADAS
TESTE 1	1	1.01	7	2	2
TESTE 2	1	1.01	7	13	17
TESTE 3	100	1.01	7	1	1 a 3

Para imagens com rostos grandes como no teste 1, o algoritmo detectou o corretamente os dois rostos. Para a imagem do teste 2, todos os 13 rostos foram detectados, porém, alguns pixels da imagem foram confundidos como rosto, resultando em 17 faces detectadas.

Para o teste com o vídeo da atividade *eating*, o rosto da pessoa foi detectado em 87 frames corretamente, porém, em alguns frames objetos do cenário eram confundidos com um rosto e também recebiam o borrão. Houveram frames onde até 3 faces foram detectas, sendo 2 erroneamente. Na Figura 30 é possível ver uma região próxima da pessoa, onde os pixels foram confundidos com um rosto.



Figura 30-
Frame
com faces
detectadas

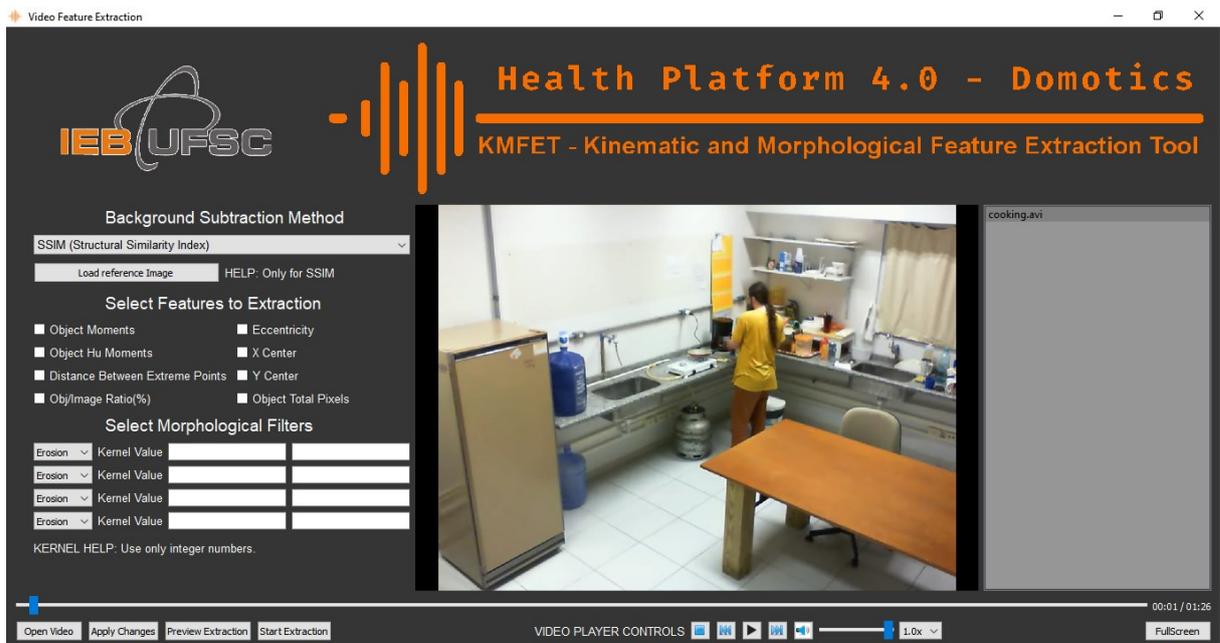
Fonte: autor.

4.2.3 Kinematic and Morphological Feature Extract Tool - KMFET

4.2.3.1 Overview da interface

A ferramenta de extração de características cinemáticas e morfológicas permite ao usuário carregar um vídeo, selecionar filtros para o pré-processamento com transformações morfológicas, extrair características relacionadas ao maior objeto em cena (que serão descritas abaixo) e salvar em um arquivo .CSV. A tela principal pode ser vista na Figura 31.

Figura 31 - Tela principal da KMFET



Fonte: autor

Inicialmente o usuário carrega um ou mais vídeos através do botão “Open Video”, de qualquer diretório disponível, que pode ser nos formatos .AVI, .WMV e MPEG. Existe um controle multimídia de vídeo, com botões de play, pause\stop, avançar e retroceder, inclusive um controle de velocidade de reprodução, que pode ser 0.5 (metade da velocidade normal), 1 (velocidade normal) ou 2x (dobro da velocidade normal).

É possível ver a lista com o nome dos vídeos carregados ao lado direito da tela de reprodução do vídeo. Ao clicar em cima do nome do vídeo ele será selecionado para reprodução e para a extração de características (só é possível realiza a extração de características de um vídeo por vez). Existe também o botão “Fullscreen” para reproduzir o vídeo em tela cheia (ao ativar o modo tela cheia, não é possível controlar o tempo de vídeo, somente assistir).

Depois que o vídeo foi carregado o primeiro passo é escolher qual o método de extração de fundo vai ser utilizado, na seção “*Background Subtraction Method*” através de um botão do tipo *dropdown*, podendo escolher entre duas opções: SSIM e KNN.

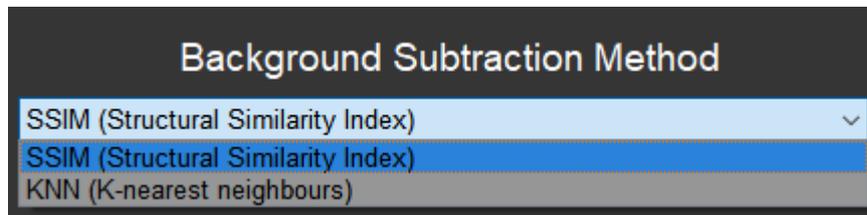
SSIM - *Structural Similarity Index Measure* (Medida do Índice de Similaridade Estrutural) é um método utilizado para medir semelhanças entre duas imagens, baseada em uma imagem inicial de referência (uma imagem sem modificações ou distorções). Esse modelo considera uma distorção na imagem como uma mudança na informação estrutural, e também considera fenômenos como iluminância e contraste (JIN, et.al, 2020) (NASR, et.al, 2017). O algoritmo utilizado pode ser definido pela função abaixo:

$$SSIM(x, y) = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)} \quad (2)$$

KNN - *K-nearest neighbours* (Método dos vizinhos mais próximos) é um método não paramétrico utilizado para classificação e regressão. O aprendizado é baseado na similaridade de um vetor com o outro, e esses vetores podem ser de n dimensões (FERREIRA, et.al, 2020). A biblioteca OpenCv oferece um algoritmo chamado *backgroundSubtractorKNN*, que realiza subtração de fundo da imagem atual pela imagem anterior, de acordo com os vizinhos mais próximos. No caso de um vídeo, o frame atual poderá ter sempre um frame anterior como imagem de “referência”.

Na Figura 32 contém um print da tela principal da interface, que quando clicado no botão *dropdown*, surgem as opções para o método de extração de fundo. Como *default* está o método SSIM.

Figura 32 - Opções de métodos de extração de fundo



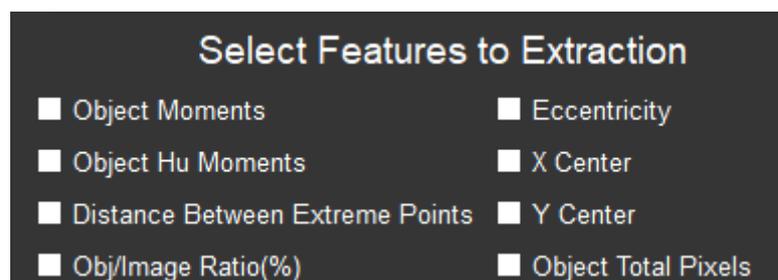
Fonte: autor.

Caso a escolha seja a opção SSIM, será necessário carregar uma imagem de referência através do botão “*Load Reference Image*” (que também está na seção “*Background Subtraction Method*”) nos formatos de arquivo de imagem .JPG ou .PNG. Ao lado desse botão existe uma mensagem de ajuda, instruindo o usuário a carregar a imagem apenas para o método SSIM.

Após escolher o método de subtração de imagem, é necessário escolher quais serão os atributos extraídos da imagem. Na seção “*Select Features to Extraction*” é possível escolher quais serão essas características clicando no botão do tipo *checkbox* referente a cada característica. Os *checkbox* selecionados definem quais características serão extraídas para o arquivo .CSV ao final de todas as etapas de extração.

Algumas características dependem de valores de outras (*e.g* Centro em X depende de valores dos momentos do objeto), nesse caso, mesmo que o atributo não tenha sido selecionado, ele será extraído para realizar o cálculo da outra característica, porém não será salvo no arquivo final. Na Figura 33 podemos ver um print da tela com a seção de seleção de *features*.

Figura 33 – Opções de atributos para serem extraídos do objeto de interesse.



Fonte: autor.

Abaixo serão explicadas cada característica que pode ser extraída do objeto de interesse e o algoritmo utilizado.

- **Object Moments:** No campo de processamento de imagens, um momento é classificado como uma média ponderada da intensidade dos pixels de uma imagem. Esses momentos servem para descrever um objeto após o processo de segmentação e podem extrair parâmetros como orientação, centroide e área do objeto. Esses momentos podem ser definidos para uma imagem digital $f(x,y)$ através de (3) e os momentos centrais, que são invariantes translacionais, podem ser definidos através de (4) (HU, 1962) (TEAGUE, 1980) (BOYCE, HOSSACK, 1983) (HUANG, LENG, 2010).

$$m_{pq} = \sum_{x=1}^{nx} \sum_{y=1}^{ny} x^p y^q f(x,y) \quad (3)$$

Onde o $p+q$ é a ordem do momento (e.g m11), nx é a largura digital da imagem e ny é a altura digital da imagem. Com esses momentos definidos, podemos encontrar informações como o centroide do objeto. Os momentos extraídos do objeto neste trabalho são calculados automaticamente utilizando a função `cv2.moments`, da biblioteca `openCV`, e vão até a ordem 3.

$$u_{pq} = \sum_{x=1}^{nx} \sum_{y=1}^{ny} (x - x_c)^p (y - y_c)^q f(x,y) \quad (4)$$

- **Object Hu Moments:** Os momentos de Hu, ou momentos invariantes (Im), são representados por sete equações (5), que definem sete momentos invariantes de um objeto de interesse (Hu0, Hu1, Hu2, Hu3, Hu4, Hu5 e Hu6), onde os seis primeiros são invariantes à rotação, escala reflexão e translação. O sétimo momento sofre mudança de sinal quando a imagem é refletida (Hu6 = 2 passa a ser Hu6= -2 após o reflexo da imagem). Os momentos invariantes extraídos neste trabalho são calculados

automaticamente utilizando a função cv2.HuMoments, da biblioteca openCv (HU, 1962) (BOYCE, HOSSACK, 1983) (HUANG, LENG, 2010).

$$\begin{aligned}
 I_1 &= n_{20} + n_{02} \\
 I_2 &= (n_{20} - n_{02})^2 + (2n_{11})^2 \\
 I_3 &= (n_{30} - 3n_{12})^2 + (3n_{21} - n_{03})^2 \\
 I_4 &= (n_{30} + n_{12})^2 + (n_{21} + n_{03})^2 \\
 I_5 &= (n_{30} - 3n_{12})(n_{30} + n_{12})[(n_{30} + n_{12})^2 - 3(n_{21} + n_{03})^2] \\
 &\quad + (3n_{21} - n_{03})(n_{21} + n_{03})[3(n_{30} + n_{12})^2 - (n_{21} + n_{03})^2] \\
 I_6 &= (n_{20} - n_{02})[(n_{30} + n_{12})^2 - (n_{21} + n_{03})^2] + 4n_{11}(n_{30} + n_{12})(n_{21} + n_{03}) \\
 I_7 &= (3n_{21} - n_{03})(n_{30} + n_{12})[(n_{30} + n_{12})^2 - 3(n_{21} + n_{03})^2] \\
 &\quad + (n_{30} - 3n_{12})(n_{21} + n_{03})[3(n_{30} + n_{12})^2 - (n_{21} + n_{03})^2]
 \end{aligned} \tag{5}$$

- **X Center:** O Centro em X representa o centro de massa em X do objeto de interesse, que pode ser calculado a partir dos momentos regulares m_{10} e m_{00} através de (7).

$$x_c = \frac{m_{10}}{m_{00}} \tag{7}$$

- **Y Center:** O Centro em Y representa o centro de massa em Y do objeto de interesse, que pode ser calculado a partir dos momentos regulares m_{01} e m_{00} através de (8)

$$y_c = \frac{m_{01}}{m_{00}} \tag{8}$$

Com os valores de centro em X e valores de centro em Y, podemos calcular o centro de massa de um objeto de interesse.

- **Distance Between Extreme Points:** Representa a distância entre os pontos extremos de contorno de um objeto de interesse. Os pontos de contorno extremos são: o ponto mais distante encontrado na parte superior do contorno de um objeto (PS – Ponto superior); o ponto mais distante encontrado na parte inferior do contorno de um objeto (PI – Ponto Inferior); o ponto mais

distante encontrado à direita do contorno de um objeto (PD – Ponto a direita); o ponto mais distante encontrado à esquerda de um objeto (PE – Ponto a esquerda). Ao encontrar esses pontos, uma linha é traçada entre PS e PI e outra linha é traçada entre PD e PE. O valor da distância dessas linhas em pixels compõe os valores que serão extraídos, que são: “*Distance between lowest point to higher*” e “*Distance between rightmost point to leftmost point*”. Esses 4 pontos são extraídos automaticamente através da função `cv2.contourArea`, existente na biblioteca `openCv`.

- ***Eccentricity***: A excentricidade do objeto é representada pela relação entre dois pontos extremos do objeto quem passem pelo maior eixo e pelo eixo ortogonal. Como o objeto de interesse (humano) não possui forma geométrica definida, utilizamos uma aproximação em formato de elipse para calcular a excentricidade, definindo uma função fixa onde o menor eixo será sempre o numerador b , definido em (9).

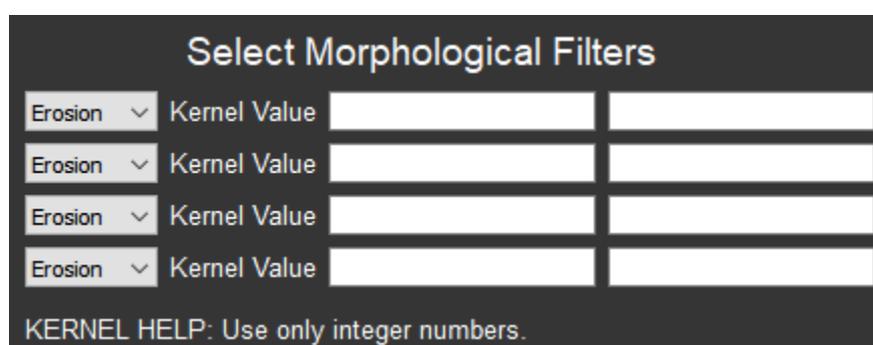
$$Eccentricity = \sqrt{1 - \frac{b^2}{a^2}} \quad (9)$$

- ***Object Total Pixels***: É a quantidade total de pixels presentes no objeto de interesse, calculado a partir do contorno, pela função `cv2.contourArea`. É importante destacar que os pixels do contorno também são considerados como parte do total de pixels presentes no objeto.
- ***Obj/Image Ratio (%)***: É a relação da quantidade de pixels do objeto pela quantidade total de pixels do quadro, obtido pela divisão da quantidade de pixels do objeto pela quantidade total de pixels no quadro.

O usuário pode escolher algumas técnicas de morfologia matemática, que serão aplicadas logo após o método de extração de fundo e antes da extração dos atributos acontecer. Os métodos de extração de morfologia matemática estão na seção “*Select Morphological Filters*”, entre eles é possível utilizar o método de erosão, dilatação, abertura e fechamento.

A interface possui 4 botões do tipo *dropdown*, que permitem ao usuário utilizar até 4 técnicas de morfologia matemática, podendo ser todas iguais ou todas diferentes. Ao lado de cada botão existem dois campos correspondentes aos valores da máscara (*kernel*) que será utilizada, que devem receber apenas números inteiros (e.g. 1,2,3,10,1000) conforme a mensagem de ajuda ao usuário. A ordem de aplicação das técnicas será na mesma ordem dos botões *dropdown* (e.g. o botão superior corresponde a primeira técnica a ser aplicada). Na Figura 34 é possível ver um print da interface contendo a seção de seleção de técnicas de morfologia.

Figura 34 - Opções de técnicas de morfologia matemática e os seus respectivos *Kernels*



Select Morphological Filters		
Erosion ▾	Kernel Value	<input type="text"/>
Erosion ▾	Kernel Value	<input type="text"/>
Erosion ▾	Kernel Value	<input type="text"/>
Erosion ▾	Kernel Value	<input type="text"/>

KERNEL HELP: Use only integer numbers.

Fonte: autor.

Após escolher toda a parametrização, é necessário apertar o botão “*Apply Changes*” para que toda a parametrização seja utilizada para a extração. É possível simular qual será o impacto dos filtros e métodos utilizados, apertando o botão “*Preview Extraction*”, que abrirá uma tela simulando exatamente como será a saída da imagem para ser os atributos serem extraídos. O botão “*Start Extraction*” serve para para iniciar a extração e ao ser pressionado abrirá uma tela para escolher o nome e diretório em que o arquivo .CSV será salvo. Os dados salvos no arquivo .CSV possuem cabeçalho e são separados por vírgula, facilitando na exportação. Na Figura 35 é possível ver como esses dados em .CSV podem ser abertos.

Figura 35 - Dados no arquivo.CSV

	L	M	N	O	P	Q	R	S	T	U	V	W
1	Moment 10	Moment 01	Moment 20	Moment 11	Moment 02	Moment30	Moment21	Moment12	Moment03	Hu moment [0]	Hu moment [1]	Hu moment [2]
2	3935954	2420319.167	1379244062	843355251.2	565516555.6	4.8441E+11	2.9456E+11	1.96242E+11	1.41727E+11	0.380929698	0.112072152	0.005446994
3	3918170	2418506.333	1368054702	840019278.3	565286410.5	4.78737E+11	2.92442E+11	1.95599E+11	1.41724E+11	0.381712597	0.11266937	0.005420326
4	3894094.5	2406660.667	1357078488	834677341.9	562279559.5	4.73996E+11	2.90152E+11	1.94333E+11	1.40956E+11	0.383683444	0.113641992	0.005979351
5	3900893.167	2414224.333	1358072877	836583309.5	564493854.3	4.73868E+11	2.90564E+11	1.94948E+11	1.41673E+11	0.384803942	0.114282564	0.006412214
6	3873534	2401166.667	1348245183	831920732.1	562043728.8	4.70323E+11	2.88891E+11	1.94074E+11	1.41168E+11	0.386695298	0.1156511	0.006505274
7	3875292.667	2405242.333	1348510749	833206777.3	563482566.4	4.70297E+11	2.89295E+11	1.94561E+11	1.41625E+11	0.386647644	0.115476198	0.006370579
8	3873650.667	2403707.667	1347916759	832781638	562852690.5	4.70084E+11	2.89183E+11	1.94394E+11	1.41384E+11	0.385564166	0.114560306	0.00641444
9	3886344.833	2408301.167	1352149828	834384576	563067251.3	4.71513E+11	2.89752E+11	1.94499E+11	1.41254E+11	0.383558022	0.112656419	0.006684957
10	3876191.167	2395621.167	1348920849	830308547.7	558679772.2	4.70503E+11	2.88451E+11	1.93086E+11	1.39842E+11	0.382769458	0.111564601	0.007184397
11	3848800.833	2377891.833	1339697978	824501792.4	554484315.1	4.67391E+11	2.8655E+11	1.91738E+11	1.38802E+11	0.385430774	0.113060152	0.007701836
12	3848052.5	2375533.5	1339025709	823477962.2	553336550.2	4.67023E+11	2.86127E+11	1.91306E+11	1.38393E+11	0.385008471	0.112432042	0.007951177
13	3817520.667	2360414.5	1329025652	818603194.9	551014215.1	4.63736E+11	2.84553E+11	1.90574E+11	1.38072E+11	0.389857829	0.115949215	0.007819735
14	3834194.667	2375167.333	1333198087	822822033.1	554422910.9	4.64642E+11	2.85718E+11	1.91577E+11	1.38856E+11	0.38548533	0.112532658	0.007559463
15	3824526.167	2366901.333	1328641096	819395097.6	551329961.2	4.62655E+11	2.84341E+11	1.90426E+11	1.37785E+11	0.384193518	0.110964667	0.007441319
16	3830486	2375007.167	1329717146	821674372.2	553740723.4	4.62685E+11	2.8495E+11	1.91156E+11	1.38482E+11	0.383349667	0.110439232	0.007162028
17	3832230	2380911.5	1328638030	822773449.6	555465533.6	4.61722E+11	2.85002E+11	1.91556E+11	1.38974E+11	0.382480325	0.110050388	0.00684108
18	3834345.5	2385868.833	1328186130	823744460.8	556886952.9	4.61153E+11	2.8508E+11	1.91875E+11	1.39363E+11	0.381599541	0.109590628	0.006637576
19	3840374.667	2392400.667	1329817110	825503613.7	558854035.8	4.60875E+11	2.84646E+11	1.91076E+11	1.38885E+11	0.380458762	0.108538066	0.006373720

Fonte: autor.

Ao todo podem ser obtidas 27 colunas de informações, são elas: Frame, Frame Time(s), Image (pixels), Object(pixels), Ratio Obj/Image(%), X center, Y center, Eccentricity, Distance between lowest point to higher, Distance between rightmost point to leftmost point, Moment00, Moment01, Moment20, Moment11, Moment02, Moment30, Moment21, Moment12, Moment03, Hu Moment[0], Hu Moment[1], Hu Moment[2], Hu Moment[3], Hu Moment[4], Hu Moment[5], Hu Moment[6].

A coluna *Frame* apresenta o número do *frame* que está sendo analisado; a coluna *Frame time* apresenta o tempo referente a divisão de frames por segundo (*e.g* para um vídeo de 10fps, 1 frame vai ter um tempo correspondente a 0.1 segundos); A coluna *Image Pixels* apresenta o total de pixels que a imagem contém (*e.g.* uma imagem de 640x480 terá um total de 307200 pixels); As outras colunas são referentes aos atributos escolhidos para serem extraídos, cada um em sua respectiva coluna (cada momento e momento de HU é apresentado em uma coluna separada).

4.2.3.2 Algoritmos utilizados

Os algoritmos utilizados foram todos de bibliotecas de código aberto com as suas respectivas licenças de uso incluídas no cabeçalho do código. Para o método de extração de fundo SSIM, as funções pertencem à biblioteca do scikit-image. Para os métodos de extração de características e métodos de morfologia matemática, foram utilizadas funções pertencentes à biblioteca OpenCv. Para a desenhar a interface (botões, telas, labels) foi utilizado o PyQt.

4.2.3.3 Testes de funcionamento

Teste de extração de *features* completo: Foi realizado uma extração de todos os atributos de um vídeo de teste, utilizando o método de subtração de fundo SSIM. Não foi utilizada nenhuma técnica de morfologia matemática. O vídeo utilizado faz parte do acervo de vídeos do laboratório de domótica do IEB-UFSC, e foi gravado de forma controlada contendo uma arena com fundo escuro e apenas uma bolinha de isopor na cor branca, de raio igual a 0,75cm (MARCOLAN, COLUSSI, 2018). A bolinha permanece estática durante todo o tempo do vídeo, com uma duração total de 20 segundos, largura do quadro igual a 852 e altura do quadro igual a 480, com taxa de 29.97fps (*frames* por segundo). Além disso existe, foi utilizada uma imagem de referência do cenário, sem a bolinha branca.

Como se trata de um cenário e um objeto estático, os resultados esperados seriam valores semelhantes para todos os frames e para verificar a veracidade dos dados, foi calculado o desvio padrão de algumas colunas de resultados. Se tratando de valores todos iguais, o desvio padrão teria que ser zero. Os resultados estão na TABELA 3:

Tabela 3 - Resultados do teste completo de extração de *features*

Teste Extração Completa - Vídeo Bolinha Branca	Atributo							
	Image(pixels)	Object(pixels)	Ratio Obj/Image (%)	X center	Y center	Eccentricity	DBLPH	DBRPLP
Média	408960	4035.07508	0.989840256	280	232.089	0.290031949	72.2798083	70.3728
Mediana	408960	4034.5	0.99	280	232	0.29	72.03	70.41
Desvio Padrão	0	3.624024558	0.001254767	0	0.28563	0.004381664	0.5118812	0.173037

Fonte: autor.

Observando os valores da Tabela 3 de alguns atributos, verificamos que o método de extração de fundo, detecção de contorno e posicionamento estão realizando a detecção e extração de dados corretamente, e que o desvio padrão diferente de zero, em diversos atributos, pode ter acontecido por alterações na iluminação, afetando diretamente o objeto de interesse. Outro fator relevante que pode ter contribuído é a variação na frequência de amostragem de vídeo durante a aquisição, pode ser que uma parte tenha sido adquirida com uma velocidade e outra parte com outra (MARCOLAN, COLUSSI, 2018). A arena utilizada e a bolinha de isopor podem ser vistas na Figura 36, junto com o método de extração de fundo aplicado.

Figura 36 - Extração de fundo com detecção do contorno do objeto



Fonte: autor.

4.2.4 Execução da Coleta de Dados

A coleta de dados foi iniciada após recebermos a aprovação do CEPH, o que nos permitiu iniciar os agendamentos dos voluntários. A coleta de amostras para cada voluntário foi realizada de forma individual e com apenas um voluntário por dia, pois foi necessário preparar o ambiente posicionando os objetos que teria interação com o ambiente e o voluntário. Além disso foi necessário utilizar o *Checklist* para garantir que não faltasse nenhum item.

Alguns avisos em forma de cartazes foram anexados na porta de entrada do laboratório, para alertar que estávamos em processo de coleta de amostras e que o silêncio era necessário para não prejudicar o módulo de áudio.

Para a coleta de amostras o ambiente de testes teve suas janelas fechadas e cobertas por cortinas do tipo *blackout*, impedindo totalmente a entrada de iluminação externa. A iluminação do local foi feita com as lâmpadas do próprio laboratório, mantendo-as sempre ligadas. Segundo a norma da ABNT 5413/1992, a iluminação para cozinha de uma residência possui diferentes faixas (medidas em lux) conforme os fatores determinantes. Observando a TABELA 4, verificamos que existem pesos conforme características da tarefa e do observador:

Tabela 4 - Fatores determinantes da iluminação adequada

Características da tarefa e do observador	Peso		
	-1	0	+1

Idade	Inferior a 40 anos	40 a 55 anos	Superior a 55 anos
Velocidade e precisão	Sem importância	Importante	Crítica
Refletância do fundo da tarefa	Superior a 70%	30 a 70%	Inferior a 30%

Fonte: Adaptado de NBR ISO/CIE 8995-1:2013 (ABNT, 2013)

No nosso caso consideramos a idade superior a 55 anos, velocidade e precisão da tarefa sem importância, e refletância do fundo de tarefa média (50%), pois temos paredes brancas mas com eletrodomésticos, móveis e a cortina *blackout* de cores diferentes que tornam as superfícies mais escura. Observando os pesos temos: +1 por ser acima de 55 anos; -1 por velocidade e precisão sem importância; 0 para refletância, pois está entre 30% e 70%. Totalizando o peso como 0 ((+1) + (-1) + (0) = 0). Nesse caso adotamos a iluminação média, que seria 150 lux para o ambiente em geral e 300 lux para locais específicos do ambiente como pia, fogão e mesa.

Para verificar se os valores da iluminação se enquadram nas normas, foram tomadas medidas com o auxílio de um luxímetro (devidamente calibrado), de diferentes locais dentro do ambiente, mantendo sempre na altura da mesa, pois não existem atividades no roteiro de testes que necessitem visualizar objetos no chão, logo consideramos o campo visual de atividade da pessoa sendo da mesa, pia e balcão para cima. Obtemos um valor mínimo de 162 lux (local próximo do equipamento e da câmera) e um valor máximo de 230 lux (sobre a mesa, aproximadamente no centro da visão da câmera), que nos garante que a iluminação está adequada para um ambiente do tipo cozinha de uma residência (entre 150 e 300 lux). Os testes foram realizados apenas durante o dia para facilitar o acesso dos voluntários ao local conforme o horário comercial do IEB-UFSC (8:00 às 17:30).

Esta sequência de atividades que cada voluntário deveria realizar foi explicada e ensaiada com os mesmos de modo *ad libitum*, antes da coleta real de amostras. Os sistemas de registro de dados foram acionados um minuto antes de iniciar a sequência, e a coleta de dados foi interrompida um minuto após a saída do sujeito do local do ambiente. Cada sujeito repetiu a sequência por um número mínimo de vezes, com um intervalo de três minutos entre cada repetição. O intervalo foi importante para deixar o voluntário a vontade e não se cansar (o que poderia prejudicar a coleta). Estas sequências foram registradas durante o período iluminado do dia, com as lâmpadas acesas e as cortinas foscas cobrindo a janela.

Após a coleta de amostras de 7 voluntários diferentes, foi obtido um total de 29 vídeos roteiros completos de atividades. Cada vídeo completo contém todas as atividades previstas (abrir\fechar geladeira, cozinhar, comer, lavar-louça) para realizar a análise e extrair possíveis atributos.

4.2.5 Preparação dos vídeos para construção do banco de dados

Para facilitar a análise, foi utilizada a funcionalidade da ferramenta construída que permite cortar e salvar vídeos, indicando o tempo inicial e final. Com isso foi realizado uma série de cortes em cada vídeo original do roteiro, separando em pequenas amostras contendo somente a atividade prevista (*e.g* cozinhar, lavar louça). Como são 4 atividades diferentes, obtemos um total de 116 vídeos de amostras de atividades, sendo 29 amostras de cada atividade.

Dessas 116 amostras de vídeo, foram extraídos os arquivos .csv referentes a cada uma, contendo todos os atributos disponíveis para serem extraídos. A parametrização utilizada para extração de atributos em cada vídeo, está disponível em um arquivo .txt, dentro de cada pasta referente àquela amostra. Na parametrização consta quais foram os valores de máscara utilizados (Kernels), valores para cada operação de morfologia na imagem (*e.g.* Erosão e dilatação).

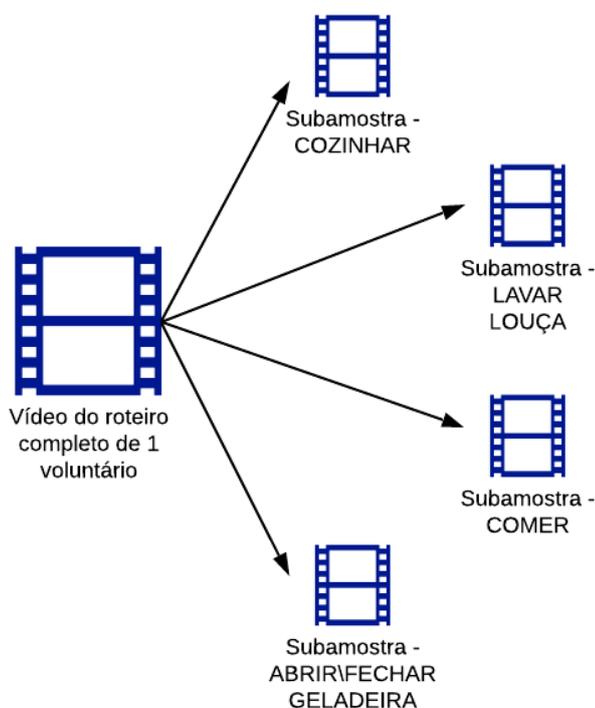
Registrar a parametrização é de extrema importância, pois cada vídeo pode receber uma parametrização diferente, dependendo da iluminação, cores, entre outros aspectos, visto que a análise da entrada dos parâmetros é feita de forma manual, pelo próprio observador, através de testes como verificar se os contornos estão sendo extraídos ou se os ruídos estão sendo minimizados corretamente com o filtro utilizado, tudo isso antes de começar a extrair os atributos. Como a ferramenta permite uma pré-visualização, é possível testar os valores antes de iniciar.

4.2.6 Utilização dos dados para encontrar padrões em atividades domésticas

O objeto dessa validação é tentar encontrar padrões em atividades domésticas realizadas por um voluntário, utilizando vídeos diferentes do mesmo voluntário realizando a mesma atividade. Para tentar encontrar padrões em atividades domésticas, foram selecionados 5 vídeos de roteiros completos de um mesmo voluntário. Cada vídeo de roteiro escolhido possui todas as atividades propostas conforme a coleta de dados.

De cada vídeo de roteiro foram extraídos 4 vídeos de sub amostras diferentes, contendo uma atividade específica, ou seja, a sub amostra não tem tempos de inatividade, somente da atividade escolhida. Para extrair as subamostras do vídeo principal do roteiro, foi utilizada a interface construída VSST (*Video Sample Split tool*). Foram extraídos vídeos de 4 atividades diferentes: abrir\fechar geladeira, cozinhando, comendo e lavando louças. Os nomes utilizados para cada subamostra foram na língua inglês, respectivo a cada atividade: open_close_fridge, cooking, eating, washing_dishes. A extração de subamostras dos vídeos de roteiro pode ser exemplificada através da Figura 37.

Figura 37 – Representação da subdivisão de vídeos do roteiro principal em quatro atividades



Fonte: autor.

Os vídeos das subamostras foram organizados na pasta junto com o vídeo do roteiro original. Na Figura 38 é possível ver como foi organizado o diretório dos vídeos:

Figura 38 - Organização do diretório das subamostras

```

AVI 000003_20-01-2020_15-50-25.avi
AVI cooking_000003_20-01-2020_15-50-25.avi
AVI eating_000003_20-01-2020_15-50-25.avi
AVI open_close_fridge_000003_20-01-2020_15-50-25.avi
AVI washing_dishes_000003_20-01-2020_15-50-25.avi

```

Fonte: autor

Com os vídeos das subamostras preparados, foi possível iniciar a parametrização ideal (tipo de extração de fundo, kernels, tipos de técnicas de morfologia matemática) para extrair as características cinemáticas e morfológicas utilizando a KMFET (*Kinematic and Morphological Feature Extract Tool*). Para encontrar a parametrização ideal para todos os vídeos das subamostras, o autor fez uma série de testes que são descritos a seguir.

Para descobrir qual o método de extração de fundo ideal, o autor testou os dois tipos disponíveis na interface: KNN (*K-nearest neighbours*), o SSIM (*Structural Similarity Index*). O método de KNN não necessita de uma imagem inicial de referência (imagem sem o objeto de interesse), então é uma opção a ser utilizada quando o usuário não possui imagens de referência. Esse método foi testado para a atividade “cozinhar”, sem utilizar nenhuma técnica de morfologia matemática. O resultado podem ser vistos na Figura 39.

Figura 39 – Frame com o resultado da subtração de fundo utilizando KNN para a atividade "cozinhar"



Fonte: autor

Todas as amostras de vídeos gravados dos roteiros de testes possuem 1 minuto de gravação de cenário, antes e depois das atividades do voluntário, com isso foi possível obter um frame de referência para realizar um teste com o método SSIM. A imagem de referência utilizada foi um frame com cenário vazio (sem o voluntário) e com os objetos de interação intactos (e.g. panelas, pratos, cadeira) como pode ser visto na Figura 40.

Figura 40 - Imagem de referência utilizada para o método SSIM



Fonte: autor.

Com a imagem de referência foi possível testar o método o SSIM para a atividade “cozinhar, sem utilizar nenhuma técnica de morfologia matemática. O resultado do teste pode ser visto na Figura 41.

Figura 41 – Frame com o resultado da subtração de fundo utilizando SSIM para a atividade "cozinhar"



Fonte: autor.

Com as imagens resultantes no vídeo, foi possível perceber que a quantidade de pequenos ruídos no resultado do método KNN é superior a quantidade de ruídos do resultado do método SSIM. Além disso, como os movimentos realizados pelo voluntário não são

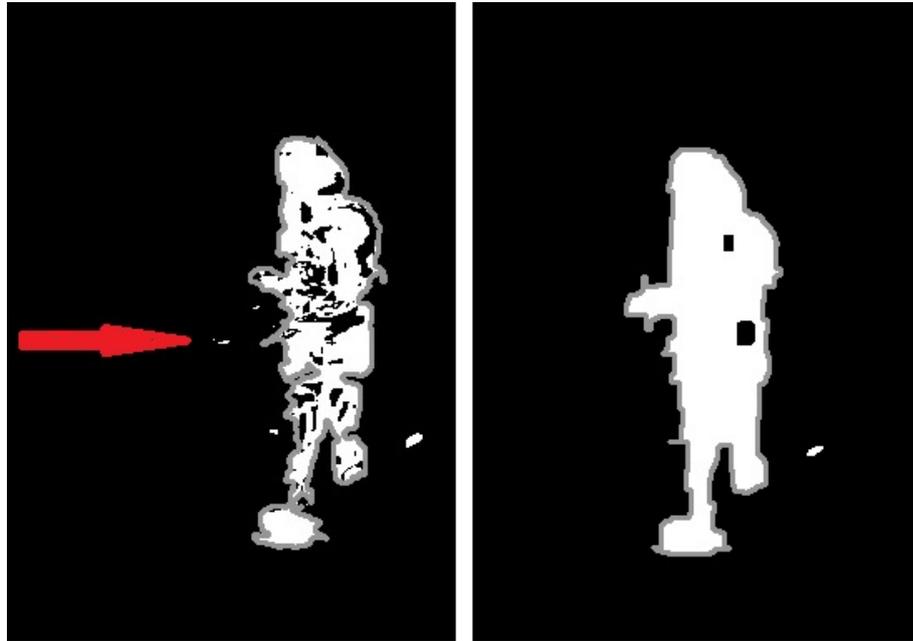
rápidos (são movimentos sutis ao cozinhar), o método KNN não identifica essa movimentação em alguns frames, resultando na ausência de alguns pixels que fazem parte do objeto de interesse (voluntário). Na Figura 39 fica evidente a ausência dos membros inferiores no resultado do método KNN.

Com o método SSIM, a quantidade de ruídos foi reduzida, o que facilita nas próximas etapas e impacta na qualidade dos atributos que serão extraídos, logo, o método SSIM foi o método de extração de fundo escolhido para a atividade “cozinhar”. Os métodos também foram testados para as outras atividades e tiveram resultados semelhantes, logo o método SSIM foi escolhido para ser utilizado como parâmetro em todos os vídeos de todas as atividades.

Como o método de subtração utilizado retornou alguns ruídos, foi testado inicialmente a técnica de morfologia Erosão para a subamostra “comendo”, com o intuito de remover os pequenos ruídos permanentemente. Porém, ao remover os ruídos, muitos pixels que faziam parte do objeto de interesse também foram removidos, deixando alguns “buracos”, logo, ficou claro que para este caso, utilizar a erosão como primeira técnica de morfologia não seria uma boa escolha. Então, ao invés do método de erosão, foi aplicado o método de “Dilatação” como primeira técnica, que agrupou os pixels do objeto de interesse que estavam próximos um do outro. Como o método vale para todos os pixels da imagem, os ruídos também sofreram uma dilatação, então foi necessário aplicar logo em seguida uma Erosão, removendo alguns ruídos permanentemente.

Como é possível ver na Figura 42, ao aplicar a segunda técnica de morfologia (erosão), surgiram alguns “buracos” na região de interesse, porém os ruídos foram removidos. Foram testados valores de Kernel quadrados (*e.g* (1,1) (2,2)) para ambas as técnicas, encontrando valores ideais de (8,8) para a dilatação e (10,6) para a erosão.

Figura 42 - Correção de ruídos utilizando técnicas de morfologia matemática



Fonte: autor.

Para as outras atividades (abrir\fechar geladeira, comer e lavar louça) foi testada a técnica “Fechamento” (que é uma dilatação seguida de uma erosão) com alguns valores de Kernels não quadrados, até remover os ruídos de forma que a região de interesse ficasse contínua (sem buracos). Os resultados foram satisfatórios, pois eliminaram a maioria dos ruídos e conservaram os pixels pertencentes ao objeto de interesse. Alguns objetos que também podemos considerar como ruídos permaneceram na imagem, conforme pode ser visto na Figura 43, na atividade “comendo”. O contorno do voluntário acabou se unindo com o contorno dos objetos da mesa (panela e prato).

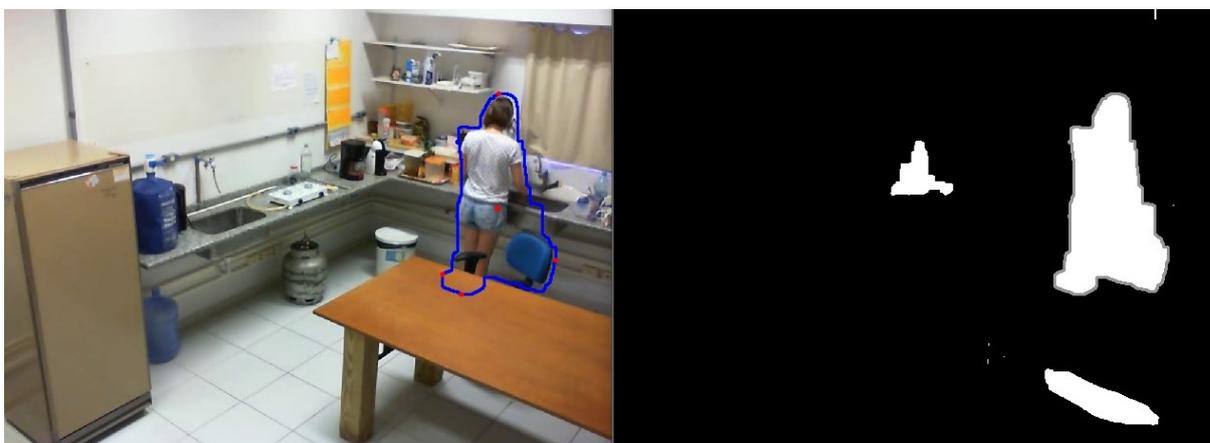
Figura 43 - Resultado da remoção de ruídos utilizando a técnica "fechamento", para a atividade “comer”.



Fonte: autor.

Para a atividade “lavar louça” ocorreu um problema de oclusão, onde a cadeira acabou se tornando parte do contorno do voluntário e o local onde havia um prato azul inicialmente, acabou se tornando um ruído, pois o prato foi colocado em outro local. O resultado pode ser visto na Figura 44.

Figura 44 - Resultado da remoção de ruídos utilizando a técnica "fechamento" para a atividade “lavar louça”.



Fonte: autor.

Como a atividade abrir\fechar geladeira é a primeira atividade a ser executada pelo voluntário no cenário durante a coleta de dados, todos os objetos em cena estão intactos, o que evita ruídos, visto que eles estão exatamente como a imagem de referência. A região da geladeira não possui objetos capazes de obstruir o voluntário, porém ao abrir a geladeira, toda a região passa a ser considerada como um ruído, e o contorno do voluntário se mistura com o contorno da geladeira (Figura 45).

Figura 45 - Resultado da remoção de ruídos utilizando a técnica "fechamento" para a atividade "abrir\fechar geladeira".



Fonte: autor

Ao fechar a geladeira, os contornos vão se separando, até que sobre somente o contorno no voluntário (Figura 46). Então, em alguns frames da atividade "abrir\fechar" geladeira, o que temos é somente um contorno de geladeira aberta.

Figura 46 - Resultado da remoção de ruídos utilizando a técnica "fechamento" para a atividade "abrir\fechar geladeira".



Fonte: autor.

Mesmo que existam outros objetos considerados como ruídos em cena, o algoritmo utilizado na extração, identifica qual é o maior contorno em toda a imagem, entre os objetos identificados, e aplica os métodos de extração de atributos somente no objeto de maior contorno. A parametrização para realizar a extração foi definida após todos os testes e pode ser vista na TABELA 5.

Tabela 5 - Parametrização utilizada para extrair os atributos dos 20 vídeos de atividades

ATIVIDADE	MORFOLOGIA MATEMÁTICA APLICADA	KERNEL UTILIZADO
ABRIR\FECHAR GELADEIRA	Fechamento	Kernel (35,10)
COZINHAR	Dilatação/Erosão	Kernel (8,8)
	Erosão	Kernel (10,6)
COMER	Fechamento	Kernel (35,10)
LAVAR LOUÇAS	Fechamento	Kernel (35,10)

Fonte: autor.

Com a parametrização escolhida, todos os atributos disponíveis no menu da interface de extração foram selecionados e a extração realizada para os 5 vídeos de cada atividade, mantendo o mesmo nome do arquivo de vídeo de cada amostra para o seu respectivo .CSV. É importante manter os arquivos na mesma pasta, pois possuem o nome codificado para cada amostra, que podem se misturar facilmente com outras amostras. Na Figura 47 é possível ver a organização das sub amostras no diretório.

Figura 47 - Organização de arquivos de teste no diretório

Nome	Data de modificação
 cooking_000003_20-01-2020_14-46-13.csv	09/12/2020 14:48
 cooking_000003_20-01-2020_15-01-25.csv	09/12/2020 14:40
 cooking_000003_20-01-2020_15-16-48.csv	09/12/2020 15:14
 cooking_000003_20-01-2020_15-33-10.csv	09/12/2020 15:24
 cooking_000003_20-01-2020_15-50-25.csv	09/12/2020 15:30
 eating_000003_20-01-2020_14-46-13.csv	09/12/2020 14:47
 eating_000003_20-01-2020_15-01-25.csv	09/12/2020 14:40
 eating_000003_20-01-2020_15-16-48.csv	09/12/2020 15:15
 eating_000003_20-01-2020_15-33-10.csv	09/12/2020 15:23
 eating_000003_20-01-2020_15-50-25.csv	09/12/2020 15:30
 open_close_fridge_000003_20-01-2020_14-46-13.csv	09/12/2020 14:44
 open_close_fridge_000003_20-01-2020_15-01-25.csv	09/12/2020 14:37
 open_close_fridge_000003_20-01-2020_15-16-48.csv	09/12/2020 15:10
 open_close_fridge_000003_20-01-2020_15-33-10.csv	09/12/2020 15:20
 open_close_fridge_000003_20-01-2020_15-50-25.csv	09/12/2020 15:26
 washing_dishes_000003_20-01-2020_14-46-13.csv	09/12/2020 14:47
 washing_dishes_000003_20-01-2020_15-01-25.csv	09/12/2020 14:40
 washing_dishes_000003_20-01-2020_15-16-48.csv	09/12/2020 15:12
 washing_dishes_000003_20-01-2020_15-33-10.csv	09/12/2020 15:23
 washing_dishes_000003_20-01-2020_15-50-25.csv	09/12/2020 15:28

Fonte: autor.

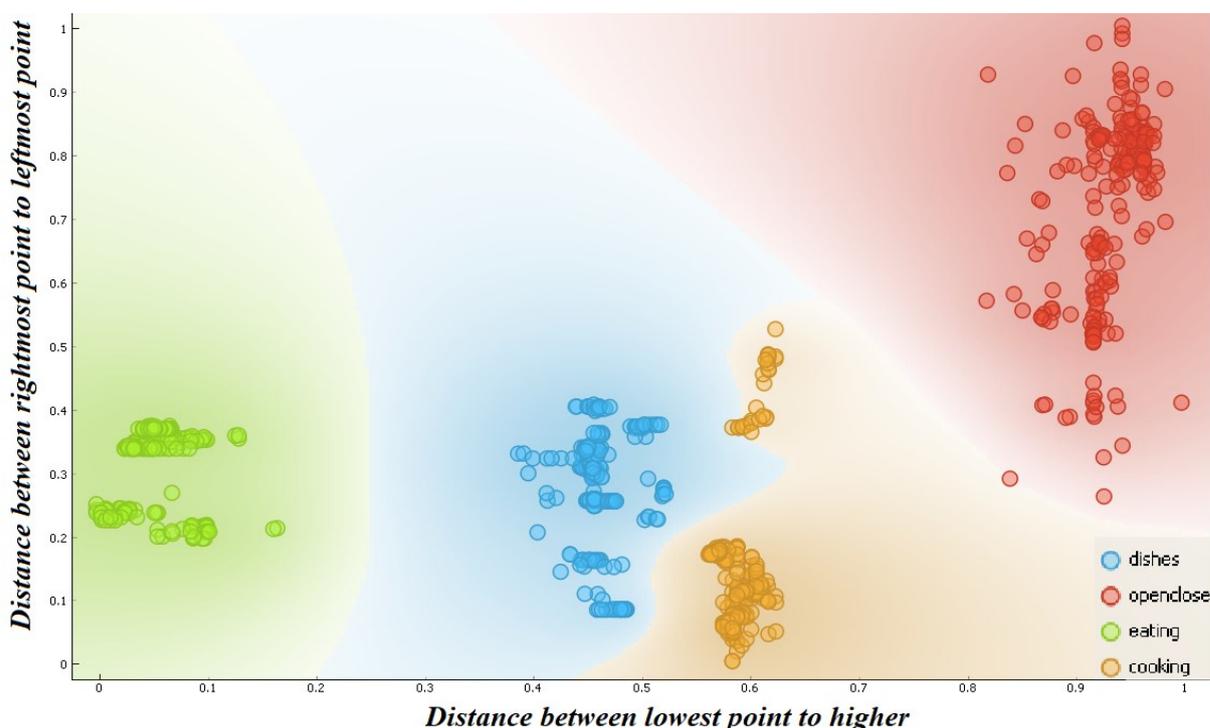
Com a extração de todos as atividades, foi obtido um total de 20 arquivos .CSV, cada arquivo com 27 atributos diferentes. Para realizar uma análise mais detalhada desses dados, foi utilizado um kit de ferramentas chamado Orange (ORANGE, 2020), que permite visualização e mineração de dados em código aberto, além de aprendizados de máquina (supervisionados e não supervisionados).

Com o Orange foi possível inserir os 20 arquivos .CSV simultaneamente, para concatenar todos os arquivos em um único, considerando a mesma faixa de tempo. Foram definidas 4 classes distintas “dishes”, “openclose”, “eating”, “cooking”, utilizando 23 atributos diferentes para aplicar o método K-means (ZHAO, CAO, 2018), que é um método de clusterização que, dentre k grupos, particiona n observações que pertencem a um grupo mais próximo da média. Foram definidos 4 clusters fixos, 300 repetições e um máximo de 500 iterações como parametrização do método de K-means da ferramenta Orange.

O gráfico de dispersão (*Scatter-plot*) para o método utilizado pode ser visto na Figura 48, com representação de dois atributos: no eixo x o atributo “*Distance between lowest point to higher*” e no eixo y o atributo “*Distance between rightmost point to leftmost point*”. É possível observar que os valores se agrupam em 4 clusters diferentes, caracterizando uma atividade específica agrupada.

Os valores agrupados em “dishes” e “cooking” estão próximos entre si e distante das outras duas atividades, em resumo, as duas atividades foram realizadas pelo voluntário em pé, o que caracteriza que os eixos verticais da morfologia extraída são maiores que os eixos horizontais. Para a atividade “openclose” (abrir e fechar a geladeira), os pontos se agrupam distantes mesmo que o voluntário tenha realizado a atividade em pé, isso pode ter ocorrido pelo fato do contorno da geladeira se unir com o contorno do voluntário, retornando uma morfologia em formato “quadrado”, como uma geladeira. A atividade “eating” foi realizada com o usuário sentado, logo a morfologia é distinta das outras, com o eixo horizontal maior ou igual ao vertical em alguns casos.

Figura 48 - Gráfico de dispersão de dados (*Scatter-plot*) dos resultados do método K-means



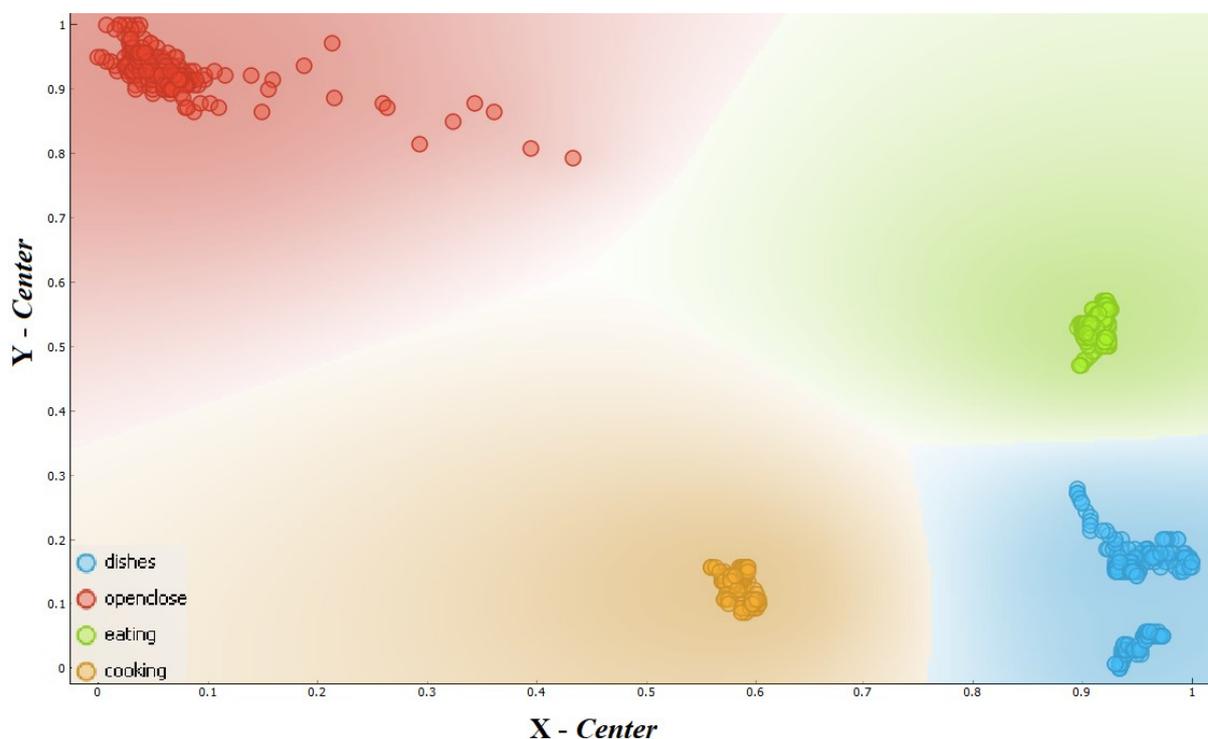
Fonte: autor

Na Figura 49 podemos ver o gráfico de dispersão para os atributos “X- Center” e “Y – Center”, que representam respectivamente o centro em X e o centro em Y. Verificamos que os pontos se agruparam em 4 classes de forma bem próxima entre si e distantes uma classe da outra. As atividades “cooking” e “eating” foram realizadas pelo voluntário sem realizar deslocamentos no cenário, principalmente a atividade “eating” que o voluntário realizou sentado, e através do gráfico verificamos que as duas classes ficaram bem concentradas em uma região.

Para a atividade “dishes” também não foi realizado deslocamento pelo voluntário, porém a divisão em dois grupos pequenos dentro da mesma classe “dishes” pode ter ocorrido devido ao prato e a cadeira, que como relatado anteriormente, se comportaram como ruídos e os seus contornos se uniram com o contorno do voluntário, resultando em uma morfologia diferente e consequentemente com centro em outros pontos.

Para a atividade “openclose”, houve deslocamento do voluntário, no início da atividade e no final da atividade. É possível verificar que existe uma maior concentração de pontos em um local dentro dessa mesma classe, que pode ter acontecido devido ao fato de que na maioria do tempo de execução dessa atividade o usuário ficou parado em frente a geladeira retirando os objetos.

Figura 49 - Gráfico de dispersão de dados (*Scatter-plot*) dos resultados do método *K-means*



Fonte: autor.

Através dos testes realizados utilizando a ferramenta Orange, percebemos que o método *K-means* ilustra a dissociação entre os 4 eventos (cozinhar, comer, lavar louça, abrir\fechar geladeira) e informa que esses parâmetros se mostram promissores para o uso como entradas em redes neurais classificadoras, dada a clusterização que o método *K-means* nos apresenta.

4.2.7 Covid-19.

É importante explicar como a pandemia do COVID – 19 afetou no desenvolvimento do presente trabalho. Conforme o projeto submetido e aprovado pelo Comitê de Ética Humano (CEPH), a primeira etapa da coleta de dados seria com pessoas saudáveis, entre 24 e 55 anos e a segunda etapa de coleta de dados seria realizada com um grupo de idosos (pertencentes ao Núcleo da Terceira Idade da UFSC). A primeira etapa da coleta de dados foi realizada com 7 pessoas entre os meses de dezembro de 2019 e fevereiro de 2020. A segunda etapa da coleta de dados estaria prevista para ser iniciada após a análise dos dados obtidos na primeira etapa, porém, aconteceu o *Lockdown* em março de 2020, onde todas as instituições de ensino públicas foram fechadas para assegurar a saúde de todos os estudantes.

Com isso, os participantes do projeto ficaram impossibilitados de realizar qualquer atividade dentro da universidade, logo o laboratório ficou inacessível para a testes. Todas as atividades previstas para serem realizadas no laboratório, ficarão como propostas de trabalhos futuros, pois até o dia da defesa deste trabalho, a Universidade ainda estará fechada para atividades presenciais (dezembro de 2020). Entre as atividades mais importantes que não puderam ser executadas estão a coleta de dados com voluntários idosos e a gravação de objetos com tamanhos conhecidos (em metros) para realizar o mapeamento do laboratório de testes e realizar a conversão dos pixels para tamanhos aproximado em metros de algumas características, permitindo o cálculo de medidas como a área em metros.

5 DISCUSSÃO

O presente trabalho teve como objetivo geral propor e implementar uma ferramenta para registro e extração de atributos cinemáticos e morfológicos utilizados para detecção de padrões de atividades domésticas em humanos, fundamentado em plataformas de prototipagem eletrônica e de programação em código aberto.

O hardware construído utilizou somente plataformas de código aberto, o que permite a reprodutibilidade do projeto e a aplicação em outras áreas. Em resumo, implementar, usando hardware e software de código aberto, um sistema para aquisição e armazenamento de amostras de imagem obtidas em ambiente doméstico (cozinha), contribui para a comunidade acadêmica realizar estudos utilizando esse tipo de ferramenta, que já existe no mercado porém em formato comercial. Com o desenvolvimento de novos modelos do Raspberry Pi, será possível ampliar os limites de processamento que temos no Raspberry Pi model 3, e utilizar algoritmos cada vez mais robustos em unidades pequenas e de fácil instalação como o RPi.

A ferramenta VSST criada para subdividir amostras de vídeos, contribui para a divisão de amostras grandes de vídeo em amostras menores, como no caso da nossa coleta de dados para pesquisa, onde um vídeo do roteiro completo de atividades contém diversas atividades (*e.g.* lavar louça, cozinhar, comer). Além das atividades definidas no roteiro, existem diversas micro atividades que poderiam ser extraídas do vídeo principal para serem analisadas de forma isolada, como por exemplo a forma de andar dentro da cozinha ou até mesmo a forma com que um voluntário serve a comida da panela para o prato. Talvez uma melhoria interessante para esta ferramenta seria realizar diversas extrações de subamostras ao mesmo tempo e também realizar a reprodução frame a frame dos vídeos.

Os algoritmos utilizados para a detecção facial na ferramenta FBDT utilizam um modelo pré treinado disponibilizado pelo OpenCv e conseguiram detectar as faces dos voluntários no ambiente doméstico cozinha, porém, houveram frames onde objetos do cenário foram confundidos com rostos e acabaram sendo detectados. Uma possível solução que poderia reduzir esse tipo de problema seria criar um modelo pré treinado de faces utilizando somente imagens de faces de pessoas em situações domésticas ou até mesmo criar um modelo de treinamento individual para cada pessoa. Ao pensarmos em uma aplicação da Plataforma de Saúde Ubíqua na casa de um idoso que mora sozinho, utilizando somente imagens dele para treinar a rede e as mesmas imagens dele do dia a dia funcionando como validação, poderia garantir uma proteção robusta de privacidade.

Outra solução para este problema de detecção facial seria a implementação de redes neurais artificiais, que apesar de consumirem mais memória do dispositivo embarcado, podem retornar uma solução mais robusta. Se utilizarmos um modelo como RPi 4, que foi lançado recentemente com 8 GB de memória RAM, poderíamos ampliar a quantidade de algoritmos utilizados sem comprometer o processamento de dados.

Atualmente existem diversos algoritmos (*open source*) que permitem ao usuário extrair atributos cinemáticos e morfológicos, mas muitas vezes o usuário precisa realizar uma implementação, o que exige certo conhecimento específico em programação. Com uma ferramenta concentrando todos esses métodos de processamento de imagens em uma única interface, como a KMFET, o usuário consegue aplicar técnicas de morfologia matemática, realizar a extração de contornos e inclusive optar pelos métodos que mais se encaixam ao projeto que ele está desenvolvendo.

Para casos como o do nosso projeto, onde as amostras obtidas contêm humanos realizando atividades doméstica em um ambiente, (cozinha), os métodos utilizados podem ser aqueles voltados para movimentação lenta do objeto de interesse, pois um humano sentado e comendo, por exemplo, não realiza movimentos rápidos. É possível verificar isso ao tentar utilizar a técnica de extração de fundo como a KNN nos vídeos de amostras coletadas de atividades domésticas, onde objeto de interesse acaba vindo pela metade como aconteceu na amostra “cozinando”, pois havia ausência de movimentação nos membros inferiores do voluntário.

As técnicas de morfologia utilizadas na extração de características em cada atividade, bem como toda a parametrização utilizada, serviram para os tipos de atividades e para o ambiente utilizado, porém, poderia não servir para outros tipos de ambientes, como um banheiro por exemplo. Os atributos encontrados podem não apresentar relevância para representar padrões em outras situações, e para cada situação poderá ser necessário um engenheiro para levantar os parâmetros necessários para identificar padrões. Caso esses parâmetros não existam na ferramenta proposta neste trabalho, será possível realizar uma implementação de melhoria, pois se tratando de uma plataforma *open source*, temos acesso a bibliotecas de visão computacional que se atualizam a todo momento.

As amostras de vídeos obtidas e o banco de dados resultante, representam uma grande evolução para a Plataforma de Saúde Ubíqua como um todo, pois com as amostras é

possível seguir para a segunda etapa do projeto, que é utilizar redes neurais para detectar padrões em atividades domésticas de forma automática.

Definir um padrão mínimo de amostras necessárias para obter dados suficientes para treinar a rede, pode facilitar a aquisição de dados realizada com idosos. Alguns idosos escolhidos para realizar o teste podem ser debilitados, cansar ou esquecer atividades facilmente. Talvez, repetir cinco atividades diferente em um roteiro por cinco vezes seja cansativo para um idoso. Além disso pode ser que não seja necessário utilizar cinco amostras de um mesmo voluntário para obter um padrão de atividades, talvez uma quantidade menor já seja suficiente.

É possível analisar e realizar melhorias na dinâmica de coleta de dados, bem como a mudança em posicionamento e atividades executadas. Na coleta de dados atual, ao realizar a atividade “lavar louça”, o usuário tem parte dos seus membros inferiores obstruídos pela mesa e cadeira, o que causou ruídos e mistura de contornos do humano com o contorno da cadeira. É fato que em uma situação de vida real, existem diversos objetos e móveis que podem obstruir uma obtenção de dados, mas nesta fase de extração de características, uma visão completa da articulação do voluntário no cenário, poderia proporcionar um volume maior de dados em relação a morfologia.

O posicionamento da câmera também é um fator que pode ser ajustado para obter imagens de ângulos diferentes, ângulos que talvez representem melhor as características e cinemáticas do humano no ambiente (*e.g.* uma visão lateral da atividade lavar-louça poderia representar uma visão lateral de movimentação dos braços e mãos do voluntário).

A obtenção de um banco de dados de imagens e sons de atividades domésticas em humanos (como os vídeos obtidos neste trabalho) pode permitir um conjunto de ensaios sobre modos de extração de outros parâmetros de imagem e também de sons (como as amostras de áudio obtidas por SIRIDAKIS, 2020), bem como a exploração de diferentes modos de fusão destes dados com vistas à sua classificação por diferentes algoritmos de aprendizado de máquina.

Como plano de estudos futuro para a Plataforma de Saúde Ubíqua temos o desenvolvimento de um método para embarcar todos os algoritmos de extração de atributos cinemáticos e morfológicos (apresentados neste trabalho), bem como os algoritmos de extração de atributos de áudio e dados multissensoriais (SIRIDAKIS, 2020) em um hardware que possa ser adaptável a diferentes condições de uso (*e.g.* ambientes externos e locais

úmidos como um banheiro), que seja capaz de extrair, processar e analisar dados domóticos que sejam úteis em uma Plataforma de Saúde Ubíqua.

Em resumo, como um teste geral para a capacidade de os parâmetros escolhidos serem úteis para o uso em redes neurais, usamos a ferramenta Orange em uma sequência de etapas de extração e classificação utilizando o método *k-means*. As distâncias encontradas entre esses grupos são grandes suficientes para enunciar que teríamos condições promissoras de uso em redes neurais classificadoras, porém, nas condições dos nossos experimentos e para esses parâmetros. Talvez em outros ambientes e outras condições de iluminação e cores, não teríamos o mesmo resultado.

Além disso, os algoritmos de extração, quando aplicados a objetos imóveis mostram que temos uma certa margem de erro e variação, o que é comum, e pode ser causado por uma cadeia de erros que devem ser investigadas. Essas aparentes fragilidades do processo, são apenas aparentes, pois para cada situação é esperado que o desempenho mude, no entanto, o fato de todo esse processo ser desenvolvido baseado em plataformas de código aberto (*open-source*) indica que pode ser rápido e facilmente modificado para se adaptar a novas situações ou novos ambientes com facilidade, pois essas plataformas são continuamente atualizadas.

6 CONCLUSÃO

O sistema de aquisição de dados de áudio, sensores e principalmente de vídeo, adquiriu com sucesso todas as amostras necessárias, permitindo o controle de parâmetro específicos, antes da coleta, como a quantidade frames/pacotes por segundo. Além disso, foi possível construir um sistema utilizando *hardware* e software livre, de forma estrutural discreta.

O laboratório de testes, estruturado especialmente para realizar a coleta de dados, serviu para o seu propósito, permitindo aos voluntários a interação com o ambiente de forma que atividades domésticas fossem de fato simuladas em um ambiente do tipo cozinha. O protocolo criado com as etapas para a coleta de dados, serviu para o seu propósito, proporcionando ao voluntário da coleta uma sensação de liberdade ao realizar a atividade, além de garantir a segurança da proteção de imagens através da proteção de dados faciais.

O projeto submetido e aprovado pelo CEPH, garantiu a estrutura de planejamento ética do presente trabalho, bem como a Plataforma como um todo. Com a aprovação, é possível que os próximos trabalhos sejam desenvolvidos utilizando voluntários idosos, dentro das mesmas condições propostas no planejamento de coleta de dados contida no presente trabalho. Atualmente o projeto conta com um total de 29 vídeos de roteiro de atividades completos (além dos áudios e dados multissensoriais respectivos a esses vídeos), que podem ser utilizados para diversas análises de comportamento, bem como realizar a extração de características. A condição de utilização é a proteção facial do voluntário.

A interface de *Video Sample Split tool - VSST* serviu para separar os vídeos de roteiros completos em sub amostras, contendo somente atividades específicas como lavar louça, cozinhar, comer, abrir e fechar uma geladeira. A separação de uma amostra grande de vídeo em sub amostras permite analisar especificamente uma atividade em relação ao tempo, separando, por exemplo, de uma amostra onde não existem atividades de interesse sendo realizadas.

A interface *Face Blur and Detection Tool – FBDT* serviu para proteger a imagem de cada voluntário nas sub amostras analisadas, e mesmo que não tenha sido possível alcançar um desempenho de detecção de 100% das faces, reduziu o trabalho manual que o usuário teria para borrar um rosto frame a frame.

A interface KMFET de extração de características cinemáticas e morfológicas, que foi a interface principal deste trabalho, permitiu a extração de diversos atributos relevantes e utilizados para representar um padrão em alguma atividade específica. É possível identificar padrões em atividades que representam padrões para uma pessoa em específico, como a forma de abrir a geladeira, por exemplo. Uma alteração nesse padrão pode representar que algo está errado com a pessoa, e utilizando a interface, foi possível obter os dados brutos que representam esses padrões.

REFERÊNCIAS

- ABD ELAZIZ, Mohamed; LU, Songfeng. Many-objectives multilevel thresholding image segmentation using knee evolutionary algorithm. *Expert Systems With Applications*, v. 125, p. 305-316, 2019.
- ABTOY, Anouar et al. Ambient Assisted living system's models and architectures: A survey of the state of the art. *Journal of King Saud University-Computer and Information Sciences*, v. 32, n. 1, p. 1-10, 2020.
- ACHARYA, Tinku; RAY, Ajoy K. *Image processing: principles and applications*. John Wiley & Sons, 2005.
- AHMED, Imran et al. Person detector for different overhead views using machine learning. *International Journal of Machine Learning and Cybernetics*, v. 10, n. 10, p. 2657-2668, 2019.
- ALONSO, Mercedes et al. Background-Subtraction Algorithm Optimization for Home Camera-Based Night-Vision Fall Detectors. *IEEE Access*, v. 7, p. 152399-152411, 2019.
- ALSMADI, Mutasem K. Content-Based Image Retrieval Using Color, Shape and Texture Descriptors and Features. *Arabian Journal for Science and Engineering*, p. 1-14, 2020.
- AMMAR, Sirine et al. Moving objects segmentation based on deepsphere in video surveillance. In: *International Symposium on Visual Computing*. Springer, Cham, 2019. p. 307-319.
- ARMDEVELOPER, ArmDeveloper Cortex-A53: Overview. Página inicial. Disponível em: <<https://developer.arm.com/ip-products/processors/cortex-a/cortex-a53>>. Acesso em: 3 de outubro de 2020.
- ASLAN, Muhammet Fatih; DURDU, Akif; SABANCI, Kadir. Human action recognition with bag of visual words using different machine learning methods and hyperparameter optimization. *Neural Computing and Applications*, v. 32, n. 12, p. 8585-8597, 2020.
- ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS. NBR ISO/CIE 8995-1:2013.
- AZODO, Ijeoma et al. Opportunities and Challenges Surrounding the Use of Data From Wearable Sensor Devices in Health Care: Qualitative Interview Study. *Journal of Medical Internet Research*, v. 22, n. 10, p. e19542, 2020.
- BADRINARAYANAN, Vijay; KENDALL, Alex; CIPOLLA, Roberto. Segnet: A deep convolutional encoder-decoder architecture for image segmentation. *IEEE transactions on pattern analysis and machine intelligence*, v. 39, n. 12, p. 2481-2495, 2017.
- BARDRAM, Jakob E.; MATIC, Aleksandar. A decade of ubiquitous computing research in mental health. **IEEE Pervasive Computing**, v. 19, n. 1, p. 62-72, 2020.

BELSHAW, Michael et al. Towards a single sensor passive solution for automated fall detection. In: 2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE, 2011. p. 1773-1776.

BIBI, Shafina; ANJUM, Nadeem; SHER, Muhammad. Automated multi-feature human interaction recognition in complex environment. *Computers in industry*, v. 99, p. 282-293, 2018.

BOULAY, Bernard; BRÉMOND, Francois; THONNAT, Monique. Applying 3d human model in a posture recognition system. *Pattern Recognition Letters*, v. 27, n. 15, p. 1788-1796, 2006.

BOUWMANS, Thierry; EL BAF, Fida; VACHON, Bertrand. Background modeling using mixture of gaussians for foreground detection-a survey. *Recent patents on computer science*, v. 1, n. 3, p. 219-237, 2008.

BOYCE, James F.; HOSSACK, W. J. Moment invariants for pattern recognition. *Pattern Recognition Letters*, v. 1, n. 5-6, p. 451-456, 1983.

BOZAN, Karoly; BERGER, Andrew. Revisiting the technology challenges and proposing enhancements in ambient assisted living for the elderly. In: *Proceedings of the 52nd Hawaii International Conference on System Sciences*. 2019.

BRADSKI, Gary. The opencv library. *Dr Dobb's J. Software Tools*, v. 25, p. 120-125, 2000.

BROWN, Ian et al. The ethical challenges of ubiquitous healthcare. *International Review of Information Ethics*, v. 8, n. 12, p. 53-60, 2007.

CAO, Yu et al. Recognize human activities from partially observed videos. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2013. p. 2658-2665.

CAO, Yuanyuan; TAO, Linmi; XU, Guangyou. An event-driven context model in elderly health monitoring. In: *2009 Symposia and Workshops on Ubiquitous, Autonomic and Trusted Computing*. IEEE, 2009. p. 120-124.

CASTAÑEDA, William Alberto Cruz et al. Novo Paradigma Da Engenharia Clínica Na Integração De Tic'S Para Criação De Ambientes Ubiquos E De Interoperabilidade Na Saúde. *Dissertação (Dissertação em Engenharia Elétrica) – UFSC*. Santa Catarina, p. 35-36. 2011.

CHEN, Jay et al. Wearable sensors for reliable fall detection. In: *2005 IEEE Engineering in Medicine and Biology 27th Annual Conference*. IEEE, 2006. p. 3551-3554.

CHITYALA, Ravishankar; PUDIPEDDI, Sridevi. *Image processing and acquisition using Python*. CRC Press, 2020.

CHO, Nam-Gyu et al. Compositional interaction descriptor for human interaction recognition. *Neurocomputing*, v. 267, p. 169-181, 2017.

COLUSSI, Alex; MARCOLAN, João A.; Marino-Neto, José. Algoritmos para detecção automática de múltiplos objetos e segmentos em animais de laboratório. In: Anais do V Congresso Brasileiro de Eletromiografia e Cinesiologia e X Simpósio de Engenharia Biomédica. Anais. Uberlândia(MG) Center Convention Uberlândia, 2018. Disponível em: <<https://www.even3.com.br/anais/cobecseb/79056-ALGORITMOS-PARA-DETECCAO-AUTOMATICA-DE-MULTIPLOS-OBJETOS-E-SEGMENTOS-EM-ANIMAIS-DE-LABORATORIO>>. Acesso em: 11/06/2019 10:41.

COOK, Diane J.; DAS, Sajal K. How smart are our environments? An updated look at the state of the art. *Pervasive and mobile computing*, v. 3, n. 2, p. 53-73, 2007.

CUCCHIARA, Rita; GRANA, Costantino; PRATI, Andrea; TARDINI, Giovanni; VEZZANI, Roberto. Using computer vision techniques for dangerous situation detection in domotic applications. 2004. 1 - 5. 10.1049/ic:20040088.

DAWADI, Prafulla Nath; COOK, Diane Joyce; SCHMITTER-EDGEcombe, Maureen. Automated cognitive health assessment from smart home-based behavior data. *IEEE journal of biomedical and health informatics*, v. 20, n. 4, p. 1188-1194, 2016.

DE MIGUEL, Koldo et al. Home camera-based fall detection system for the elderly. *Sensors*, v. 17, n. 12, p. 2864, 2017.

DE QUEIROZ, José Eustáquio Rangel; GOMES, Herman Martins. Introdução ao processamento digital de imagens. *Rita*, v. 13, n. 2, p. 11-42, 2006.

DIN SPEC, Deutsches Institut für. DIN SPEC 77002:2013-04. 2013.

D-LAB, D-Lab Video, ERGONEES, from Science to innovation. Disponível em: <<https://www.ergoneers.com/en/data-capture-software-and-analysis-software/d-lab/video/>>. Acesso em: 23 de novembro de 2020.

ESQUEF, Israel Andrade; ALBUQUERQUE, Márcio Portes de; ALBUQUERQUE, Marcelo Portes de. Processamento digital de imagens. *Rio de Janeiro*, v. 12, 2003.

EYE TRACKING, Eye tracking: screen-based, Imotions. Disponível em: <<https://imotions.com/biosensor/eye-tracking-screen-based/>>. Acesso em: 23 de novembro de 2020.

FERREIRA, Paulo JS; CARDOSO, João MP; MENDES-MOREIRA, João. kNN Prototyping Schemes for Embedded Human Activity Recognition with Online Learning. *Computers*, v. 9, n. 4, p. 96, 2020.

GARCIA-GARCIA, Belmar; BOUWMANS, Thierry; SILVA, Alberto Jorge Rosales. Background subtraction in real applications: Challenges, current models and future directions. *Computer Science Review*, v. 35, p. 100204, 2020.

GHAZAL, Bilal; AL-KHATIB, Khaled. Smart home automation system for elderly, and handicapped people using XBee. **International Journal of Smart Home**, v. 9, n. 4, p. 203-210, 2015.

GOMEZ, Carles et al. Internet of Things for enabling smart environments: A technology-centric perspective. *Journal of Ambient Intelligence and Smart Environments*, v. 11, n. 1, p. 23-43, 2019.

GONZALES, R. C.; WOODS, R. E. *Processamento de Imagens Digitais*. Tradução de Cristina Yamagami e Leonardo Piamonte. 3ª Edição. São Paulo: Pearson Prentice Hall, 2010.

GRUOSSO, Monica; CAPECE, Nicola; ERRA, Ugo. Human segmentation in surveillance video with deep learning. *Multimedia Tools and Applications*, p. 1-25, 2020.

GUNALE, Kishanprasad; MUKHERJI, Prachi. Indoor Human Fall Detection System Based on Automatic Vision Using Computer Vision and Machine Learning Algorithms. *Journal of Engineering Science and Technology*, v. 13, n. 8, p. 2587-2605, 2018.

HU, Ming-Kuei. Visual pattern recognition by moment invariants. *IRE transactions on information theory*, v. 8, n. 2, p. 179-187, 1962.

HUANG, Zhihu; LENG, Jinsong. Analysis of Hu's moment invariants on image scaling and rotation. In: 2010 2nd International Conference on Computer Engineering and Technology. IEEE, 2010. p. V7-476-V7-480.

JÄHNE, Bernd. *Digital Image Processing 5th revised and extended edition*. Berlin: Springer-Verlag, doi, v. 10, p. 3-540, 2002.

JALAL, Ahmad et al. A spatiotemporal motion variation features extraction approach for human tracking and pose-based action recognition. In: 2015 International Conference on Informatics, Electronics & Vision (ICIEV). IEEE, 2015. p. 1-6.

JALAL, Ahmad; KAMAL, Shaharyar; AZURDIA-MEZA, Cesar A. Depth maps-based human segmentation and action recognition using full-body plus body color cues via recognizer engine. *Journal of Electrical Engineering & Technology*, v. 14, n. 1, p. 455-461, 2019.

JALAL, Ahmad; KAMAL, Shaharyar; KIM, Daijin. Depth Silhouettes Context: A new robust feature for human tracking and activity recognition based on embedded HMMs. In: 2015 12th International Conference on Ubiquitous Robots and Ambient Intelligence (URAI). IEEE, 2015. p. 294-299.

JALAL, Ahmad; KAMAL, Shaharyar; KIM, Daijin. Shape and motion features approach for activity tracking and recognition from kinect video camera. In: 2015 IEEE 29th International Conference on Advanced Information Networking and Applications Workshops. IEEE, 2015. p. 445-450.

JALAL, Ahmad; KAMAL, Shaharyar; KIM, Dong-Seong. Detecting complex 3D human motions with body model low-rank representation for real-time smart activity monitoring system. *KSII Transactions on Internet and Information Systems*, v. 12, n. 3, p. 1189-1204, 2018.

JALAL, Ahmad; KIM, Jeong Tai; KIM, Tae-Seong. Human activity recognition using the labeled depth body parts information of depth silhouettes. In: *Proceedings of the 6th International Symposium on Sustainable Healthy Buildings*, Seoul, Korea. 2012.

JIN, Chanwoo et al. Similarity measurement on human mobility data with spatially weighted structural similarity index (SpSSIM). *Transactions in GIS*, v. 24, n. 1, p. 104-122, 2020.

JUNIOR, Carlos Fernando Crispim et al. ETHOWATCHER: validation of a tool for behavioral and video-tracking analysis in laboratory animals. *Computers in biology and medicine*, v. 42, n. 2, p. 257-264, 2012.

KABBAI, Leila; ABDELLAOUI, Mehrez; DOUIK, Ali. Image classification by combining local and global features. *The Visual Computer*, v. 35, n. 5, p. 679-693, 2019.

KAMAL, Shaharyar; JALAL, Ahmad. A hybrid feature extraction approach for human detection, tracking and activity recognition using depth sensors. *Arabian Journal for science and engineering*, v. 41, n. 3, p. 1043-1051, 2016.

LAKSHMI, Bagavathi; PARTHASARATHY, S. Human action recognition using median background and max pool convolution with nearest neighbor. *International Journal of Ambient Computing and Intelligence (IJACI)*, v. 10, n. 2, p. 34-47, 2019.

LAN, Zhenzhong et al. Deep local video feature for action recognition. In: *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*. 2017. p. 1-7.

LIFECAM HD, LifeCam HD - 3000, Microsoft. Disponível em: <<https://www.microsoft.com/pt-br/accessories/products/webcams/lifecam-hd-3000?activetab=overview%3aprimaryr2>>. Acesso em: 23 de novembro de 2020.

LOPEZ, Natalia M. et al. From hospital to home care: Creating a domotic environment for elderly and disabled people. *IEEE pulse*, v. 7, n. 3, p. 38-41, 2016.

LUBITZ, James et al. Health, life expectancy, and health care spending among the elderly. *New England Journal of Medicine*, v. 349, n. 11, p. 1048-1055, 2003.

MARCOLAN, João A.; COLUSSI, Alex; Marino-Neto, José. Algoritmo de camera virtual móvel para a redução do tempo de processamento de vídeos em diferentes resoluções. In: *Anais do V Congresso Brasileiro de Eletromiografia e Cinesiologia e X Simpósio de Engenharia Biomédica*. Anais. Uberlândia (MG) Center Convention Uberlândia, 2018. Disponível em: <<https://www.even3.com.br/anais/cobecseb/78861-ALGORITMO-DE-CAMERA-VIRTUAL-MOVEL-PARA-A-REDUCAO-DO-TEMPO-DE-PROCESSAMENTO-DE-VIDEOS-EM-DIFERENTES-RESOLUCOES>>. Acesso em: 11/06/2019 10:05

MOVIEPY, MoviePy: repositório do MoviePy no Github. Disponível em: <<https://github.com/Zulko/moviepy>>. Acesso em: 14 de novembro de 2020.

NASR, M. Abdel-Salam; ALRAHMAWY, Mohammed F.; TOLBA, A. S. Multi-scale structural similarity index for motion detection. *Journal of King Saud University-Computer and Information Sciences*, v. 29, n. 3, p. 399-409, 2017.

NIMMAKAYALA, Satish et al. Modern Health Monitoring System Using IoT. In: ICCCE 2020. Springer, Singapore. p. 1135-1144.

NOLDUS, Noldus TrackLab, Spatial Behavior Tracking. Disponível em: <<https://www.noldus.com/tracklab-human>>. Acesso em: 23 de novembro de 2020.

OPEN CV, Open Source Computer Vision Library. Disponível em: <<https://opencv.org/>>. Acesso em: 02 de novembro de 2020 . OpenCV. (2015).

ORANGE, data Mining Fruitful and Fun. Disponível em: < <https://orange.biolab.si/>>. Acesso em: 09 de novembro de 2020.

PANINI, L.; CUCCHIARA, Rita. A machine learning approach for human posture detection in domotics applications. In: 12th International Conference on Image Analysis and Processing, 2003. Proceedings. IEEE. Mantova, Italy, 2003. p. 103-108.

PAVITHRA, S.; RAMESH, R.; SUNDRAMBAL, B. Survey on Face Recognition with pre filtering Techniques and their comparative study. In: 2020 6th International Conference on Advanced Computing and Communication Systems (ICACCS). IEEE, 2020. p. 1089-1096.

PICCARDI, Massimo. Background subtraction techniques: a review. In: 2004 IEEE International Conference on Systems, Man and Cybernetics (IEEE Cat. No. 04CH37583). IEEE, 2004. p. 3099-3104.

PONCE, Sergio et al. Wearable sensors and domotic environment for elderly people. In: World Congress on Medical Physics and Biomedical Engineering 2018. Springer, Singapore, 2019. p. 195-200.

PRATI, Andrea; SHAN, Caifeng; WANG, Kevin I.-Kai. Sensors, vision and networks: From video surveillance to activity recognition and health monitoring. *Journal of Ambient Intelligence and Smart Environments*, v. 11, n. 1, p. 5-22, 2019.

RASPBERRYPI, RaspberryPi Org. Products Raspberry Pi 3 Model B. Disponível em: <<https://www.raspberrypi.org/products/raspberry-pi-3-model-b/>>. Acesso em: 3 de outubro de 2020.

SARWAR, Omair; RINNER, Bernhard; CAVALLARO, Andrea. A privacy-preserving filter for oblique face images based on adaptive hopping Gaussian mixtures. *IEEE Access*, v. 7, p. 142623-142639, 2019.

SENNSLAB, SeensLab, Bitbrain. Disponível em: <<https://www.bitbrain.com/neurotechnology-products/software/sennslab>>. Acesso em: 23 de novembro de 2020.

SHARMA, Lavanya. Human Detection and Tracking Using Background Subtraction in Visual Surveillance. Towards Smart World: Homes to Cities Using Internet of Things, p. 317, 2020.

SIRIDAKIS, E. J. Ferramenta de aquisição e exploração de dados multissensoriais para monitoramento de atividades humanas em ambiente doméstico. 2020. Dissertação (Mestrado em engenharia elétrica e eletrônica) - Instituto de Engenharia Biomédica, Universidade Federal de Santa Catarina, Florianópolis.

SOHN, Misook; HAHN, Dongwon; LEE, Jeonwoo. The strategy development of u-health service. In: Technology Management for the Global Future, 2006. PICMET 2006. IEEE, 2006. p. 1931-1934.

TAPIA, Emmanuel Munguia et al. Real-time recognition of physical activities and their intensities using wireless accelerometers and a heart rate monitor. In: 2007 11th IEEE international symposium on wearable computers. IEEE, 2007. p. 37-40.

TEAGUE, Michael Reed. Image analysis via the general theory of moments. JOSA, v. 70, n. 8, p. 920-930, 1980.

TINKERCAD, TinkerCad Autodesk Inc. Projetos 3D. Disponível em: <<https://www.tinkercad.com/>>. Acesso em: Julho de 2018.

VIOLA, Paul; JONES, Michael. 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. IEEE, 2001. p. I-I.

WANG, Cunrui et al. Facial feature discovery for ethnicity recognition. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, v. 9, n. 1, p. e1278, 2019.

WANG, Guiming. Machine learning for inferring animal behavior from location and movement data. Ecological Informatics, v. 49, p. 69-76, 2019.

WU, Winston H. et al. MEDIC: Medical embedded device for individualized care. Artificial intelligence in medicine, v. 42, n. 2, p. 137-152, 2008.

XU, Wenyao et al. Co-recognition of human activity and sensor location via compressed sensing in wearable body sensor networks. In: 2012 Ninth International Conference on Wearable and Implantable Body Sensor Networks. IEEE, 2012. p. 124-129.

YOUNGBLOOD, G. Michael et al. Automation intelligence for the smart environment. In: International Joint Conference On Artificial Intelligence. LAWRENCE ERLBAUM ASSOCIATES LTD, 2005. p. 1513.

ZHAO, Shizhen; LI, Wenfeng; CAO, Jingjing. A user-adaptive algorithm for activity recognition based on k-means clustering, local outlier factor, and multivariate gaussian distribution. *Sensors*, v. 18, n. 6, p. 1850, 2018.

ZOUBA, Nadia; BREMOND, Francois; THONNAT, Monique. An activity monitoring system for real elderly at home: Validation study. In: 2010 7th IEEE International Conference on Advanced Video and Signal Based Surveillance. IEEE, 2010. p. 278-285.

ANEXO A – Questionário utilizado para a coleta de dados

ETAPA -1

O seguinte questionário constitui a **Etapa 1** da pesquisa intitulada: **“Plataforma de saúde ubíqua voltada à assistência domiciliar de idosos: detecção multissensorial de eventos em ambientes domésticos”** e deve ser preenchido somente pelos participantes que já assinaram o **“TCLE - Termo de Consentimento Livre e Esclarecido”**.

•Dados cadastrais

1.Sexo: () Masculino() Feminino

2.Idade: _____

3.Contato: () _____

•Dados clínicos

1.Você faz uso de algum medicamento de forma contínua?

() Não() SimSe sim, qual? _____

2.Você tomou algum medicamento nas últimas 24 horas que possa prejudicar a sua coordenação motora?

() Não() SimSe sim, qual? _____

3.Você tomou algum medicamento nas últimas 24 horas que possa prejudicar a sua concentração ou memória?

() Não() SimSe sim, qual? _____

4.Você sofreu alguma lesão recentemente que possa prejudicar a sua locomoção ou movimento corporal?

() Não() SimSe sim, em qual parte do corpo? _____

5.Você possui alguma doença que prejudique a sua locomoção ou movimento corporal?

() Não() SimSe sim, qual? _____

6.Você possui alguma doença que prejudique a sua concentração ou memória?

() Não() SimSe sim, qual? _____

Campo de preenchimento exclusivo do laboratório!

PARTICIPANTE Nº _____ ID: _____ DATA: ____ / ____ / ____

ANEXO B – Checklist utilizado para a coleta de dados

Data: __/__/__ Hora: __:__

ID voluntário: _____

Responsável: _____

Obs: Chegar no mínimo 1 hora antes da coleta de dados:

- 1 – Confirmar com o voluntário o horário e a presença;
- 2 – Fazer as anotações no caderno de experimento (Data, hora);
- 3 – Criar código para o voluntário (ex: ID 12345_09dez2019)

Checklist estrutural

- Interditar banheiro (Anexar aviso nas portas)
- Anexar aviso na porta de principal e trancar
- Verificar Tomadas
- Verificar interruptores
- Verificar Lâmpadas
- Verificar Janelas (fechar)
- Verificar *Blackout* (Vedação total da luz solar)
- Verificar chão (seco)
- Verificar Geladeira (Potes de água e bolinhas)
- Verificar Fogareiro
- Verificar caixinha de fósforo
- Verificar panela
- Verificar prato
- Verificar talheres
- Verificar copo
- Verificar pote para descarte das bolinhas
- Verificar mesa
- Verificar cadeira
- Verificar torneira (água)
- Verificar detergente
- Verificar esponja
- Verificar escorredor de louças

Checklist de equipamentos

- Verificar Câmera
- Verificar placa de sensores (teensy)
- Verificar sensor umidade
- Verificar sensor de temperatura
- Verificar microfone
- Verificar sensor de luminosidade
- Verificar extensor de USB
- Verificar Computador
- Verificar HD externo
- Verificar criação de pastas de armazenamento de amostras
- Verificar código do voluntário
- Verificar Software
- Coletar dados de teste de funcionamento
- Inserir ID do voluntário no software

Após a coleta de dados:

- Verificar os vídeos coletados (Assistir começo, meio e fim)
- Verificar os áudios coletados (Ouvir começo, meio e fim)
- Verificar as tabelas .CSV (Continuidade de dados\linha)
- Verificar se os dados foram salvos
- Distribuir os dados em 3 computadores
- Bloquear as amostras com senha

ANEXO B – Termo de Consentimento Livre e Esclarecido
TERMO DE CONSENTIMENTO LIVRE E ESCLARECIDO - TCLE

Caro voluntário(a), você está sendo convidado(a) a participar de uma pesquisa intitulada **“Plataforma de saúde ubíqua voltada à assistência domiciliar de idosos: detecção multissensorial de eventos em ambientes domésticos”**, cujo objetivo é o desenvolvimento e implementação de uma ferramenta de reconhecimento automático de comportamentos, baseada em imagem, som e sensores (temperatura, umidade e luminosidade). Esta pesquisa está associada aos projetos de Dissertação de Mestrado dos alunos Eduardo Jorge Siridakis, Lennon José Gazola dos Santos e Suzana Carnielli do Prado, mencionados neste documento como “pesquisadores assistentes”, do programa de Pós-Graduação em Engenharia Elétrica na subárea de Engenharia Biomédica da Universidade Federal de Santa Catarina (UFSC), sob orientação da Prof^a. Dr. Daniela Ota Hisayasu Suzuki (UFSC), mencionada neste documento como “Professora responsável” e sob coordenação do Prof. Dr. José Marino Neto (UFSC), mencionado neste documento como “Professor corresponsável”.

Este documento que você está lendo é chamado de Termo de Consentimento Livre e Esclarecido (TCLE). O TCLE contém explicações sobre o estudo ao qual você está sendo convidado(a) a participar. Antes de você aceitar participar por livre e espontânea vontade, você deverá ler e compreender todo o seu conteúdo. Os professores responsáveis e os pesquisadores assistentes acreditam que esta pesquisa seja importante para o desenvolvimento de uma plataforma de saúde ubíqua que possa auxiliar no monitoramento de idosos e portadores de necessidades especiais em ambientes domésticos complexos.

Procedimentos e participação na pesquisa

Os procedimentos éticos desta pesquisa se baseiam no que aponta a resolução nº 466/2012 do Conselho Nacional de Saúde (CNS), que regula os cuidados éticos em pesquisas envolvendo seres humanos. Tal resolução incorpora a consideração ao respeito, dignidade, liberdade e autonomia do ser humano, assegurando os direitos e deveres que dizem respeito aos participantes da pesquisa, à comunidade científica e ao Estado.

A participação no estudo é voluntária, o que significa que você não é obrigado a aceitá-la e pode se sentir absolutamente à vontade em deixar de participar a qualquer momento, sem ter que apresentar qualquer justificativa e sem ter qualquer prejuízo. Caso você decida participar, será então denominado neste documento como: “o(a) participante”.

A sua participação na pesquisa terá duração de aproximadamente quarenta minutos, durante os quais, como participante, você será convidado a:

Etapa 1 – Responder a um questionário solicitando: dados pessoais como idade, sexo e informações de contato; dados clínicos referentes à coordenação motora e movimentos corporais.

Etapa 2 - Em seguida o(a) participante será convidado a realizar um conjunto de ações cotidianas, definidas por um roteiro (veja abaixo) em um laboratório do IEB-UFSC, que simula um ambiente doméstico (cozinha), estruturado com eletrodomésticos e utensílios (copos, talheres, pratos, geladeira, fogareiro, cafeteira). Para registrar os eventos e coletar os dados necessários, o local foi equipado com uma câmera, um microfone, e sensores de temperatura, de umidade e de luminosidade. O roteiro de atividades do participante é composto por cinco sequências de ações:

1 - O(A) participante entra no ambiente pela porta e se dirige para a geladeira; abre a geladeira e retira dois objetos (uma garrafa plástica contendo água e um pote de plástico contendo esferas plásticas); deposita os dois objetos sobre a mesa e fecha a porta da geladeira.

2 – O(a) participante recolhe os dois objetos da mesa e se dirige até o balcão (que contém um fogareiro com uma panela (em cima), uma caixa de fósforos sem a lixa de ignição e um pote com diversas colheres); deposita água e esferas plásticas dentro da panela; simula a ignição de um fósforo com a caixa e o direciona ao fogareiro simulando o ato de acender o fogão. Em seguida, escolhe uma colher, pega e realiza movimentos circulares (10 a 15 movimentos) dentro da panela, movimentando as esferas na água; em seguida, simula o “desligamento” do fogão.

3 – O(a) participante se dirige ao balcão da pia, escolhe um prato e o leva até a mesa; se dirige ao fogareiro, recolhe a panela e deposita a água e as esferas dentro do prato que está sobre a mesa; deixa a panela sobre a mesa e coloca a colher dentro do prato.

4 – O(a) participante se senta na cadeira e utiliza a colher para simular o ato de “comer”, movimentando a colher, do prato em direção próxima da boca, por cerca de 10 vezes.

5 – O(a) participante se levanta, recolhe a panela, o prato e a colher e os deposita dentro da pia (no balcão da pia contém: um pote plástico de cor branca, uma embalagem de detergente (com água no lugar de detergente), uma esponja e escorredor de louças); deposita a água e as esferas dentro do pote plástico de cor branca; simula o ato de “lavar louça”: abre a torneira, coloca o “detergente” na esponja, esfrega o utensílio e enxagua em água corrente; deposita os utensílios lavados no escorredor de louças; seca as mãos com um pano seco ao lado da pia; se retira do cenário pela porta. Fim do roteiro de atividades.

Esta sequência de ações será explicada pelos pesquisadores assistentes e ensaiada com o(a) participante antes da coleta real de amostras. Os sistemas de registro de dados serão acionados um minuto antes de iniciar a sequência, e a coleta de dados será interrompida um minuto após a saída do participante do cenário. Cada participante repetirá a sequência três vezes, com um intervalo de três minutos entre cada repetição. As atividades serão gravadas em formato de áudio, vídeo e tabela de dados.

Benefícios, riscos e possíveis desconfortos

É importante ressaltar que o(a) participante não deve esperar nenhum benefício pessoal proveniente desta pesquisa e que toda pesquisa tem um risco associado.

Risco de quebra de sigilo: A primeira parte do protocolo do presente estudo envolve o preenchimento do questionário (dados cadastrais e clínicos referentes a coordenação motora). A segunda parte do protocolo consiste na coleta de dados no formato de áudio (.wav), vídeo (.mp4) e tabela de dados (.CSV e .XML) pelo sistema de registro de dados (câmera, microfone e sensores). Os pesquisadores assistentes e os professores responsáveis serão os únicos a ter acesso aos dados conforme coletados e tomarão os devidos cuidados para manter o sigilo da identidade do(a) participante, identificando o(a) participante apenas por meio de códigos (cada participante terá um número de referência). Os dados coletados (áudio, vídeo e tabelas) serão armazenados em uma unidade de disco rígido utilizado unicamente para este fim, e cujo acesso será permitido apenas aos pesquisadores assistentes e professores responsáveis, minimizando os riscos.

Todas as imagens gravadas em formato de vídeo passarão por um processo de proteção de privacidade, no qual será aplicado um borrão no rosto do(a) participante, de forma que não seja possível realizar o reconhecimento facial. Os vídeos originais serão descartados, e os vídeos processados para impedir reconhecimento facial serão identificados apenas por números. O arquivo de áudio coletado conterá apenas os sons emitidos durante a realização das atividades propostas no roteiro, não sendo necessário gravar diálogos ou qualquer tipo de som da voz do(a) voluntário(a), que é mais uma garantia de proteção de privacidade. No entanto, sempre existe a possibilidade de quebra de sigilo, mesmo que seja remota e não intencional, cujas consequências são tratadas nos termos da lei. Os resultados dessa pesquisa poderão ser apresentados em trabalhos científicos que mostrarão os resultados obtidos sem revelar, contudo, quaisquer informações relacionadas à identidade do(a) participante. Se algum dado do(a) participante em específico for mostrado, tal dado será identificado a partir de um código numérico. Se alguma imagem das gravações for utilizada em trabalhos científicos, a proteção de privacidade será garantida ocultando o rosto do participante. Estimamos esta atividade como de baixo risco ao participante.

Riscos durante o roteiro de atividades: Na realização das atividades definidas pelo roteiro, existe um risco de alergia associado ao contato com os utensílios utilizados (plástico, metal, madeira, tecido e água). Se o(a) participante vier a apresentar um possível quadro de alergia, profissionais da saúde como do Serviço de Atendimento Móvel de Urgência (SAMU) serão acionados pelos pesquisadores assistentes. Os eletrodomésticos utilizados no cenário estarão desligados e desativados da rede elétrica, evitando qualquer risco de choque elétrico; o fogareiro não terá gás conectado; a caixa de fósforos estará com a lixa de ignição selada com fita adesiva, e os fósforos serão substituídos por palitos de madeira, sem ponta perfurante ou combustível; a embalagem de detergente será desinfetada e preenchida com água para minimizar o risco de alergia ao detergente; os utensílios como potes e

pratos serão de plástico, dificultando a quebra em caso de queda; o chão do laboratório não terá tapetes e a cada intervalo de teste o chão será seco para evitar o risco de queda por chão liso/molhado. Se o(a) participante vier a sofrer uma queda, corte ou qualquer lesão, as atividades serão interrompidas e os profissionais da saúde (SAMU) serão acionados pelos pesquisadores assistentes. O chão possui piso áspero, dificultando acidentes por escorregamento; a pia possui uma borda elevada nas laterais evitando que caia água no chão; não há utensílios de louça, facas, ou objetos cortantes; assim, estimamos esta atividade como de baixo risco ao participante.

Outros riscos: Existe o risco de alterações na autoestima do(a) participante provocada pelo cansaço, desconforto, constrangimento ou alterações de comportamento durante as gravações de áudio e vídeo; existe o risco de frustração do(a) participante por não conseguir lembrar das atividades descritas no roteiro de testes durante a execução. Para minimizar estes riscos, durante o protocolo acima descrito estarão presentes, além do(a) participante, apenas os pesquisadores assistentes, que poderão interromper as atividades, rememorar o roteiro e auxiliar o(a) participante em qualquer dúvida, inclusive sugerindo uma pausa caso necessário.

Caso ocorram danos causados pela participação na pesquisa devido aos riscos previstos acima ou não previstos, o(a) participante terá direito a acompanhamento, tratamento médico, indenização e reparo nos termos da lei.

Ressarcimento e indenização

A legislação brasileira vigente não permite que o(a) participante tenha qualquer compensação financeira pela sua participação na presente pesquisa, mas o(a) participante será integralmente ressarcido(a) pelas despesas com transporte e alimentação previstas no orçamento do projeto; o(a) participante não terá nenhuma despesa advinda da sua participação no projeto. Caso alguma despesa extraordinária associada à pesquisa venha a ocorrer, o(a) participante será ressarcido(a) nos termos da lei. Caso o(a) participante tenha algum prejuízo material ou imaterial em decorrência da pesquisa poderá solicitar indenização de acordo com a legislação vigente e amplamente consubstanciada

Liberdade de participação

Antes de concordar em participar da pesquisa e assinar, você deverá realizar todas as perguntas que desejar aos pesquisadores assistentes, sobre o que não tiver ficado claro para você. A sua participação é voluntária, portanto, você tem livre acesso e deve se sentir à vontade em deixar de participar da pesquisa a qualquer momento. A sua eventual desistência nesta pesquisa não trará prejuízo algum ou penalidade alguma a você. Caso você decida desistir da sua participação, basta entrar em contato com algum dos pesquisadores assistentes ou com o professor responsável – cujas formas de contato serão apresentadas a seguir - e informar que gostaria de retirar o consentimento. Os

pesquisadores assistentes estarão sempre acompanhando, antes, durante e depois da coleta de dados, para prestar toda a assistência necessária.

Contato

Os pesquisadores assistentes responderão suas dúvidas a qualquer momento, seja ela antes, durante ou após a coleta dos dados. Caso tenha alguma dúvida sobre os procedimentos ou sobre o projeto, você poderá entrar em contato com os pesquisadores assistentes: Eduardo Jorge Siridakis, Lennon José Gazola dos Santos e Suzana Carnielli do Prado, ou com o Professora responsável e o Professor corresponsável, respectivamente Prof^a. Dr. Daniela Ota Hisayasu Suzuki e Prof. Dr. José Marino Neto, ambos vinculados ao Instituto de Engenharia Biomédica da Universidade Federal de Santa Catarina (IEB - UFSC), cujos contatos são: Prof. Dr. Daniela Ota Hisayasu Suzuki: telefone: (48) 3721-8686, e-mail suzuki@eel.ufsc.br; Prof Dr. José Marino Neto: telefone (48) 3721-8686, e-mail j.marino@ufsc.br; Eduardo Jorge Siridakis: telefone (48) 3721-8686 , e-mail eduardosiridakis@gmail.com, Lennon José Gazola dos Santos: telefone (44) 99990-3482, e-mail lennongazola@gmail.com; Suzana Carnielli do Prado: telefone (48) 99159-5876, e-mail suzanacarnielli@hotmail.com. O endereço físico profissional dos pesquisadores assistentes e dos professores responsáveis é: Instituto de Engenharia Biomédica (IEB), Universidade Federal de Santa Catarina, Cep 88040-900, Fundos do Hospital Universitário da UFSC, Florianópolis – SC. Horário de funcionamento de segunda a sexta-feira – das 8h às 12h e das 13h30min às 17h30min. Telefone para contato: (48) 3721-8686.

Todos os protocolos aqui descritos serão realizados no Laboratório de Bioengenharia, instalado no 5º andar, sala 56, do Instituto de Engenharia Biomédica (IEB), Universidade Federal de Santa Catarina, Cep 88040-900, Fundos do Hospital Universitário da UFSC, Florianópolis – SC. Horário de funcionamento de segunda a sexta-feira – das 8h às 12h e das 13h30min às 17h30min. Email: ceged.tmh@ieb.ufsc.br. Horário de funcionamento de segunda a sexta-feira – das 8h00min às 12h00min e das 13h30min às 17h30min. Telefone para contato: (48) 3721-8686.

Comitê de Ética em Pesquisa com Seres Humanos - CEPESH

Dúvidas sobre a pesquisa envolvendo os princípios éticos poderão ser direcionadas ao Comitê de Ética em Pesquisa com Seres Humanos da UFSC (CEPSH-UFSC). O CEPESH é um órgão colegiado interdisciplinar, deliberativo, consultivo e educativo, vinculado à Universidade Federal de Santa Catarina, mas independente na tomada de decisões, criado para defender os interesses dos participantes da pesquisa em sua integridade e dignidade e para contribuir no desenvolvimento da pesquisa dentro de padrões éticos com atendimento no Prédio Reitoria II, 4º andar, sala 401, localizado na Rua Desembargador Vitor Lima, nº 222, Trindade, Florianópolis. – SC. Telefone para contato (48) 37216094, e-mail: cep.propesq@contato.ufsc.br. Horário de funcionamento de segunda a sexta-feira – 07h00min às 19h00min.

DECLARAÇÃO

Este documento foi elaborado em duas vias, datadas e com páginas numeradas. Uma das vias será destinada ao(a) participante, outra aos professores responsáveis e pesquisadores assistentes. O participante, os professores responsáveis e o pesquisador assistente (presente no dia dos testes) devem assinar ambas as vias. Todas as páginas deverão ser rubricadas pelo(a) participante, pelo Professor responsável e pelo pesquisador assistente. O(a) participante deverá guardar cuidadosamente a sua via, pois é este documento que traz informações importantes de contato e garante seus direitos como participante da pesquisa.

Os Professores responsáveis e pesquisadores assistentes, que também assinam este documento, comprometem-se a conduzir a pesquisa de acordo com o que preconiza a resolução 466/12 de 12/06/2012, que trata dos preceitos éticos e da proteção aos participantes da pesquisa.

Eu, Sr(a) _____, portador(a) do documento de identidade nº _____, considero-me informado(a) sobre a pesquisa “Plataforma de saúde ubíqua voltada à assistência domiciliar de idosos: detecção multissensorial de eventos em ambientes domésticos” e obtive dos pesquisadores todas as informações que julguei necessárias para me sentir esclarecido(a) e optar por livre espontânea vontade participar da pesquisa. Aceito participar da mesma, sendo denominado de “o(a) participante” nos termos anteriormente descritos, consentindo que o questionário e a coleta dos dados via áudio, vídeo e tabela de dados sejam realizados. Expresso minha concordância com a divulgação pública dos resultados uma vez que recebi garantias sobre o sigilo de minha identidade, imagem e o anonimato das informações prestadas.

Florianópolis, ____ de _____ de _____

Assinatura do(a) participante

Professora responsável
Prof.ª. Dr Daniela Ota Hisayasu Suzuki

Assinatura
CPF: _____
RG _____

Professor corresponsável
Prof. Dr. José Marino Neto

Assinatura
CPF: _____
RG _____

Pesquisador assistente

Nome: _____

Assinatura
CPF: _____
RG _____

ANEXO C – Parecer Consubstanciado do CEP

UNIVERSIDADE FEDERAL DE
SANTA CATARINA - UFSC

**PARECER CONSUBSTANCIADO DO CEP****DADOS DO PROJETO DE PESQUISA**

Título da Pesquisa: PLATAFORMA DE SAÚDE UBÍQUA VOLTADA À ASSISTÊNCIA DOMICILIAR DE IDOSOS: DETECÇÃO MULTI-SENSORIAL DE EVENTOS EM AMBIENTES DOMÉSTICOS

Pesquisador: Daniela Ota Hisayasu Suzuki

Área Temática:

Versão: 1

CAAE: 24153319.0.0000.0121

Instituição Proponente: UNIVERSIDADE FEDERAL DE SANTA CATARINA

Patrocinador Principal: Financiamento Próprio

DADOS DO PARECER

Número do Parecer: 3.700.866

Apresentação do Projeto:

Projeto de pesquisa realizado pelo laboratório de Bioengenharia do Instituto de Engenharia Biomédica (IEB), no Programa de PósGraduação em Engenharia Elétrica da Universidade Federal de Santa Catarina, tendo como coordenadores Daniela Ota Hisayasu Suzuki José Marino Neto.

O estudo será desenvolvido em duas etapas de coleta de dados. A primeira etapa consistirá em um estudo- piloto realizado em 6-8 voluntários (mulheres e homens adultos, entre 26-60 anos de idade, que incluem essencialmente colaboradores do projeto e colegas do IEB-UFSC). Esta etapa visa testar os protocolos de coleta e processamento de amostras e construir um banco de dados preliminar contendo as variáveis e atributos primários extraídos dos registros de som e imagem. Este banco de dados permitirá o teste/validação em situação real do sistema de sensores em desenvolvimento, bem como a prospecção dos procedimentos mais adequados para a obtenção e preparação de dados e de seu uso em um conjunto de algoritmos de aprendizado de máquina. Este estudo deverá guiar o detalhamento dos procedimentos de análise.

Neste piloto, cada participante deve repetir o roteiro de testes (uma sequência de diferentes atos

comuns em uma cozinha, e.g., abrir a porta da geladeira, abrir a torneira da pia) por um mínimo de 5 e um máximo de 7 vezes, durante os quais os registros em vídeo, em áudio e a partir dos demais sensores serão realizados. As amostras de cada atividade, coletadas de cada indivíduo serão processadas com o intuito de encontrar características que revelem um padrão particular desse participante. Assim, pretendemos verificar a possibilidade de construir, a partir destes dados, um modelo para detecção das atividades daquele indivíduo em particular, e não um modelo genérico para o conjunto dos participantes. Estes modelos individuais dos padrões de atividade poderiam favorecer a detecção de mudanças mais sutis nos eventos comportamentais exibidos por uma pessoa atuando no ambiente.

As amostras serão analisadas e submetidas a extração de diversos parâmetros, principalmente descritores relacionados à áudio (e.g., potências em bandas determinadas de frequência) e vídeo (mudanças de forma, localização, velocidade), que posteriormente serão utilizados para alimentar uma rede neural artificial. Os resultados do registro piloto devem sugerir soluções para definir o número aproximado de repetições dos testes que o grupo alvo da investigação, os idosos (6-8, homens e mulheres em igual número, hígidos e aptos a locomoção não assistida, entre 60 e 75 anos de idade.) Será solicitado a realizar, para que as amostras sejam suficientes para extração de atributos importantes e para formar os grupos de teste, treinamento e validação da rede neural.

Estes participantes serão convidados entre o conjunto de indivíduos idosos que frequentam o Núcleo de Estudos da Terceira Idade (NETI) da Universidade Federal de Santa Catarina (UFSC). No planejamento amostral do registro com idosos, o número de participantes será definido de acordo com os resultados do registro piloto. Novamente, pretendemos realizar a modelagem de cada um dos comportamentos de maneira individual, para revelar um padrão particular desse idoso. É muito importante notar que o modelo e a implementação deste em um detector automático de eventos não devem ter características generalistas, mas ser adequado ao usuário específico da ferramenta.

Cada participante repetirá a sequência três vezes, com um intervalo de três minutos entre cada repetição. As atividades serão gravadas em formato de áudio, vídeo e tabela de dados.

Critério de Inclusão:

Para o grupo de registro piloto serão incluídos de 6 a 8 participantes, mulheres e homens adultos, entre 26 e 60 anos de idade, que se sintam dispostos e capazes de realizar as atividades propostas e que sejam capazes de locomoção não assistida. Para o grupo de registro da segunda etapa, serão incluídos de 6 a 8 participantes, mulheres e homens, entre 60 e 75 anos de idade, que se sintam dispostos e capazes de realizar as atividades propostas e que sejam capazes de locomoção não assistida.

Critério de Exclusão:

- Indivíduos com dificuldades de deambulação ou restrições de movimentos que impeçam ou prejudiquem o desempenho das atividades propostas no roteiro; - Indivíduos que tenham consumido algum medicamento nas últimas 24 horas que impeça ou prejudique o desempenho das atividades propostas no roteiro; - Indivíduos com cortes, feridas ou queimaduras que impeçam ou prejudiquem o desempenho das atividades propostas no roteiro;

Objetivo da Pesquisa:

Objetivo Primário:

Prospecção, desenvolvimento e implementação de ferramentas de reconhecimento automático de comportamento baseadas em imagem/som e em RNAs, destinadas ao monitoramento de idosos e de portadores de necessidades especiais em ambientes domésticos complexos. Estes instrumentos serão, por sua vez, baseados em plataformas de desenvolvimento de código aberto, e implementados como parte de uma rede de sensores vinculada a uma plataforma de saúde ubíqua.

Objetivo Secundário:

- Criação e testes de uma ferramenta (hardware e software) para detecção de padrões de atividades domésticas através do processamento de imagens: registro e extração de atributos cinemáticos e morfológicos (registros em vídeo) e sonoros (registros em áudio).- Desenvolvimento e testes de algoritmos de aprendizado profundo de máquina para detecção de atividades domésticas.

Avaliação dos Riscos e Benefícios:

Primorosamente detalhada e adequada, tanto no projeto quanto no TCLE.

Comentários e Considerações sobre a Pesquisa:

Sem comentários adicionais.

Considerações sobre os Termos de apresentação obrigatória:

Folha de rosto assinada pela pesquisadora responsável e pelo coordenador do PPG em Engenharia Elétrica da UFSC.

Declaração de anuência e infraestrutura assinada pelo coordenador do Instituto de Engenharia Biomédica da UFSC.

Coleta de dados da etapa piloto prevista para novembro de 2019, com a adição, por parte dos proponentes, de que "o cronograma de execução para a coleta de dados depende inteiramente da data de aprovação do comitê de ética. Caso o processo atrase, as datas serão adiadas."

O orçamento informa despesas de R\$ 4130,00 com financiamento próprio.

Questionário para a etapa 1 da pesquisa, com 3 perguntas sobre dados cadastrais (sexo, idade, contato) e 6 perguntas associadas medicação, coordenação motora, concentração, locomoção e memória.

O TCLE está muito bem redigido, esclarecedor a respeito de objetivos, procedimentos, riscos e direitos, atendendo essencialmente a todas as exigências da res. 466/12.

Conclusões ou Pendências e Lista de Inadequações:

Pela aprovação.

Considerações Finais a critério do CEP:

Este parecer foi elaborado baseado nos documentos abaixo relacionados:

Tipo Documento	Arquivo	Postagem	Autor	Situação
Informações Básicas do Projeto	PB_INFORMACOES_BASICAS_DO_PROJETO_1431908.pdf	17/10/2019 15:16:02		Aceito
Outros	Questionario_Etapa1_coleta_de_dados.pdf	17/10/2019 15:12:45	LENNON JOSE GAZOLA DOS SANTOS	Aceito
Declaração de Instituição e Infraestrutura	Declaracao_instituicao.pdf	17/10/2019 15:11:17	LENNON JOSE GAZOLA DOS SANTOS	Aceito
Projeto Detalhado / Brochura Investigador	Projeto_detalhado.pdf	17/10/2019 15:10:33	LENNON JOSE GAZOLA DOS SANTOS	Aceito
TCLE / Termos de Assentimento / Justificativa de Ausência	TCLE_Termo_de_consentimento.pdf	17/10/2019 15:09:53	LENNON JOSE GAZOLA DOS SANTOS	Aceito
Folha de Rosto	Folha_de_rosto.pdf	17/10/2019 15:08:24	LENNON JOSE GAZOLA DOS SANTOS	Aceito

Situação do Parecer:

Aprovado

Necessita Apreciação da CONEP:

Não

FLORIANOPOLIS, 12 de Novembro de 2019

**Assinado por:
Maria Luiza Bazzo
(Coordenador(a))**

Endereço: Universidade Federal de Santa Catarina, Prédio Reitoria II, R: Desembargador Vitor Lima, nº 222, sala 401	
Bairro: Trindade	CEP: 88.040-400
UF: SC	Município: FLORIANOPOLIS
Telefone: (48)3721-6094	E-mail: cep.propesq@contato.ufsc.br