

GUILHERME HOLSBACH COSTA

**DETECÇÃO FACIAL:
AUTOFACES VERSUS ANTIFACES**

**FLORIANÓPOLIS
2003**

UNIVERSIDADE FEDERAL DE SANTA CATARINA

**PROGRAMA DE PÓS-GRADUAÇÃO
EM ENGENHARIA ELÉTRICA**

**DETECÇÃO FACIAL:
AUTOFACES VERSUS ANTIFACES**

Dissertação submetida à
Universidade Federal de Santa Catarina
como parte dos requisitos para a
obtenção do grau de mestre em Engenharia Elétrica.

GUILHERME HOLSBACH COSTA

Florianópolis, maio de 2003.

DETECÇÃO FACIAL: AUTOFACES VERSUS ANTIFACES

Guilherme Holsbach Costa

‘Esta dissertação foi julgada adequada para a obtenção do Título de Mestre em Engenharia Elétrica, Área de Concentração em Processamento Digital de Sinais, e aprovada em sua forma final pelo Programa de Pós-Graduação em Engenharia Elétrica da Universidade Federal de Santa Catarina.’

Prof. Rui Seara, Dr.
Orientador

Prof. Edson Roberto De Pieri, Dr.
Coordenador do Programa de Pós-Graduação em Engenharia Elétrica

Banca Examinadora:

Prof. Rui Seara, Dr.
Presidente

Prof. Marcus Vinicius Lamar, Ph.D.
Co-orientador

Prof. Orlando José Tobias, Dr.

Prof. Leonardo da Silva Resende, Dr.

Prof. Walter Pereira Carpes Júnior, Dr.

AGRADECIMENTOS

- Ao meu orientador, Prof. Rui Seara, e co-orientador, Prof. Marcus Lamar, pelo apoio, excelente relacionamento e pelo voto de confiança incondicional durante o decorrer deste trabalho.
- À minha mãe, irmã e,
- em especial, ao meu irmão, pelas incansáveis discussões e pelas contribuições dadas em todos os âmbitos, durante esses dois anos de mestrado.
- A todos os professores, funcionários e colegas que, de uma forma ou de outra, contribuíram no andamento deste trabalho,
- em especial a Elton Fontão, Orlando Tobias e Leonardo Rezende.
- Aos colegas Alessandro Lima, Cláudia Sedrez, Felipe Santos, Leandro Sobreiro e Lisandre Pontes, pela amizade, incentivo e compreensão.
- Ao CNPq, pelo apoio financeiro.

Resumo da Dissertação apresentada à UFSC como parte dos requisitos necessários para obtenção do grau de Mestre em Engenharia Elétrica.

DETECÇÃO FACIAL: AUTOFACES VERSUS ANTIFACES

Guilherme Holsbach Costa

Maio/2003

Orientador: Prof. Rui Seara, Dr.

Co-orientador: Prof. Marcus Vinicius Lamar, Ph.D.

Área de Concentração: Comunicações e Processamento de Sinais

Palavras-chave: Detecção facial, reconhecimento facial, Autofaces, Antifaces, PCA.

Número de Páginas: 84

RESUMO: No presente trabalho, é desenvolvido um estudo comparativo entre duas técnicas de detecção facial baseadas em projeções vetoriais: Autofaces e Antifaces. O método de Autofaces tem sido significativamente estudado nos últimos anos, enquanto o de Antifaces é ainda considerado o estado-da-arte para a detecção de objetos. Ambos os métodos são descritos de forma detalhada e, para o método de Antifaces, é proposto um procedimento que permite obter os detectores subótimos. Ambos os métodos são avaliados em condições idênticas de teste. Tais avaliações consideram detecções de características faciais, de objetos tridimensionais e de uma face específica, vista de um ângulo frontal. Finalmente, é feita uma análise de sensibilidade dos métodos ao ruído branco Gaussiano aditivo, a distorções no foco e a alterações na cena em que se apresenta o objeto de interesse. Através dos resultados obtidos, é possível constatar que, no método de Antifaces, os critérios para a determinação de algumas variáveis de projeto não estão ainda bem estabelecidos. Além disso, esse método apresenta alta seletividade durante o processo de detecção. O método de Autofaces possui maior capacidade de generalização e menor sensibilidade à adição de ruído, distorções no foco e alterações na cena.

Abstract of Dissertation presented to UFSC as a partial fulfillment of the requirements for the degree of Master in Electrical Engineering.

FACIAL DETECTION: EIGENFACES VERSUS ANTIFACES

Guilherme Holsbach Costa

May/2003

Advisor: Prof. Rui Seara, Dr.

Co-advisor: Prof. Marcus Vinicius Lamar, Ph.D.

Area of Concentration: Communications and Signal Processing.

Keywords: Facial detection, facial recognition, Eigenfaces, Antifaces, PCA.

Number of Pages: 84

ABSTRACT: This dissertation presents a comparative study between two techniques of facial detection based on vector projections: Eigenfaces and Antifaces. The Eigenfaces method has been significantly studied in the last years, while the one of Antifaces is still considered the state-of-art in object detection. Both methods are described in details, and for the Antifaces method, we have proposed a procedure that permits to obtain the suboptimal detectors. Both methods have been assessed in identical test conditions. Such evaluations consider facial characteristic detections, three-dimensional object detection, and specific face detection seen of a frontal angle. Finally, an analysis of the methods' sensitivity is accomplished with respect to the white additive Gaussian noise, focus distortions, and changes in the scene of the desired object. By the results obtained it is possible to verify that, in the Antifaces method, the criteria for obtaining some design variables are not yet well established. Moreover, such a method presents high selectivity in the detection process. The Eigenfaces method presents a larger generalization capacity and smaller sensitivity for additive noise, focus distortions and changes in the scene.

SUMÁRIO

1. INTRODUÇÃO	1
1.1 Definições envolvendo detecção e reconhecimento facial.....	2
1.2 Detecção de faces em uma imagem.....	3
1.2.1 Métodos baseados em conhecimento humano e morfologia.....	4
1.2.2 Métodos baseados em casamento de padrões	6
1.2.3 Métodos baseados em aparência	7
1.3 Representação de imagens.....	8
1.4 Notação.....	9
1.5 Organização da dissertação.....	9
2. AUTOFACES	11
2.1 Descrição das Autofaces	12
2.2 Determinação das Autofaces	14
2.3 Classificação de uma imagem	17
2.4 Considerações	22
3. ANTIFACES.....	24
3.1 Descrição do método de Antifaces	24
3.2 Fundamentação matemática do método	27
3.2.1 Valor esperado do produto interno entre imagens alisadas.....	27
3.2.2 Independência entre detectores	28
3.3 Determinação dos detectores.....	29
3.3.1 Detectores subóticos	31
4. RESULTADOS E CONSIDERAÇÕES	32
4.1 Condições de teste e treinamento.....	32
4.1.1 Detecção de uma face frontal específica.....	32
4.1.2 Detecção de características faciais	34
4.1.3 Detecção de objetos tridimensionais.....	36
4.1.4 Análise de robustez	37

4.2	Resultados obtidos pelo método de Autofaces	38
4.2.1	Detecção de faces frontais.....	39
4.2.2	Detecção de características faciais.....	41
4.2.3	Detecção de objetos tridimensionais.....	43
4.2.4	Análise de robustez.....	44
4.2.5	Considerações gerais sobre o método de Autofaces.....	47
4.3	Resultados obtidos pelo método de Antifaces	47
4.3.1	Detecção de faces frontais.....	47
4.3.2	Detecção de características faciais.....	49
4.3.3	Detecção de objetos tridimensionais.....	51
4.3.4	Análise de robustez.....	52
4.3.5	Considerações gerais sobre o método de Antifaces.....	55
4.4	Considerações gerais	56
5.	DISCUSSÕES E CONCLUSÕES	58
5.1	Discussão dos resultados.....	58
5.2	Contribuições.....	59
5.3	Propostas para trabalhos futuros.....	60
	APÊNDICE 1 – FUNDAMENTAÇÃO METEMÁTICA	61
	APÊNDICE 2 – DEMONSTRAÇÕES ENVOLVENDO AUTOVALORES ...	65
	APÊNDICE 3 – FUNDAMENTOS DA ANÁLISE DE COMPONENTES PRINCIPAIS (PCA)	67
	APÊNDICE 4 – DECOMPOSIÇÃO EM VALORES SINGULARES (SVD) ..	71
	APÊNDICE 5 – DETERMINAÇÃO DOS DETECTORES SUBÓTIMOS	75
	REFERÊNCIAS BIBLIOGRÁFICAS	80

LISTA DE FIGURAS

Figura 1.1 — Modelo de vetorização de uma imagem bidimensional.	9
Figura 2.1 — Exemplo de conjunto de treinamento.	16
Figura 2.2 — Exemplo de média entre as imagens do conjunto de treinamento.....	16
Figura 2.3 — Exemplo de Autofaces (imagens ortogonais).....	17
Figura 2.4 — Versão simplificada do espaço de faces para ilustrar os quatro resultados possíveis de projeção de uma imagem no espaço de faces (U) — neste caso existem dois autovetores principais (\mathbf{u}_1 e \mathbf{u}_2) e três faces conhecidas (ω_1 , ω_2 e ω_3).....	19
Figura 2.5 — Espectro típico dos autovalores e sua divisão entre os dois subespaços ortogonais.	21
Figura 3.1 — Exemplos de imagens de treinamento referentes ao multipadrão \mathbb{T} (linha superior) e exemplos de detectores (linha inferior).	31
Figura 4.1 — Imagem de treinamento.	34
Figura 4.2 — Imagem natural para teste de detecção de uma face frontal específica (tamanho reduzido).	34
Figura 4.3 — Imagem de teste usada para a detecção de características faciais.	35
Figura 4.4 — Imagem de teste usada para a detecção de objetos tridimensionais. A imagem do pato é o objeto de interesse para a detecção.....	36
Figura 4.5 — Imagem de teste sintética em que uma cena é sobreposta por imagens do objeto de interesse com fundo preservado e com fundo transparente.	38
Figura 4.6 — Resultado de detecção de uma face frontal específica: (a) espaço com 50 Autofaces e (b) com 30 Autofaces.	40
Figura 4.7 — Resultados da detecção de uma face frontal específica usando 20 Autofaces.	40
Figura 4.8 — Resultado da detecção de uma face específica em uma imagem natural.	41
Figura 4.9 — Resultado da detecção de olhos para um conjunto de treinamento com 1296 imagens.....	42
Figura 4.10 — Resultado da detecção de olhos para um conjunto de treinamento com 81 imagens.....	42

Figura 4.11 — Resultado da detecção de olhos para conjunto de treinamento com 10 imagens de olhos (olho esquerdo das primeiras 10 imagens).....	43
Figura 4.12 — Detecção, pelo método de Autofaces, de objetos tridimensionais: (a) exata e (b) com alarme falso-positivo, gerado por um limiar de detecção menos rigoroso.....	43
Figura 4.13 — Ocorrência de alarme falso-negativo na detecção. Imagem submetida a ruído branco aditivo; PSNR = 28 dB.	44
Figura 4.14 — Resultado da detecção para uma imagem submetida a ruído branco aditivo; <i>PNSR</i> = 22 dB.	45
Figura 4.15 — Resultado da detecção para uma imagem submetida a ruído branco aditivo; <i>PNSR</i> = 19 dB.	45
Figura 4.16 — Resultado da detecção em que alarmes falso-positivos são gerados em razão da distorção no foco. Imagem submetida a um filtro PB com resposta ao impulso de dimensão (2×2).....	46
Figura 4.17 — Resultado da detecção pelo método de Autofaces.....	46
Figura 4.18 — Resultados de detecção de uma face frontal específica. Para uma detecção sem alarmes falsos, foram necessários: (a) 15 detectores e (b) 16 detectores.	48
Figura 4.19 — Resultado de detecção inexata (100 detectores foram aplicados).	48
Figura 4.20 — Resultados de detecção de uma face frontal específica. Para uma detecção sem alarmes falsos, foram necessários: (a) 13 detectores e (b) 7 detectores.	48
Figura 4.21 — Resultados da detecção de olhos para um conjunto de treinamento com 1296 imagens: (a) melhor resultado obtido sem alarmes falso-positivo; (b) presença de alarmes falso-positivo sem a detecção de todos os olhos.....	49
Figura 4.22 — Resultado da detecção de olhos para um conjunto de treinamento composto por 81 imagens.	50
Figura 4.23 — Resultado da detecção de olhos, usando Antifaces, para um conjunto de treinamento composto por 10 imagens naturais (10 primeiros olhos esquerdos).....	50
Figura 4.24 — Resultados da detecção de objetos tridimensionais: (a) detecção exata; (b) detecção inexata.....	51
Figura 4.25 — Presença de alarmes falsos na detecção por Antifaces. Imagem de teste submetida a ruído branco aditivo; PSNR = 38 dB.	53

Figura 4.26 — Presença de alarmes falsos na detecção por Antifaces. Imagem de teste submetida a ruído branco aditivo; PSNR = 28 dB.	53
Figura 4.27 — Resultado da detecção por Antifaces em que houve uma redução no número de alarmes; imagem submetida a um filtro PB, com resposta ao impulso de dimensão (2×2)	54
Figura 4.28 — Resultado da detecção por Antifaces; imagem submetida a um filtro PB, com resposta ao impulso de dimensão (8×8)	54
Figura 4.29 — Resultados da detecção pelo método de Antifaces: (a) detecção com alarme falso-negativo; (b) detecção com alarmes falso-positivos.	55
Figura D.1 — Exemplo de SVD em que: (a) a matriz não-singular A mapeia o vetor x no espaço de b ; (b) a matriz singular A mapeia um espaço vetorial em outro de menor dimensão. Nesse caso, o espaço nulo de A é mapeado para o zero, a solução de $\mathbf{Ax} = \mathbf{d}$ consiste em qualquer solução particular somada a qualquer vetor do espaço nulo e $\mathbf{Ax} = \mathbf{c}$ é a solução com o menor erro, no sentido dos mínimos quadrados. ...	74

LISTA DE TABELAS

Tabela 4.1 — Valores do alisamento e do limiar M para os casos de (a) detecção exata e (b) detecção com alarmes falso-negativos	52
Tabela 4.2 — Resultados da detecção de características faciais.....	56
Tabela 4.3 — Resultados da avaliação de robustez	57

Com a expansão das pesquisas na área de processamento digital de sinais, uma nova forma de interação entre homens e máquinas vem surgindo. Concorrendo com os tradicionais teclado e *mouse*, aparecem outros canais de comunicação, como voz e imagem. Dentro desse contexto de crescimento tecnológico, o processamento de imagens vem sendo aplicado desde a indústria até as tarefas mais cotidianas na sociedade. Algumas dessas aplicações envolvem a interpretação direta de ações realizadas por pessoas, em que se faz necessária a detecção automática de indivíduos presentes em uma cena. Exemplos dessas aplicações podem ser:

- Compressão de imagens para transmissão em videofonia móvel celular: a imagem do objeto de interesse (por exemplo, a face do interlocutor) é detectada e enviada com maior resolução, enquanto os demais objetos da cena, menos relevantes para que se estabeleça uma conversação, podem ser transmitidos com menor quantidade de detalhes, aproveitando melhor os canais de transmissão.
- Câmera inteligente para videoconferências: acompanha o conferencista, permitindo uma maior mobilidade durante uma palestra, por exemplo.
- Sistemas automáticos de identificação de pessoas: aplicáveis tanto em sistemas de segurança quanto em sistemas que procuram por pessoas desaparecidas em grandes cidades, por exemplo.
- Sistema de reconhecimento de linguagem de sinais para surdos e mudos: nesses casos, é importante a interpretação da expressão facial, assim como a leitura labial.

Tratando-se de aplicações mais clássicas, como os sistemas automáticos de identificação de pessoas, diversos métodos podem ser empregados — chamados sistemas biométricos. Dentre esses sistemas, destacam-se o reconhecimento de impressões digitais, palmas de mão, voz, face e íris. O reconhecimento facial associado a um eficiente sistema de detecção facial aparece como a opção menos invasiva, pois não requer uma ação específica do usuário. Tendo-se, por exemplo, uma entrada monitorada por uma câmera, um sistema de controle de acesso pode liberar a passagem de determinada pessoa a uma área restrita, antes que ela alcance a maçaneta da porta de entrada.

No escopo das aplicações que envolvem comunicação homem-máquina, a localização do usuário é fundamental, seja ela determinante da posição do corpo inteiro ou apenas do rosto de uma ou mais pessoas presentes na cena analisada pelo sistema. Nos últimos cinco anos, o reconhecimento de faces e expressões faciais tem chamado mais a atenção, embora esse seja objeto de estudo há mais de vinte anos por físicos, neurocientistas e engenheiros [1].

1.1 Definições envolvendo detecção e reconhecimento facial

Dada uma imagem arbitrária, o objetivo da detecção facial é determinar a existência de uma ou mais faces na imagem e, em caso positivo, fornecer a(s) sua(s) localização(ões) e dimensões. O reconhecimento ou identificação facial, por sua vez, compara uma imagem de entrada (imagem de teste) com um banco de imagens (padrões) e verifica o melhor casamento entre elas. Existem ainda alguns outros conceitos relacionados à detecção e ao reconhecimento facial. A localização facial busca a posição de uma única face em uma imagem. Esse caso trata-se de uma simplificação da detecção facial em que a cena é assumida como possuindo apenas uma face. A localização de características faciais consiste em detectar componentes como olhos, nariz, sobrancelhas, narinas, boca, lábios ou outras características faciais, também assumindo a existência de apenas uma face na imagem. O rastreamento facial é a estimação contínua da localização e orientação de uma face em uma seqüência de imagens adquiridas em tempo-real (entenda-se por tempo-real, o tempo que decorre entre a chegada de duas informações distintas, alvos de processamento,

na entrada em um sistema). O reconhecimento de expressões faciais consiste em identificar estados de afeição, como alegria, tristeza, desgosto, dentre outros [1].

A detecção de uma face é considerada correta se pelo menos uma determinada porcentagem da área da face (em uma imagem bidimensional) for encontrada. Quando um processo de detecção gera um resultado sem erros de qualquer espécie, a detecção é chamada “exata”. Em geral, dois tipos de erros de detecção podem ocorrer: os alarmes falso-positivos, nos quais uma região da imagem é classificada como contendo uma face quando não a contém, e os alarmes falso-negativos, nos quais uma face presente na imagem não é detectada. Entre os dois tipos, o falso-positivo é menos grave, pois os alarmes gerados podem ainda passar por uma nova etapa de processamento, mais rigoroso, em que se venha a identificar dentre eles apenas os alarmes verdadeiros. Já, na ocorrência de falso-negativos, o resultado final da detecção está fadado a erro.

1.2 Detecção de faces em uma imagem

A detecção facial apresenta-se como uma tarefa de processamento de sinais não trivial em função de as faces serem objetos que não possuem formas bem definidas (rígidas). Em aplicações industriais como, por exemplo, a inspeção automática de uma peça automotiva em uma linha de produção, além de possuírem posição e dimensões bem específicas, as peças normalmente estão em um ambiente com iluminação controlada. Diferente disso, as faces podem apresentar-se com oclusão parcial, variação de posição, diferenças de iluminação (que podem tornar certas características invisíveis, fazendo inclusive com que a face pareça assimétrica), além de se ter normalmente uma cena poluída por outros artefatos provenientes da presença de óculos, barba, chapéu, lenços, etc. Devido à sua complexidade, esses problemas não são sempre considerados e, muitas vezes, são feitas considerações restringindo a ocorrência de alguns deles, como, por exemplo, a consideração de detectar apenas faces que apareçam na posição frontal. Outra razão para que algumas restrições sejam feitas é a aplicação a que se destina o sistema. Em um sistema de controle de acesso, por exemplo, é razoável estipular que não haja oclusão parcial da face.

Diversas podem ser as abordagens para se realizar a detecção ou localização de faces em imagens. A seguir são apresentados os principais métodos utilizados.

1.2.1 Métodos baseados em conhecimento humano e morfologia

Neste tipo de abordagem, os métodos de detecção facial são desenvolvidos a partir do conhecimento humano sobre a caracterização das faces. Por exemplo, uma face pode ser caracterizada por dois olhos dispostos com simetria vertical, nariz e boca, interiores a uma região aproximadamente elíptica. Pelo tamanho e posição de algumas dessas características, também é possível estimar o tamanho e a posição relativa das demais características. Um problema nesse tipo de abordagem é a dificuldade de traduzir em regras simples e bem definidas o conhecimento humano. Se as regras forem muito detalhadas (específicas), algumas faces podem ser rejeitadas (por exemplo, havendo a oclusão de alguma característica facial). Se ao contrário, as regras forem muito genéricas, alarmes falso-positivos podem ser gerados. Além disso, esta abordagem restringe-se apenas à detecção de faces que apareçam em uma (ou poucas) posição(ões). Por exemplo, considerando a possibilidade de as faces aparecerem rotacionadas, a simetria dos olhos pode aparecer horizontalmente e não mais verticalmente. Havendo também a possibilidade da ocorrência de faces vistas de perfil, seria difícil enumerar todas as regras possíveis a serem testadas. Já, para casos de detecção de faces frontais, bons resultados têm sido alcançados [1].

Em razão dos problemas apresentados, alguns métodos são desenvolvidos de forma a procurar por características invariantes nas faces. Características como sobrancelhas, olhos, nariz e boca são facilmente destacadas através de detecção de bordas. Destacadas tais características, métodos estatísticos são aplicados para verificar a existência de uma face. Os problemas que mais afetam esses métodos são variações na iluminação (sombras), ruído e oclusão de características. Outro problema relevante em métodos que procuram por características faciais aparece quando as imagens das faces são relativamente pequenas. Nesses casos, as características ficam representadas por um pequeno número de *pixels* e, normalmente, não podem ser reconhecidas pelo sistema.

A análise de textura pode ser utilizada para diferenciar as faces do fundo ou de outros objetos presentes em uma cena. Tratando-se da detecção de características faciais, a textura pode ser utilizada para detectar a região dos cabelos. O uso associado das informações de textura e cor na segmentação de faces é apresentado em [2], baseado no fato de que cenas contendo faces são geralmente ricas em cores e texturas.

A informação de cor da pele é uma característica já consagrada na detecção facial [3]-[21]. Embora pessoas diferentes possuam cores de pele diferentes, vários estudos comprovam que essa diferença está concentrada na informação de luminância, sendo a informação de crominância muito similar para a pele de todas as pessoas. Imagens em tons de cinza são facilmente influenciadas pela iluminação (sombras) [9], enquanto a cor da pele não varia com a intensidade luminosa e é uma característica insensível a oclusões parciais, diferenças de escala e de orientação das faces [10]. Apesar da cor da pele poder ser percebida com distorções, em função da cor da fonte que a ilumina e de características que variam de câmera para câmera, a identificação de cor tem a vantagem de ser um processo de baixo custo computacional [22]. Por esse motivo, tal abordagem pode servir como uma etapa de pré-processamento, utilizada para reduzir a quantidade de informação a ser processada por um outro método mais eficiente, porém, de maior complexidade computacional.

Das referências consideradas, a maioria agrega ao processamento de cores algum outro tipo de técnica. Em [6] e [8], são combinadas a procura de cor da pele e de cor dos cabelos. São identificadas as regiões de pele e cabelo, e a posição relativa entre ambas é analisada (via lógica nebulosa), avaliando a presença da face. Em [9], são construídos mapas de similaridade com pele e cabelos. *Integral projection* é usada para localização de características faciais. Outras ferramentas, como mapas de bordas, padrões deformáveis (*snakes*) e certa heurística para determinar a relação de medida entre as características da face (razão entre largura e altura do rosto), também são empregadas. Em [13], além da construção de um mapa de similaridade de pele, algoritmos de fechamento, filtros de mediana e *region growing* são aplicados para a eliminação de ruído e da cena em que aparece a face. É feita a procura por olhos nas regiões de pele, através de correlação. Também é usada heurística na eliminação de regiões que “não possuem estrutura de face”

e para estimar a posição da boca a partir da posição dos olhos e região de pele. Filtro de borda e *integral projection* são usados para definir a posição correta da boca. A procura pelo nariz é feita através da estimativa da região e buscando dentro dela encontrar as narinas (pontos escuros). Uma vez detectada a face, o algoritmo faz o seu rastreamento, repetindo o processo de detecção apenas em regiões da imagem próximas à primeira localização. Em [20], é combinada a detecção de cores com simetria local (calculada pelo gradiente em uma pequena vizinhança), gerando dois mapas — mapa de simetria e mapa de cores — que são fundidos por meio de lógica nebulosa, gerando um mapa de possibilidade da existência de uma face na imagem. Em [21], a informação de cor também é utilizada associada a outros métodos. A segmentação de objetos candidatos à face é feita através da limiarização de mapas de probabilidade da presença de pele (baseados na informação de cor). Em seguida, são empregados métodos heurísticos que descartam objetos pequenos demais para constituírem uma face ou que possuam largura e altura não consistentes com um padrão pré-definido. Por fim, a procura por olhos é realizada na região candidata à face, usando correlação. Uma vez localizada a face, ela é rastreada utilizando redes neurais artificiais.

As imagens coloridas são geralmente representadas no espaço de cores RGB (*red*, *green* e *blue*). Dentro desse sistema, todas as cores são compostas a partir desses três componentes principais. Porém, quando o processamento de imagens coloridas visa identificar um determinado subespaço de cores, esse sistema pode não ser o mais adequado. Por esse motivo, uma transformação entre espaços costuma ser efetuada. Procurando avaliar os espaços de cores que se adaptam melhor à tarefa de reconhecimento de cor de pele, comparações entre as principais abordagens são feitas em [22]. Além disso, vários trabalhos vêm sendo desenvolvidos para ajustar os modelos de cor da pele aos mais diferentes ambientes e iluminações [23]-[26].

1.2.2 Métodos baseados em casamento de padrões

No casamento de padrões, um modelo de face padrão (normalmente frontal) é construído manualmente ou através de representações matemáticas apropriadas [1]. Por exemplo, o contorno de uma face pode ser parametrizado pela equação de uma elipse ou

por diversos segmentos de reta. Dada uma imagem de entrada, o valor de correlação com a imagem padrão é calculado para o contorno da face, olhos, boca e nariz, independentemente. A existência de uma face é baseada nos valores de correlação. Essa abordagem tem a vantagem de ser simples em relação à implementação. Entretanto, o casamento de padrões tem se mostrado inadequado para detecção facial, uma vez que não é eficiente frente a variações de escala, posição e forma. O uso de padrões deformáveis (*snakes*) vem tentando amenizar tais problemas [27], visto que eles se adaptam às variações morfológicas das características faciais. Entretanto, esse tipo de abordagem apresenta alguns problemas de inicialização. Além disso, seu uso muitas vezes torna-se complicado em função de o contorno das faces ser, em vários casos, pouco definido [9]. Os padrões deformáveis apresentam ainda elevada sensibilidade ao ruído e alto custo computacional [21].

1.2.3 Métodos baseados em aparência

Nos métodos baseados em aparência, os sistemas de detecção “aprendem” quais são as características das imagens de interesse através de um conjunto de imagens exemplo. Em geral, os métodos baseados em aparência recaem sobre técnicas de análise estatística e aprendizado de máquina (*machine learning*), utilizadas para encontrar características relevantes em uma imagem de face e/ou não-face. As características aprendidas são armazenadas na forma de modelos de distribuição ou de funções discriminantes. Muitos métodos baseados em aparência podem ser formulados através de uma abordagem probabilística. Uma imagem ou um vetor de características é visto como uma variável aleatória \mathbf{x} , a qual é caracterizada como face e não-face através da função de probabilidade condicional $p(\mathbf{x}|face)$ e $p(\mathbf{x}|não-face)$ [1].

Outra abordagem é encontrar uma função discriminante (i.e., uma superfície de decisão, um hiperplano de separação, ou ainda uma função limiar) que separe as classes de faces e não-faces. Por possuírem certa similaridade, as faces encontram-se agrupadas em uma região específica do espaço de imagem. Esse subespaço pode ser delimitado por uma função discriminante, construída para realizar uma classificação (comumente baseada em uma distância) em faces e não-faces. No lugar da função discriminante, uma superfície de

separação não-linear também pode ser considerada através do uso de uma rede neural artificial multicamadas.

Support vector machines têm sido muito empregadas neste tipo de abordagem [28]-[33]. Entretanto, um dos métodos mais consagrados é o de Autofaces (*Eigenfaces*), derivado da Transformada de Karhunen-Loève, também conhecida por Análise de Componentes Principais (PCA) ou ainda Transformada de Hotteling, que será visto detalhadamente no Capítulo 2. Outro método, mais recente do que as Autofaces, é o de Antifaces, que será discutido no Capítulo 3, o qual envolve conceitos similares aos *Rejectors* [34].

1.3 Representação de imagens

As imagens são, de acordo com a sua estrutura, representadas geralmente através de matrizes. A posição de cada *pixel* na matriz corresponde à sua posição no plano da imagem, e a intensidade de cada *pixel* é representada pelo valor de cada elemento da matriz. Visando submeter uma imagem a um tratamento matemático não trivial, é comum alterar essa forma de representação, organizando-se os *pixels* de uma imagem na forma de um vetor. Esse tipo de representação busca reduzir a complexidade computacional e adaptar algumas técnicas tradicionais de cálculo vetorial ao caso de processamento de imagens (cálculo matricial).

O processo de vetorização de uma imagem pode ser feito de diversas formas como, por exemplo, organizando os *pixels* de maneira que uma coluna da matriz original fique disposta sob a outra, ou uma linha ao lado da outra (ordenação lexicográfica), ou mesmo em uma ordem definida por alguma função específica.

No presente trabalho, sendo uma imagem de face $\mathbf{I}(x,y)$ representada por uma matriz de dimensão $(N \times N)$, o processo de vetorização utilizado é descrito na Figura 1.1. A partir dessa vetorização, a imagem de face passa a ser representada por um vetor \mathbf{i} de $(N^2 \times 1)$ elementos, que pode ser visto também como um ponto em um espaço de

dimensão N^2 . A notação adotada para esse processo é “ $\mathbf{i} = \text{vec}(\mathbf{I})$ ”, onde $\text{vec}(\cdot)$ representa a operação de vetorização [35].

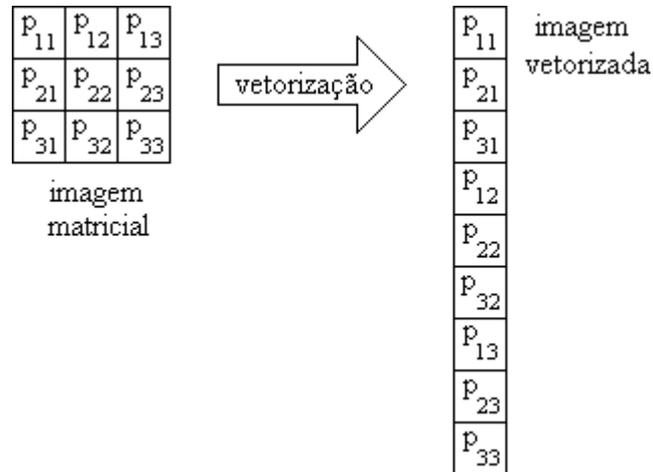


Figura 1.1 — Modelo de vetorização de uma imagem bidimensional.

1.4 Notação

A notação apresentada a seguir será adotada neste trabalho:

- *Constantes escalares*: símbolos maiúsculos em itálico.
- *Variáveis escalares*: símbolos minúsculos em itálico.
- *Vetores*: símbolos minúsculos em negrito. Os vetores serão sempre considerados vetores coluna.
- *Matrizes*: símbolos maiúsculos em negrito

1.5 Organização da dissertação

O Capítulo 2 discute o método de Autofaces, que vem a ser um método significativamente referenciado nos últimos anos. É interessante destacar que, dentre os métodos usados para reconhecimento facial, esse é o que obtém melhores resultados em termos de reconhecimento de faces frontais [36], [37]. O Capítulo 3 apresenta o método de Antifaces. Esse método é considerado o estado-da-arte em detecção de objetos. O

Capítulo 4 explora de forma detalhada as particularidades de ambos os métodos, comparando seus desempenhos para diferentes condições. Por fim, no Capítulo 5 são apresentadas uma discussão final sobre os resultados obtidos, conclusões e sugestões para futuros trabalhos envolvendo detecção e reconhecimento facial. A fundamentação matemática que introduz os Capítulos 2 e 3 é apresentada nos Apêndices 1 a 4. No Apêndice 5, é proposta uma abordagem para a determinação de detectores subótimos, introduzidos no Capítulo 5.

Desenvolver um sistema de reconhecimento de faces é uma tarefa complexa, uma vez que para tal é necessário modelar objetos multidimensionais com formas e características não bem definidas, conforme discutido no Capítulo 1. Em razão dessa complexidade, o método de Autofaces opera com um conjunto de características que não depende de formas geométricas rígidas, utilizando apenas a informação bidimensional da representação facial. Originalmente focada em tarefas de reconhecimento facial, essa abordagem, baseada na teoria da informação, busca identificar um pequeno número de características que venham a ser relevantes para distinguir uma face de outras ou de outros objetos. Uma maneira de caracterizar uma face é verificar a presença de características faciais, tais como olhos, nariz e boca. Entretanto, essas características sofrem uma considerável variação morfológica, quando analisadas para diferentes faces ou expressões faciais. Outro encaminhamento é, sem analisar as características faciais separadamente, preocupar-se apenas com a variação dos valores assumidos pelos *pixels*, em um conjunto de imagens de faces. Em linguagem matemática, esse processo pode ser obtido via componentes principais da distribuição de faces, ou seja, através dos autovetores da matriz de covariância de um conjunto de imagens de faces. Esses autovetores, dos quais alguns formam um espaço vetorial chamado “espaço de faces”, podem ser vistos como um conjunto de características que descrevem a variação entre tais imagens. Cada autovetor descreve a variação dos *pixels* em uma região diferente das imagens [38]. Isso significa dizer que cada um deles descreve a variação dos *pixels* associados a diferentes características faciais, ou ainda que haverá um autovetor relacionado com a região da boca, um com a região do nariz, e assim por diante. Os componentes principais, por serem autovetores associados a um conjunto de faces, são chamados Autofaces.

Imagens de faces de um mesmo indivíduo, vistas como pontos em um espaço de dimensão elevada, dispõem-se de forma agrupada quando projetadas no espaço de faces. Assim, conhecida a posição que uma coleção de imagens de faces (de um mesmo indivíduo) assume, quando é projetada no espaço de faces, é possível fazer o reconhecimento de outras imagens de faces que correspondam a esse mesmo indivíduo. Esse tipo de abordagem é simples, apresentando baixa sensibilidade a mudanças na imagem e demandando um relativamente baixo custo computacional. Diferente dos métodos que operam diretamente com características faciais, como, por exemplo, procura por olhos, boca e nariz, o método de Autofaces atinge bons resultados para reconhecimento de faces observadas em diferentes posições. Outra característica importante desse tipo de abordagem é a capacidade de “aprender” durante o processo de reconhecimento e mais tarde reconhecer novas faces sem que seja necessário retreinar seguidamente o sistema, conforme será discutido na próxima seção.

O método de Autofaces foi motivado por trabalhos que descrevem a representação eficiente de fotografias de faces utilizando Análise de Componentes Principais (PCA) [38]. A partir de um conjunto de imagens de faces, um conjunto ótimo de coordenadas para a compressão das imagens — cada eixo coordenado vem também a formar uma imagem, habitualmente chamada Autoimagem (*Eigenpicture*) — foi proposto. Sendo esse conjunto de imagens suficientemente representativo do universo de faces, qualquer imagem de face poderia ser armazenada (representada) a partir de um vetor de pesos e do conjunto de eixos (Autoimagens) definido. Esses pesos poderiam ser obtidos, para cada face, projetando-a em cada Autoimagem. Assim, surgiu a idéia de que, sendo possível representar por meio de componentes principais e de um vetor de pesos, a PCA poderia ser usada como núcleo em um método de reconhecimento facial baseado na comparação dos pesos que descrevem diferentes faces quando projetadas nas Autoimagens.

2.1 Descrição das Autofaces

Como visto, na linguagem da teoria de informação, as Autofaces extraem a informação relevante de uma face, codificam tal informação e a comparam com um conjunto de faces codificadas de forma similar. Em termos matemáticos, são obtidos os

componentes principais da distribuição de faces, ou seja, os autovetores principais da matriz de covariância do conjunto de imagens de faces (ver Apêndice 1 e 3). Esses autovetores podem ser vistos como um conjunto de características que definem a variação entre as faces.

Cada imagem de face do conjunto de treinamento (imagens usadas para formar o espaço de faces) pode ser representada como uma combinação linear das Autofaces. As faces podem ainda ser representadas de forma aproximada, considerando-se apenas as “melhores” Autofaces, que vêm a ser aquelas associadas aos maiores autovalores. O fato de poucas Autofaces serem usadas implica em certo ganho computacional.

Assim como uma exponencial complexa forma uma base para a decomposição de Fourier, as Autofaces são os elementos da base para a “decomposição em Autofaces”. Portanto, uma imagem de face pode ser representada a partir de um “pequeno” conjunto de pesos armazenado e de um “pequeno” conjunto de faces padrões.

Resumidamente, o método consiste em:

- a) Formar um espaço descrito pelos autovetores principais da matriz de covariância de um determinado conjunto de imagens de treinamento, chamado de espaço de faces \mathbf{U} .
- b) Projetar uma imagem de teste \mathbf{i} sobre este espaço, gerando um vetor de pesos $\boldsymbol{\omega}$ ($\mathbf{U}^T \mathbf{i} = \boldsymbol{\omega}$).
- c) Comparar o vetor de pesos $\boldsymbol{\omega}$ obtido da projeção, com vetores obtidos da projeção de faces conhecidas. Ou, em vez disso, recuperar a imagem projetada para o espaço original ($\mathbf{i}_f = (\mathbf{U}^T)^{-1} \boldsymbol{\omega}$, onde \mathbf{U} é uma matriz ortogonal, ou seja, $\mathbf{U}^{-1} = \mathbf{U}^T$ e, portanto, $\mathbf{i}_f = \mathbf{U} \boldsymbol{\omega}$) e comparar a imagem recuperada com a original.
- d) O espaço de faces pode ser atualizado, assim como o conjunto de vetores de pesos referente às faces conhecidas, caso uma face desconhecida apareça muitas vezes, fazendo com que essa passe também a figurar como uma face conhecida (opcional).

2.2 Determinação das Autofaces

Imagens de dimensão $(N \times N)$ podem ser vistas como pontos no espaço Euclidiano de dimensão N^2 . Por possuírem características comuns entre si, as faces não estarão distribuídas aleatoriamente nesse espaço e, por esse motivo, podem ser descritas alternativamente por um subespaço de menor dimensão [38], [39]. A análise de componentes principais (ver Apêndice 3) é usada para se obter os vetores que melhor representem a distribuição de faces no espaço das imagens. Esses vetores definem o subespaço de imagens de faces (espaço de faces). Seguindo essa filosofia, as Autofaces podem ser determinadas como se segue.

Sendo \mathbf{C}_x a matriz de covariância de uma variável \mathbf{x} , ela pode ser definida de acordo com (A1.7) por

$$\mathbf{C}_x = E\{[\mathbf{x} - E(\mathbf{x})][\mathbf{x} - E(\mathbf{x})]^T\}. \quad (2.1)$$

Considerando-se que \mathbf{x} é uma variável com média zero ($E(\mathbf{x}) = 0$), \mathbf{C}_x passa a ser dada por

$$\mathbf{C}_x = E(\mathbf{x}\mathbf{x}^T). \quad (2.2)$$

Sendo definido então $\mathbf{i}_1, \mathbf{i}_2, \dots, \mathbf{i}_M$, como o conjunto de treinamento das imagens de faces, a média das faces do conjunto é obtida por

$$\boldsymbol{\Psi} = \frac{1}{M} \sum_{k=1}^M \mathbf{i}_k, \quad (2.3)$$

e cada face difere da média pelo vetor $\boldsymbol{\phi}_k = \mathbf{i}_k - \boldsymbol{\Psi}$. Logo, a estimativa da matriz de covariância do conjunto de treinamento é dada por

$$\mathbf{C} \approx \mathbf{A}\mathbf{A}^T = \sum_{k=1}^M \boldsymbol{\phi}_k \boldsymbol{\phi}_k^T, \quad (2.4)$$

onde $\mathbf{A} = [\boldsymbol{\phi}_1, \boldsymbol{\phi}_2, \dots, \boldsymbol{\phi}_M]$.

A matriz \mathbf{C} , neste caso, é equivalente à estimativa da matriz de autocorrelação em (A3.15). O conjunto de vetores representado pela matriz \mathbf{A} é então submetido à análise de componentes principais, dando origem aos autovetores \mathbf{u}_k e aos autovalores λ_k da matriz de covariância \mathbf{C} .

Em razão das linhas da matriz \mathbf{A} possuírem média zero e do número de *pixels* ser maior do que o número de imagens do conjunto de treinamento, $N^2 > M$, existirão apenas $(M-1)$ autovetores não-nulos, em vez de N^2 , para a representação das faces que compõem a base. Os demais $[N^2 - (M-1)]$ autovetores estarão associados a autovalores nulos (ver Apêndice 2). Sendo necessário avaliar apenas $(M-1)$ autovalores, um método alternativo pode ser usado, no qual o custo computacional é consideravelmente reduzido quando $M \ll N^2$. Esse método determina os autovalores a partir de uma matriz de dimensão $(M \times M)$ em vez de $(N^2 \times N^2)$. Desejando-se, originalmente, encontrar os autovetores \mathbf{u}_k e autovalores λ_k , de $\mathbf{A}\mathbf{A}^T$ de dimensão $(N^2 \times N^2)$, que obedeçam à

$$\mathbf{A}\mathbf{A}^T \mathbf{u}_k = \lambda_k \mathbf{u}_k, \quad (2.5)$$

é possível fazê-lo através da determinação dos autovetores \mathbf{v}_i e dos autovalores μ_i da matriz $\mathbf{A}^T \mathbf{A}$ de dimensão $(M \times M)$, de forma que

$$\mathbf{A}^T \mathbf{A} \mathbf{v}_i = \mu_i \mathbf{v}_i. \quad (2.6)$$

Para mostrar que os $(M-1)$ autovalores não nulos e os autovetores associados à matriz $\mathbf{A}\mathbf{A}^T$ podem ser obtidos através dos M autovalores e autovetores de $\mathbf{A}^T \mathbf{A}$, procede-se da seguinte forma: pré-multiplica-se ambos os lados de (2.7) por \mathbf{A} , resultando em

$$\mathbf{A}\mathbf{A}^T \mathbf{A} \mathbf{v}_i = \mu_i \mathbf{A} \mathbf{v}_i. \quad (2.7)$$

Agrupando-se os termos convenientes, (2.7) assume a mesma forma de (2.5),

$$\mathbf{A}\mathbf{A}^T (\mathbf{A} \mathbf{v}_i) = \mu_i (\mathbf{A} \mathbf{v}_i), \quad (2.8)$$

ficando claro que os vetores $\mathbf{A}\mathbf{v}_i$, por satisfazerem o mesmo sistema de equações que (2.6), correspondem apenas a alguns autovetores da matriz de covariância $\mathbf{A}\mathbf{A}^T$. Uma vez que $\mathbf{A}\mathbf{A}^T$ e $\mathbf{A}^T\mathbf{A}$ possuem o mesmo número de autovalores não-nulos¹ e que eles satisfazem o mesmo sistema de equações, é possível observar que, dentre os autovetores \mathbf{v}_i , $(M-1)$ correspondem aos autovetores \mathbf{u}_k associados aos autovalores λ_k não-nulos.

Exemplos de um conjunto de treinamento, média e Autofaces, provenientes desse conjunto, são apresentados através das Figuras 2.1, 2.2 e 2.3, respectivamente.



Figura 2.1 — Exemplo de conjunto de treinamento.



Figura 2.2 — Exemplo de média entre as imagens do conjunto de treinamento.

¹ O posto de $\mathbf{A}\mathbf{A}^T$ é igual ao posto de $\mathbf{A}^T\mathbf{A}$ ([35], p-125), logo ambas possuem o mesmo número de autovalores não-nulos.



Figura 2.3 — Exemplo de Autofaces (imagens ortogonais).

2.3 Classificação de uma imagem

Uma imagem \mathbf{i} pode ser projetada dentro do espaço de faces, efetuando-se a operação $\omega_k = \mathbf{u}_k^T(\mathbf{i} - \boldsymbol{\psi})$ para $k = 1, 2, \dots, M$, onde M é a dimensão escolhida para o espaço de faces (número de autovetores associados aos autovalores mais significativos, escolhido experimentalmente [39]). Os pesos ω_k representam a contribuição de cada Autoface na representação da imagem \mathbf{i} . Esses pesos são organizados em um vetor $\boldsymbol{\omega} = [\omega_1, \omega_2, \dots, \omega_M]^T$.

Dentro do paradigma das Autofaces, uma imagem de face e sua reconstrução no espaço original, após ter sido projetada no espaço de faces, devem ser idênticas (para uma situação ideal). Logo, a energia do erro residual de reconstrução de uma imagem projetada no espaço de faces,

$$\varepsilon^2(\boldsymbol{\omega}) = \sum_{i=M+1}^{N^2} \omega_i^2, \quad (2.9)$$

deve ser a menor possível. Para que isso aconteça, a energia da imagem deve ser preservada durante o processo de projeção. A relação entre as energias que a imagem apresenta, quando ela se encontra nos espaços original e de faces, é definida como Distância do Espaço de Faces (DFFS — *Distance From Face Space*). A DFFS é equivalente à energia do erro residual ε^2 e pode ser facilmente computada a partir dos M componentes principais e da norma da imagem (com média normalizada) ϕ , como segue:

$$\varepsilon^2(\Phi) = \|\Phi\|^2 - \sum_{i=1}^M \omega_i^2. \quad (2.10)$$

As projeções de imagens de faces de diferentes indivíduos ocupam posições bastante diferentes no espaço de faces. Por outro lado, quando mais de uma imagem da face de um mesmo indivíduo é projetada no espaço de faces, suas projeções agrupam-se em uma pequena região desse espaço. Cada um desses agrupamentos pode ser visto como uma classe. A distância Euclidiana (ver Apêndice 1) entre um ponto qualquer, pertencente ao espaço de faces, e o centróide de uma classe é definida por Distância no Espaço de Faces (DIFS — *Distance in Face Space*).

Em resumo, a distância entre uma imagem de face e o espaço de faces (DFFS) deve ser pequena (menor do que um pré-determinado limiar). Além disso, uma imagem de face conhecida deve ter sua projeção compreendida no espaço de faces e próxima à classe correspondente. Isto é, sua distância da respectiva classe no espaço de faces (DIFS) deve ser pequena (menor do que um segundo limiar pré-estabelecido).

De modo geral, existem quatro possibilidades para a posição de uma imagem qualquer em relação ao espaço de faces (Figura 2.4): (1) perto do espaço de faces e perto de uma classe; (2) perto do espaço de faces e distante de qualquer classe; (3) fora (distante) do espaço de faces, porém com projeção no espaço de faces perto de uma das classes; e (4) distante do espaço de faces e com projeção no espaço de faces distante de qualquer classe de faces conhecidas. Nos dois primeiros casos, as imagens correspondem a faces. No primeiro, a face é detectada e identificada. No segundo, a face é detectada, porém, sua identidade não é reconhecida. Os últimos dois casos correspondem a imagens que não representam faces. Na maior parte das outras abordagens discutidas na literatura, quando ocorre a situação descrita por (3), um alarme falso-positivo é gerado. Dentro do paradigma das Autofaces, entretanto, esse alarme falso-positivo não ocorre, em razão do elevado valor da DFFS.

O modo mais simples para se reconhecer uma face, ou seja, para determinar a que classe pertence uma imagem de face \mathbf{f} , através do método de Autofaces, é determinar a classe k que minimiza a distância Euclidiana $\varepsilon_k = \|\omega - \omega_k\|$ (DIFS), onde ω é o vetor de

pesos resultante da projeção de \mathbf{f} e ω_k é um vetor que representa o centróide do agrupamento (*cluster*) relativo à k -ésima classe. A face será então considerada como pertencente à classe k quando o menor erro ε_k estiver abaixo de um pré-estabelecido limiar. Caso contrário, a face será considerada desconhecida.

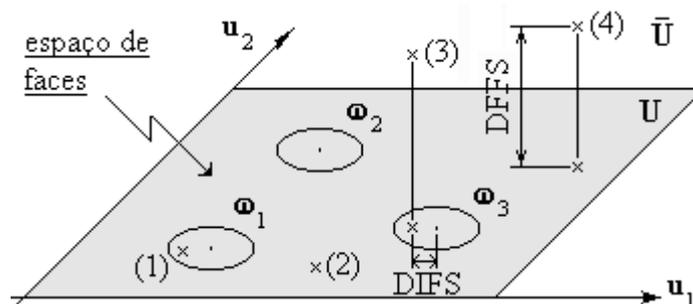


Figura 2.4 — Versão simplificada do espaço de faces para ilustrar os quatro resultados possíveis de projeção de uma imagem no espaço de faces (U) — neste caso existem dois autovetores principais (u_1 e u_2) e três faces conhecidas (ω_1 , ω_2 e ω_3).

O método de Autofaces também pode ser utilizado para detecção e localização facial, fazendo uso do componente ortogonal ao espaço de faces, DFFS (Figura 2.4). Como foi discutido, a distância entre uma imagem \mathbf{i} e sua projeção no espaço de faces (DFFS) é simplesmente a distância Euclidiana entre $\phi = \mathbf{i} - \psi$ (a imagem \mathbf{i} com média ajustada) e $\phi_f = \sum \omega_k \mathbf{u}_k$ (a reconstrução da projeção de ϕ no espaço de faces). Assim, quando se deseja detectar, em uma imagem de entrada \mathbf{I} , uma subimagem \mathbf{i} que represente uma face, é possível medir a DFFS de todas as regiões de \mathbf{I} que possuem dimensões de \mathbf{i} . Dessa forma, seria possível montar um “mapa de faces”, indicando, a partir das DFFS medidas, a probabilidade de haver ou não uma face em uma determinada posição de \mathbf{I} . Esse tipo de procura é um processo que demanda alto custo computacional. Porém, alguns métodos de simplificação podem ser utilizados a fim de aumentar a velocidade final da detecção, como será discutido a seguir.

O componente da imagem \mathbf{i} que é projetada no espaço de faces também pode ser interpretado como uma função distribuição de probabilidade. Dada uma boa estimativa da

média $\boldsymbol{\psi}$ e da matriz de covariância \mathbf{C} do conjunto de treinamento, a probabilidade ou verossimilhança de um padrão de entrada \mathbf{i} é dada por

$$P(\mathbf{i} | \boldsymbol{\omega}) = \frac{1}{(2\pi)^{\frac{N^2}{2}} |\mathbf{C}|^{\frac{1}{2}}} e^{-\frac{1}{2}(\mathbf{i}-\boldsymbol{\psi})^T \mathbf{C}^{-1}(\mathbf{i}-\boldsymbol{\psi})}. \quad (2.11)$$

A estatística utilizada para caracterizar essa vizinhança é a distância de Mahalanobis [40]

$$d(\mathbf{i}) = \boldsymbol{\phi}^T \mathbf{C}^{-1} \boldsymbol{\phi}, \quad (2.12)$$

onde $\boldsymbol{\phi} = \mathbf{i} - \boldsymbol{\psi}$. No lugar de calcular esse produto de forma direta, a distância de Mahalanobis pode ser determinada através dos autovalores e dos autovetores de \mathbf{U} , ou seja:

$$\begin{aligned} d(\boldsymbol{\phi}) &= \boldsymbol{\phi}^T \mathbf{C}^{-1} \boldsymbol{\phi} \\ &= \boldsymbol{\phi}^T [\mathbf{U} \boldsymbol{\Lambda}^{-1} \mathbf{U}^T] \boldsymbol{\phi}, \\ &= \boldsymbol{\omega}^T \boldsymbol{\Lambda}^{-1} \boldsymbol{\omega} \end{aligned} \quad (2.13)$$

onde $\boldsymbol{\Lambda}$ é a matriz diagonal de autovalores de \mathbf{C} . Em razão de $\boldsymbol{\Lambda}$ ser diagonal, (2.13) pode ser expressa da seguinte forma:

$$d(\boldsymbol{\phi}) = \sum_{i=1}^{N^2} \frac{\omega_i^2}{\lambda_i}. \quad (2.14)$$

O cálculo da distância de Mahalanobis através de (2.14) não é conveniente em termos de carga computacional, uma vez que requer a determinação de N^2 valores de ω_i . Devido a isso, o valor de $d(\boldsymbol{\phi})$ é estimado, usando-se apenas os M componentes principais. Para tanto, a soma é particionada em dois termos independentes, que correspondem ao subespaço principal $\mathbf{U} \Rightarrow \{\boldsymbol{\phi}_i\}_{i=1}^M$ e seu complemento ortogonal $\bar{\mathbf{U}} \Rightarrow \{\boldsymbol{\phi}_i\}_{i=M+1}^{N^2}$,

$$d(\boldsymbol{\phi}) = \sum_{i=1}^M \frac{\omega_i^2}{\lambda_i} + \sum_{i=M+1}^{N^2} \frac{\omega_i^2}{\lambda_i}. \quad (2.15)$$

É fácil notar que o primeiro somatório pode ser calculado pela projeção de \mathbf{i} dentro do subespaço principal M -dimensional \mathbf{U} . O segundo somatório é definido como a DFFS $\varepsilon(\phi)^2$, que pode ser calculada a partir de (2.10). Baseada nessas considerações, uma estimativa para $d(\phi)$ é dada por

$$\begin{aligned} d(\phi) &= \sum_{i=1}^M \frac{\omega_i^2}{\lambda_i} + \frac{1}{\rho} \left[\sum_{i=M+1}^{N^2} \omega_i^2 \right] \\ &= \sum_{i=1}^M \frac{\omega_i^2}{\lambda_i} + \frac{\varepsilon(\phi)^2}{\rho}, \end{aligned} \quad (2.16)$$

onde o valor de ρ é uma estimativa da soma dos $(N^2 - M)$ autovalores menos significativos. O valor de ρ pode ser estimado através dos autovalores disponíveis e de uma função do tipo $1/f$, que define um espectro típico de autovalores [40] para um conjunto de padrões morfologicamente homogêneos (Figura 2.5).

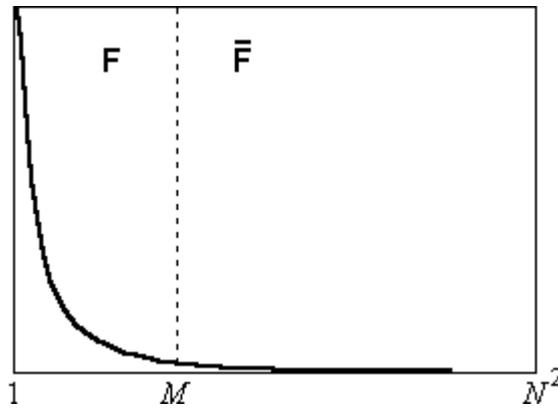


Figura 2.5 — Espectro típico dos autovalores e sua divisão entre os dois subespaços ortogonais.

Em [15], é apresentado um sistema de detecção facial em que o método de Autofaces é aplicado em um mapa de probabilidade dos *pixels* que pertencem a uma região de pele. Esse sistema é proposto para a detecção de faces frontais em imagens coloridas e, para os exemplos apresentados em tal referência, alcança resultados melhores do que o algoritmo original das Autofaces. Nesse caso, ρ é estimado como $1/(k\lambda_M)$, onde λ_M é o

menor autovalor calculado e k é uma constante que assume um pequeno valor ($1 \leq k \leq 5$) e é determinada experimentalmente.

2.4 Considerações

Foi constatado em [38] e [41] que, na prática, a cena exerce uma influência significativa sobre o reconhecimento facial. Em ambos os trabalhos, foram aplicadas máscaras Gaussianas centradas nas faces a fim de destacar as características faciais e eliminar a informação correspondente à cena em que se localiza a face. Esse tipo de problema é superado em [42], em que é proposto um método localizado de análise de componentes principais. No lugar da procura por faces, é feita a procura por características como olhos, nariz e boca, usando espaços não de faces, mas de características específicas (espaço de olhos, espaço de narizes, etc.). No mesmo trabalho, uma etapa de pré-processamento foi considerada fundamental. Foi constatada melhora na robustez do sistema a partir de um ajuste de faixa dinâmica, feito de acordo com

$$\mathbf{I}' = 255 \cdot \frac{\mathbf{I} - \min(\mathbf{I})}{\max(\mathbf{I}) - \min(\mathbf{I})}, \quad (2.17)$$

onde \mathbf{I} é uma imagem de entrada, $\min(\mathbf{I})$ e $\max(\mathbf{I})$ são o mínimo e o máximo valor dos *pixels* da imagem \mathbf{I} , respectivamente, e $\mathbf{I}' \in [0, 255]$ é o resultado final. Uma justificativa para essa melhora é o fato de que, com o ajuste feito através de (2.17), o comportamento do sistema aproxima-se ao de uma íris, que atua como um mecanismo de normalização de iluminação, ajustando o nível de contraste da informação de entrada.

Outros fatores que influenciam significativamente o desempenho do método são a presença de faces em diferentes escalas e de faces observadas de múltiplas posições, nas imagens de entrada. Em [38], esse problema é contornado com o uso de bases que reconhecem a escala da imagem de entrada, para posteriormente projetá-la nas bases que de fato realizam o reconhecimento. Em [41], foi desenvolvido um sistema que requer auxílio manual para avaliação de escala e posição das faces, como pré-processamento, baseado na distância interocular. Ambos os trabalhos destacam que diferenças de escala exercem maior influência no resultado da detecção do que variações na posição em que as

faces são observadas. Em [15], o problema de escala é abordado da seguinte forma: um mapa de probabilidade dos *pixels* representando pele é re-escalado diversas vezes. Os mapas resultantes do re-escalamento são submetidos à detecção, gerando novos mapas de DFFS. A posição do mínimo global entre todos os mapas de DFFS é tomada como sendo a posição da face detectada. Para casos em que pode ser assumida a presença de mais de uma face na imagem, um limiar é pré-estabelecido. Nesses casos, para evitar a sobreposição de duas regiões detectadas como contendo faces, primeiro o mínimo global é considerado e, depois, os demais pontos que venham a passar pelo limiar pré-estabelecido, em ordem crescente, de forma que não haja sobreposição.

O método de Antifaces foi proposto em [43] como uma ferramenta para detecção não apenas de um determinado padrão \mathbf{t} , mas de uma classe de padrões \mathbb{T} (também chamada multipadrão). Considerando que a classe \mathbb{T} seja representada por N padrões, qualquer método convencional de detecção poderá ser aplicado N vezes sobre a imagem (fazendo uma procura distinta para cada um dos N padrões) de forma a solucionar o problema de detecção. Entretanto, essa seria uma tarefa de alto custo computacional. As Antifaces, comparadas com métodos convencionais de detecção, apresentam simplicidade computacional mesmo para casos em que o número de padrões N é elevado.

3.1 Descrição do método de Antifaces

Um conceito básico referente a reconhecimento de padrões é o de função discriminante. Por exemplo, para avaliar se um ponto (x, y) , no plano, pertence a uma semi-circunferência de raio unitário, o método mais rápido é através do cálculo de $f(x, y) = x^2 + y^2 - 1$. De forma genérica, é utilizada uma função, denominada discriminante, que assume o valor zero no conjunto a ser detectado — e somente nesse conjunto. Essa função também pode ser usada para verificar a proximidade desse ponto à circunferência, pois, quanto menor o valor assumido pela função, menor o erro entre a posição do ponto (x, y) e a posição da circunferência. No exemplo anterior, apenas uma função discriminante ($f(x, y)$) é necessária para descrever a classe. Em outros casos, mais de uma função discriminante pode ser necessária. Assim, para testar se um vetor \mathbf{x} — o qual pode ser visto como um ponto no espaço Euclidiano de dimensão elevada — pertence a uma classe \mathbb{T} , é necessário que se verifique a condição $|f_i(\mathbf{x})| \leq \varepsilon_i$ para todo $1 \leq i \leq M$.

O processo de verificações pode ser encurtado uma vez que apenas seja testada a condição f_i caso a condição f_{i-1} tenha sido verificada. Esse método de detecção progressiva reduz substancialmente o tempo total de detecção. Tratando-se de detecção de imagens, é habitual o termo “função discriminante” ser substituído por “detector”.

Uma vez que as imagens submetidas à detecção de padrões (por exemplo, detecção facial) geralmente possuem dimensão elevada, um detector deve ser simples o suficiente para que o algoritmo seja rápido e possa executar sua tarefa em tempo-real. Por esse motivo, os detectores, nessa abordagem, são projetados para serem lineares e atuarem através de produto interno com tais imagens (vistas como um vetor, conforme representado na Figura 1.1).

O vetor \mathbf{d}_i é candidato a detector de um multipadrão \mathbb{T} , composto por N padrões \mathbf{t}_j , caso $|\langle \mathbf{d}_i, \mathbf{t}_j \rangle|$ seja muito pequeno para todo $\mathbf{t}_j \in \mathbb{T}$. Para que essa afirmação seja procedente, os detectores, assim como as imagens, devem ser normalizados de forma a possuírem norma unitária. No entanto, caso os detectores \mathbf{d}_i sejam escolhidos obedecendo apenas a essa condição, não existe garantia de que $|\langle \mathbf{d}_i, \mathbf{y} \rangle|$ seja grande para um padrão $\mathbf{y} \notin \mathbb{T}$. Sendo $|\langle \mathbf{d}_i, \mathbf{y} \rangle|$ também pequeno, alarmes falso-positivos são gerados. Portanto, o detector deve ser escolhido de forma que ele não somente resulte em valores baixos, quando aplicado a imagens pertencentes a \mathbb{T} , como também em valores elevados, quando aplicado a padrões não-membros de \mathbb{T} .

A maioria das imagens que ilustram cenas cotidianas é alisada (com poucos componentes de alta frequência) [43]. Como será discutido na Seção 3.2.1, o valor absoluto do produto interno entre dois vetores alisados é, em média, elevado. Se \mathbf{d}_i é um candidato a detector do multipadrão \mathbb{T} , supõe-se que não apenas $|\langle \mathbf{d}_i, \mathbf{t}_j \rangle|$ seja pequeno, como também \mathbf{d}_i seja alisado. Dessa forma, caso $\mathbf{y} \notin \mathbb{T}$, existe uma alta probabilidade de que $|\langle \mathbf{d}_i, \mathbf{y} \rangle|$ seja grande; o que permite que \mathbf{y} seja rejeitado, isto é, que ele seja classificado como não-membro de \mathbb{T} . Visto que uma das possíveis aplicações desse

método é detecção facial e, em função de o detector, nesse caso, ser um vetor que “anula” uma imagem de face, ele é chamado de Antiface.

De acordo com o que foi exposto, as Antifaces (detectores) são definidas como vetores que satisfazem as seguintes condições:

- O valor absoluto do produto interno entre os detectores e as imagens de \mathbb{T} é pequeno $|\langle \mathbf{d}_i, \mathbf{t}_j \rangle| \leq \varepsilon$.
- Os detectores devem ser alisados, o que resulta em um valor grande de $|\langle \mathbf{d}_i, \mathbf{y} \rangle|$ para todo $\mathbf{y} \notin \mathbb{T}$. Essa é a característica que permite ao detector separar imagens sem interesse das imagens pertencentes a \mathbb{T} .
- Os detectores devem atuar de forma independente, i.e., os alarmes falsos, gerados pelos diferentes detectores, devem ser decorrelacionados. Como será apresentado na Seção 3.2.2, isso não implica em que o produto interno de diferentes detectores seja zero, mas em uma condição ligeiramente diferente. A independência dos detectores é crucial para o sucesso do algoritmo, pois, sendo obedecida tal condição, o número de alarmes falsos decresce exponencialmente com o aumento do número de detectores empregados no processo.

Obtidos os detectores, o processo passa a ser de fácil implementação. Uma imagem é classificada como membro de \mathbb{T} se o valor absoluto de seu produto interno com cada detector ficar abaixo de um determinado limiar (limiar esse, diferente para cada detector). Geralmente, o limiar é escolhido como sendo duas vezes o máximo valor absoluto do detector em questão com os membros do conjunto de treinamento que representam \mathbb{T} [43]. Esse fator de dois visa permitir a detecção, não somente dos membros do conjunto de treinamento, como também de imagens semelhantes a eles, a fim de que o método possua capacidade de generalização.

Uma vez que a imagem de entrada precisa satisfazer as condições impostas por todos os detectores, é suficiente aplicar o segundo detector apenas às imagens que passarem pelo primeiro, e assim por diante.

3.2 Fundamentação matemática do método

3.2.1 Valor esperado do produto interno entre imagens alisadas

Uma imagem é considerada alisada quando possui poucos componentes de alta frequência. A transformada discreta do cosseno (DCT), de forma similar à transformada discreta de Fourier (DFT), fornece, através de seus coeficientes, uma análise em frequência de um determinado sinal. Seqüências reais, quando submetidas à DFT, geralmente geram coeficientes complexos. Quando submetidos à DCT, entretanto, sempre geram coeficientes reais. Esse fato motiva o uso da DCT na análise de alisamento de uma imagem, no que diz respeito ao método de Antifaces.

A distribuição de Boltzmann, que tem se mostrado razoável como modelo para imagens [43], assegura a uma imagem \mathbf{I} uma probabilidade proporcional à exponencial do negativo de uma medida de alisamento. Para esse tipo de análise, é preferível trabalhar-se no domínio da frequência, onde o alisamento de uma imagem \mathbf{I} de $(N \times N)$ *pixels* pode ser descrito por $S(\mathbf{I})$ como segue:

$$S(\mathbf{I}) = \sum_{k=2}^N \sum_{l=2}^N (k^2 + l^2) \mathcal{I}^2(k, l), \quad (3.1)$$

onde $\mathcal{I}(k, l)$ são os coeficientes da Transformada Discreta do Cosseno (DCT) da imagem \mathbf{I} . A probabilidade da ocorrência de componentes de baixa frequência (alisamento) de \mathbf{I} é então definida, de acordo com a distribuição de Boltzmann, como,

$$\Pr(\mathbf{I}) \propto e^{-S(\mathbf{I})}, \quad (3.2)$$

onde a imagem \mathbf{I} deve ser normalizada para média zero, o que implica em $\mathcal{I}(1, 1) = 0$.

Da definição de valor esperado e de (3.2), é possível formalizar a observação “o valor absoluto do produto interno entre dois vetores alisados é, em média, elevado”. Para uma dada imagem \mathbf{F} de dimensão $(N \times N)$ *pixels*, o valor esperado do quadrado do produto interno com outra imagem \mathbf{I} é

$$E(\langle \mathbf{F}, \mathbf{I} \rangle^2) = \int_{\mathbb{R}^{N \times N}} \langle \mathbf{F}, \mathbf{I} \rangle^2 \Pr(\mathbf{I}) d\mathbf{I}, \quad (3.3)$$

que pode ser calculado no domínio DCT via relação de Parseval, substituindo-se (3.1) e (3.2) em (3.3). Assim,

$$E(\langle \mathcal{F}, \mathcal{I} \rangle^2) = \int_{\mathbb{R}^{N-1 \times N-1}} \left[\sum_{k=2}^N \sum_{l=2}^N \mathcal{F}(k, l) \mathcal{I}(k, l) \right]^2 e^{[-\sum_{k=2}^N \sum_{l=2}^N (k^2 + l^2) \mathcal{I}^2(k, l)]} d\mathcal{I}, \quad (3.4)$$

onde \mathcal{F} e \mathcal{I} denotam as transformadas DCT de \mathbf{F} e \mathbf{I} , respectivamente. Em [43], é demonstrado, tanto de forma teórica quanto através de simulações, que (3.4) é proporcional à

$$E(\langle \mathcal{F}, \mathcal{I} \rangle^2) \propto \sum_{k=2}^N \sum_{l=2}^N \frac{\mathcal{F}^2(k, l)}{(k^2 + l^2)^{\frac{3}{2}}}. \quad (3.5)$$

A partir dessa equação, é possível observar que imagens com coeficientes da DCT elevados em baixas frequências produzirão um alto valor esperado. Isso ocorre em decorrência dos coeficientes de ponderação possuírem amplitude decrescente com o aumento da frequência.

3.2.2 Independência entre detectores

Mesmo que, para um dado detector \mathbf{d} , $|\langle \mathbf{d}, \mathbf{y} \rangle|$ atinja, em média, valores altos para imagens $\mathbf{y} \notin \mathbb{T}$, podem existir imagens não pertencentes a \mathbb{T} que resultam em um baixo valor absoluto desse produto interno. Nesses casos, essas imagens serão erroneamente classificadas como membros de \mathbb{T} . Portanto, não é apropriado considerar que um único detector seja capaz de gerar um resultado isento de alarmes falso-positivos. Por esse motivo, mais de um detector é utilizado em um processo de detecção. Tais detectores devem atuar de forma independente, o que implica em: sendo p_1 a taxa de alarmes falsos de \mathbf{d}_1 (definida como a porcentagem de alarmes falso-positivos) e p_2 a taxa de \mathbf{d}_2 , então, a taxa de alarmes falsos da aplicação de ambos detectores \mathbf{d}_1 e \mathbf{d}_2 será dada pelo produto

$p_1 p_2$. Uma vez que todo o esquema de detecção se baseia na distribuição de probabilidade definida em (3.2), a noção de independência entre os alarmes gerados por detectores \mathbf{d}_1 e \mathbf{d}_2 pode ser expressa por:

$$\int_{\mathbb{R}^{N \times N}} \langle \mathbf{d}_1, \mathbf{i} \rangle \langle \mathbf{d}_2, \mathbf{i} \rangle \Pr(\mathbf{i}) d\mathbf{i} = 0, \quad (3.6)$$

onde $\Pr(\mathbf{i})$ é a probabilidade de \mathbf{i} , definida por (3.2). Com base na determinação de (3.5) e simbolizando a integral descrita em (3.6) por $\gamma(\mathbf{d}_1, \mathbf{d}_2)$, tem-se que [43]

$$\gamma(\mathbf{d}_1, \mathbf{d}_2) = \sum_{k=2}^N \sum_{l=2}^N \frac{\mathcal{D}_1(k, l) \mathcal{D}_2(k, l)}{(k^2 + l^2)^{3/2}}, \quad (3.7)$$

onde \mathcal{D}_1 e \mathcal{D}_2 são as transformadas DCT de \mathbf{d}_1 e \mathbf{d}_2 , respectivamente.

3.3 Determinação dos detectores

Para a determinação do primeiro detector \mathbf{d}_1 , o seguinte problema de otimização deve ser resolvido:

- \mathbf{d}_1 deve possuir norma unitária.
- $|\langle \mathbf{d}_1, \mathbf{t}_j \rangle|$ deve ser pequeno para toda imagem \mathbf{t}_j pertencente ao multipadrão \mathbb{T} . Para que essa condição faça sentido, as imagens de entrada também devem possuir norma unitária.
- \mathbf{d}_1 deve ser tão alisado quanto possível, mantendo-se dentro das duas restrições anteriores.

A solução para o problema de otimização pode então ser realizada da seguinte forma: é escolhido um valor apropriado para

$$M = \max_{\mathbf{t}_j \in \mathbb{T}} |\langle \mathbf{d}_1, \mathbf{t}_j \rangle|. \quad (3.8)$$

Em [43], é sugerido que o valor exato de M não é relevante, desde que M seja substancialmente menor do que o valor absoluto do produto interno de duas imagens quaisquer. Para imagens de dimensão (20×20) *pixels*, um valor experimental sugerido para essa grandeza é da ordem de $M = 10^{-5}$. Não sendo possível atingir um valor tão baixo — o que pode acontecer no caso de, por exemplo, o multipadrão \mathbb{T} ser relativamente extenso — um valor de M mais elevado não acarretaria problema. Definido o valor de M , minimiza-se

$$\max \left| \langle \mathbf{d}_1, \mathbf{t}_j \rangle \right|_{\mathbf{t}_j \in \mathbb{T}} + \lambda S(\mathbf{d}_1), \quad (3.9)$$

onde $\lambda (> 0)$ é um escalar de ponderação da parcela de alisamento. Então, por método de bissecção (*binary search*) em λ , ajusta-se o valor da primeira parcela de forma que

$$\max \left| \langle \mathbf{d}_1, \mathbf{t}_j \rangle \right|_{\mathbf{t}_j \in \mathbb{T}} = M. \quad (3.10)$$

Sendo a otimização efetuada no domínio da freqüência, o problema torna-se bastante simples e basta que a transformada inversa do detector seja calculada ao fim do processo para que se obtenha o valor desejado de \mathbf{d}_1 .

Obtido \mathbf{d}_1 , a única mudança na determinação de \mathbf{d}_2 é a condição adicional $\gamma(\mathbf{d}_1, \mathbf{d}_2) = 0$, a qual deve ser incorporada ao processo de otimização. Os demais detectores são determinados de maneira similar.

O detector \mathbf{d}_1 deve satisfazer um número menor de restrições do que os demais detectores e, por conseqüência, níveis maiores de alisamento deverão ser alcançados por ele. Assim, como \mathbf{d}_1 terá uma capacidade maior de filtrar imagens de entrada do que os demais detectores, esse deverá ser o primeiro a ser aplicado durante o processo de detecção. Da mesma forma, \mathbf{d}_2 será mais alisado do que \mathbf{d}_3 e deve ser aplicado nas imagens que passarem pelo limiar imposto por \mathbf{d}_1 . A aplicação dos demais detectores deve seguir o mesmo procedimento.

3.3.1 Detectores subótimos

Embora os detectores sejam determinados *a priori* do processo de detecção (*off-line*), pode ser desejável usar um algoritmo mais rápido. Dessa forma, a função objetivo em (3.9) a ser minimizada pode ser substituída por

$$\sum_{\mathbf{t}_j \in \mathbb{T}} \langle \mathbf{d}_1, \mathbf{t}_j \rangle^2 + \lambda S(\mathbf{d}_1), \quad (3.11)$$

que pode ser otimizada, conforme descrito para a determinação dos detectores ótimos, avaliando-se diferentes valores de λ definidos através de bissecção. Apesar de não ótima, a função objetivo é agora quadrática e os detectores podem ser encontrados via decomposição por valores singulares, SVD (ver Apêndice 4). Resultados experimentais indicam que, tipicamente, se N detectores ótimos atingem uma determinada taxa de detecção, então $1,3N$ detectores subótimos são necessários para atingir a mesma taxa [43].

Exemplos de imagens de treinamento, de dimensão (20×20) *pixels*, assim como de seis detectores, são apresentados na Figura 3.1.

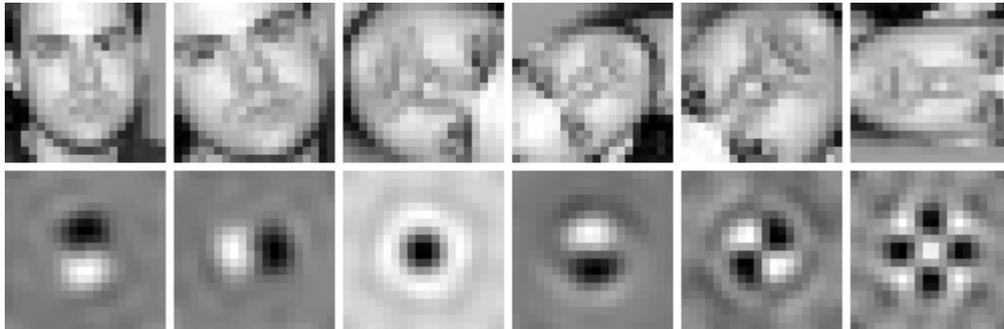


Figura 3.1 — Exemplos de imagens de treinamento referentes ao multipadrão \mathbb{T} (linha superior) e exemplos de detectores (linha inferior).

3.4 Considerações

Em [43], não é apresentada a determinação dos detectores subótimos, embora conste a sugestão do uso de SVD para tal. Uma abordagem para a obtenção de detectores subótimos é proposta no Apêndice 5.

Resultados e Considerações

Diversas situações de teste são exploradas a seguir visando analisar o comportamento das duas abordagens estudadas nos Capítulos 2 e 3, frente a diferentes tipos de aplicações. São elas: detecção de uma face específica vista de um ângulo frontal; detecção de características faciais; detecção de objetos tridimensionais; e, por fim, uma análise da sensibilidade de cada um dos métodos ao ruído aditivo e distorções no foco. Todas as avaliações realizadas, com respeito ao método de Antifaces, utilizam detectores subótimos, determinados de acordo com o Apêndice 5.

Dois termos que caracterizam a natureza das imagens serão empregados no decorrer deste capítulo: imagens sintéticas e imagens naturais. As imagens sintéticas são imagens que sofreram algum tipo de distorção intencional ou edição; as naturais, aquelas que preservam todas as características originais, sem passar por qualquer tipo de edição.

4.1 Condições de teste e treinamento

Os diferentes testes são aplicados tanto para o método de Autofaces (Seção 4.2) quanto para o método de Antifaces (Seção 4.3), conforme as condições definidas nesta seção.

4.1.1 Detecção de uma face frontal específica

Neste teste, é considerada a detecção de uma face específica. Vistos os conceitos gerais apresentados no Capítulo 1, esse tipo de aplicação envolve tanto a detecção facial, uma vez que localiza uma subimagem contendo uma face dentro de uma imagem maior, quanto o reconhecimento facial, uma vez que é feita a identificação da face detectada.

Um conjunto de imagens de treinamento é construído a partir de uma única imagem. Tendo-se uma imagem frontal da face a ser detectada, são geradas 14580 imagens de dimensão (20×20) *pixels*, através de operações de rotação e escalamento aplicadas à imagem original. As rotações são efetuadas para ângulos entre zero e 2π radianos, em intervalos de $\pi/90$ radianos. Para cada ângulo rotacionado, a imagem é escalada nas direções dos eixos horizontal e vertical, independentemente, em escalas de 0,8 a 1,2, amostradas com passos de 0,05.

A avaliação da detecção é realizada sobre uma imagem sintética. Essa imagem é composta por 100 faces dispostas em 10 colunas contendo 10 faces cada. Dentre essas 100 faces, cinco correspondem à face que se deseja detectar. Essas cinco faces a serem detectadas não pertencem ao conjunto de treinamento, porém são construídas de forma semelhante. A partir da mesma imagem original, são efetuadas rotações e escalamentos em intervalos intermediários aos definidos para a etapa de treinamento. As 95 faces que não são de interesse à detecção pertencem ao banco de dados do Departamento de Psicologia da Universidade de Stirling².

Imagens provenientes de diferentes bancos de dados são adquiridas sob diferentes condições de iluminação e digitalização. Buscando-se avaliar o comportamento dos métodos frente a essas diferenças, são realizados testes em que a face de interesse pertence ao mesmo banco de dados das demais faces, e testes em que as faces são provenientes de bancos diferentes.

Os conjuntos de treinamento e teste descritos são montados nos moldes dos testes apresentados em [43]. Nas avaliações desenvolvidas sobre esses conjuntos, as imagens de teste e treinamento são provenientes da mesma imagem original, o que certamente não acontece em aplicações que requerem a tarefa de detecção automática de faces. Visando reproduzir de forma mais fiel esse tipo de aplicação, um outro teste de detecção de faces frontais é realizado. Dessa vez, a imagem de teste, que consiste em uma cena contendo

² PICS — *Psychological Images Collection at Stirling, University of Stirling* (<http://pics.psych.stir.ca.uk/>)

várias faces, é natural. O conjunto de treinamento é gerado a partir de uma única imagem, como nos testes anteriores. Entretanto, a imagem que gera os 14580 padrões de treinamento relativos à face que se deseja detectar possui dimensão de (60×60) *pixels* e não é a mesma imagem de teste. Ambas podem ser vistas nas Figuras 4.1 e 4.2.



Figura 4.1 — Imagem de treinamento.



Figura 4.2 — Imagem natural para teste de detecção de uma face frontal específica (tamanho reduzido).

4.1.2 Detecção de características faciais

Este teste consiste na detecção de um dos olhos (olho esquerdo) de determinada pessoa. A detecção é feita sobre uma imagem formada por 20 imagens diferentes da mesma pessoa, dispostas em cinco colunas contendo quatro imagens cada (Figura 4.3).

Três conjuntos diferentes de imagens de treinamento são avaliados. Dois deles utilizam a imagem do olho esquerdo da primeira face (na posição superior esquerda) da

imagem de teste (Figura 4.3). O terceiro utiliza os olhos esquerdos das 10 primeiras faces (linhas superiores).

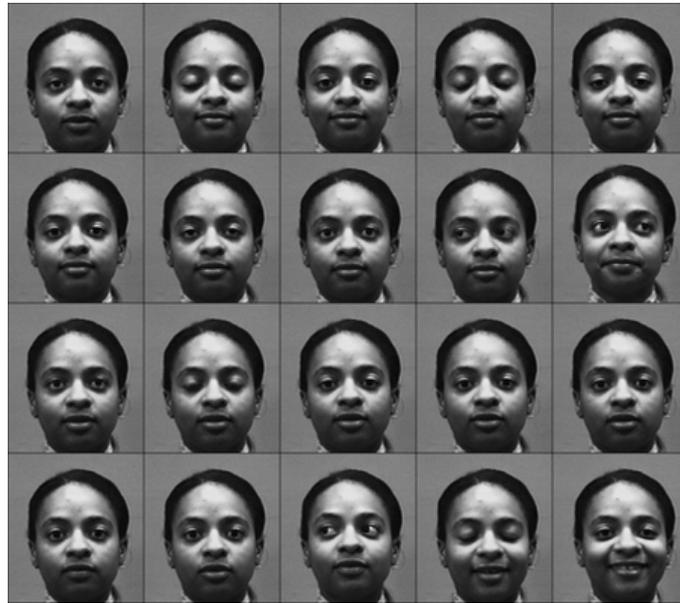


Figura 4.3 — Imagem de teste usada para a detecção de características faciais.

No primeiro conjunto de treinamento, a imagem original (olho esquerdo da primeira face) é rotacionada em ângulos entre -15° e $+15^\circ$, em passos de 2° . Para cada ângulo rotacionado, a imagem é escalada nas direções dos eixos horizontal e vertical, independentemente uma da outra, em escalas de 0,8 a 1,2, amostradas com passos de 0,05. Essas distorções realizadas sobre a imagem original geram 1296 imagens de dimensão (20×20) *pixels*.

O segundo conjunto de treinamento é montado exatamente da mesma forma que o primeiro, porém desconsiderando as rotações. Assim, as imagens que compõem esse conjunto consistem apenas de escalamentos da imagem original (olho esquerdo da primeira face) nas direções dos eixos horizontal e vertical, independentemente uma da outra, em escalas de 0,8 a 1,2, amostradas com passos de 0,05. O tamanho resultante desse conjunto de treinamento é de 81 imagens de dimensão (20×20) .

(passos angulares de $n10^\circ + 5^\circ$, $n = 0, 1, \dots, 35$) são reservadas para uso nos testes de detecção.

Uma imagem de teste é montada contendo 25 objetos, dispostos em cinco colunas (Figura 4.4). Desses objetos, cinco são de interesse à detecção, sendo que o primeiro (posição superior esquerda) pertence ao conjunto de treinamento e os outros quatro pertencem às 36 imagens reservadas para os testes de detecção.

4.1.4 Análise de robustez

A sensibilidade a determinados tipos de distorções na imagem é analisada em cada um dos dois métodos. Para tal, testes com inserção de ruído branco Gaussiano, alterações no foco e alterações da cena em que se processa a detecção de objetos são realizados.

A medida usada para avaliar o nível de ruído presente em uma imagem é a razão sinal-ruído de pico (PSNR), definida como

$$\text{PSNR [dB]} = 10 \log \left(\frac{V_{\max}^2}{\text{EQM}} \right), \quad (4.1)$$

onde V_{\max} é o máximo valor assumido por um elemento da imagem (*pixel*) e EQM é o erro quadrático médio entre a imagem original e a imagem afetada pelo ruído. O EQM é definido como

$$\text{EQM} = \frac{1}{MN} \sum_{l=1}^M \sum_{c=1}^N [\mathbf{I}_0(l, c) - \mathbf{I}_r(l, c)]^2, \quad (4.2)$$

onde \mathbf{I}_0 é a imagem original e \mathbf{I}_r é sua versão corrompida por ruído, ambas de mesma dimensão, $(M \times N)$.

As avaliações de inserção de ruído e alterações no foco são realizadas sobre a imagem de teste usada para a detecção de características faciais. Tanto o método de Autofaces quanto o de Antifaces são treinados, para essas avaliações, através de um conjunto de treinamento composto por 81 imagens (ver Seção 4.1.2). Primeiramente, um

ruído branco Gaussiano é adicionado à imagem de teste em níveis gradativos até que o comportamento do algoritmo seja alterado. Em outro teste, um filtro passa-baixas (PB) de média móvel é aplicado sobre a imagem de teste, simulando alterações no foco. Resultados obtidos com resposta ao impulso do filtro (janela) de diversas dimensões são considerados.

As avaliações de robustez frente às alterações da cena são realizadas considerando-se a detecção de objetos tridimensionais. Ambos os métodos são treinados usando-se o conjunto de treinamento discutido na Seção 4.1.3 (no qual as imagens apresentam fundo preto) e aplicados sobre uma imagem de teste criada especificamente para essa avaliação. Em [43], são apresentados alguns exemplos de detecção de objetos tridimensionais, em que a imagem de teste consiste em uma cena natural onde uma imagem menor — um padrão semelhante aos do conjunto de treinamento (com fundo preto) — é sobreposta. Em uma imagem construída dessa forma, analisando-se apenas a vizinhança do objeto de interesse, ou seja, uma janela ao redor do objeto, a cena mantém-se preta. Uma situação mais realista é a presença, na imagem de teste, do objeto de interesse desprovido do fundo utilizado no momento do treinamento. Ambos os casos são mostrados na Figura 4.5, que apresenta a imagem de teste criada para essa última avaliação envolvendo análise de robustez.



Figura 4.5 — Imagem de teste sintética em que uma cena é sobreposta por imagens do objeto de interesse com fundo preservado e com fundo transparente.

4.2 Resultados obtidos pelo método de Autofaces

Nesta seção, são apresentados os resultados obtidos através do uso do método de Autofaces, seguindo-se as condições de teste definidas na Seção 4.1.

4.2.1 Detecção de faces frontais

Cinco avaliações de detecção facial são apresentadas. Apenas nas três primeiras a face a ser detectada pertence ao mesmo banco de imagens das demais presentes na cena.

Na primeira avaliação um espaço de faces composto por 50 Autofaces é necessário para que a detecção exata seja alcançada. O resultado pode ser visto na Figura 4.6(a). Como foi discutido no Capítulo 2, uma redução da dimensão do espaço de faces acarreta perda de informação na reconstrução das imagens, podendo levar a alarmes falsos durante o processo de detecção. Reduzindo-se o espaço de faces de 50 para 30 Autofaces, gerando dessa forma alguns alarmes falso-positivos. Como pode ser observado na Figura 4.6(b), esses alarmes ocorrem cada um exatamente sobre uma única face (não sobrepostos a duas faces). Esse resultado demonstra a capacidade de generalização do método.

Nas quatro avaliações seguintes, um espaço de faces composto por 20 Autofaces é suficiente para atingir um resultado isento de alarmes falsos, assim como em todos eles um mesmo limiar de detecção (DFFS) é utilizado. Tais resultados são apresentados na Figura 4.7. Resultados de detecção exata foram também alcançados através do uso de um espaço de faces com menor dimensão. Porém, para dimensões inferiores a 20, o limiar de detecção precisa ser cuidadosamente escolhido, não sendo mais possível empregar o mesmo limiar para todas as quatro avaliações. Dessa forma, o uso de 20 Autofaces fornece um comportamento mais estável do algoritmo frente a essas condições de treinamento e teste.

Entre a primeira avaliação e as demais, é observada uma discrepância na dimensão do espaço de faces necessária para a detecção exata. Embora todas as faces sejam aparentemente similares, a análise dos autovalores associados às Autofaces que compõem o espaço de faces demonstra que na primeira avaliação o conjunto de treinamento apresenta uma complexidade maior do que nas demais (as imagens de treinamento são mais heterogêneas). Uma maior complexidade do conjunto de treinamento justifica a necessidade de um número maior de Autofaces compondo o espaço de faces.

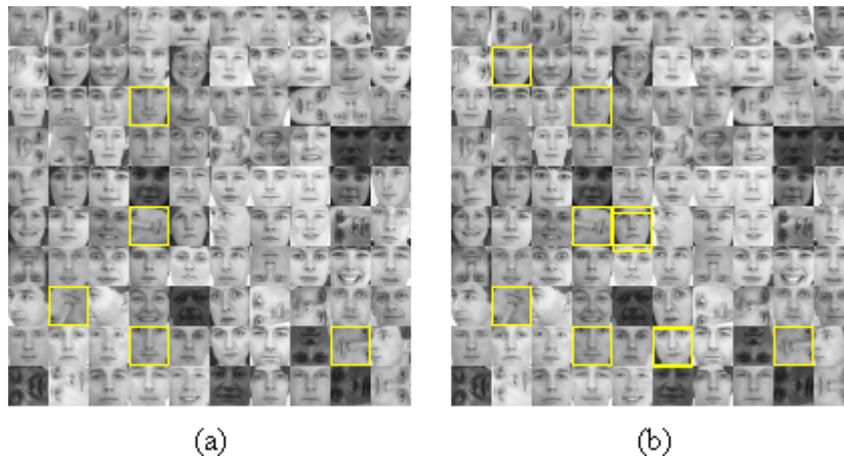


Figura 4.6 — Resultado de detecção de uma face frontal específica: (a) espaço com 50 Autofaces e (b) com 30 Autofaces.

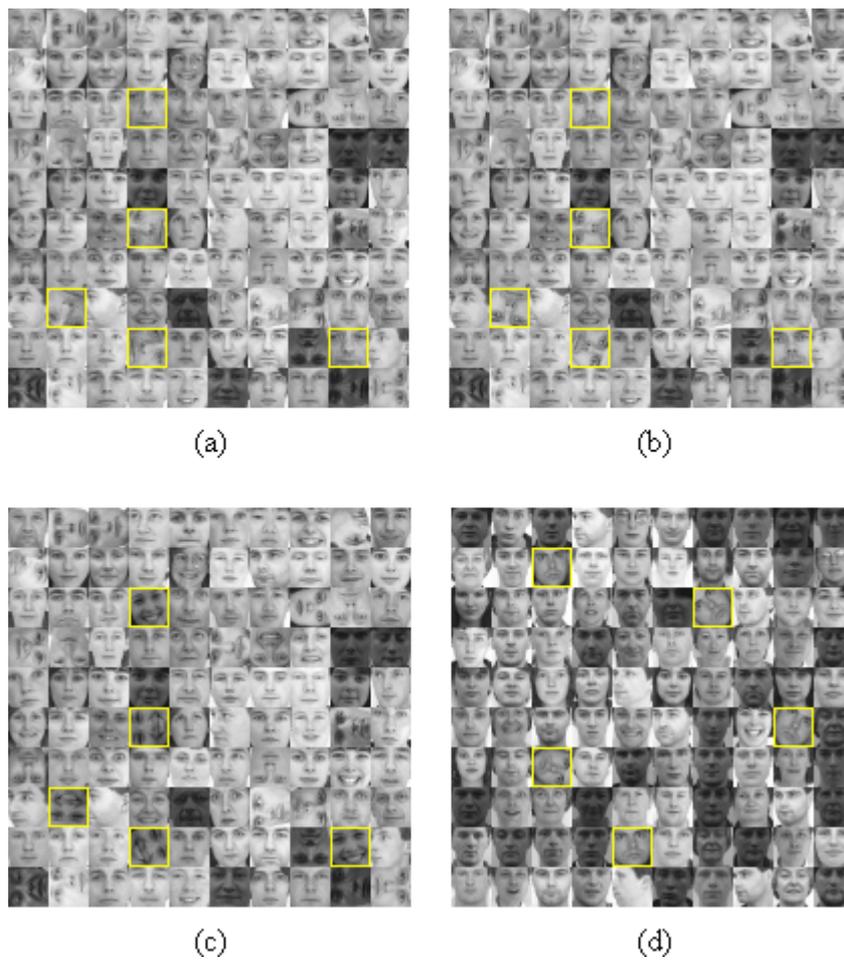


Figura 4.7 — Resultados da detecção de uma face frontal específica usando 20 Autofaces.

Usando-se as Autofaces para o caso de detecção de uma face específica em uma imagem de teste natural, chega-se à detecção exata apresentada na Figura 4.8.



Figura 4.8 — Resultado da detecção de uma face específica em uma imagem natural.

4.2.2 Detecção de características faciais

Três avaliações referentes à detecção de características faciais são apresentadas. Os conjuntos de treinamento são compostos por 1296, 81 e 10 imagens, respectivamente. No primeiro caso, 15 olhos, além do olho pertencente à face na posição superior esquerda (o qual compõe o conjunto de treinamento), são detectados. Apenas os olhos que aparecem fechados são rejeitados (Figura 4.9). Para o conjunto de treinamento de 81 olhos, o resultado é apresentado na Figura 4.10. Mesmo alguns dos olhos que aparecem fechados são detectados. Por fim, para o conjunto de treinamento composto por 10 olhos, todos os olhos esquerdos são detectados. O método de Autofaces é, nesse caso, capaz de abstrair até mesmo algumas imagens de olhos direitos (Figura 4.11). Nas três avaliações, apenas duas Autofaces são utilizadas para compor o espaço de faces. A concentração relativamente alta de alarmes ao redor de cada um dos olhos detectados, existente em ambas as imagens, demonstra que o método é capaz de detectar a presença de um padrão mesmo que esse não esteja perfeitamente centrado na subimagem analisada.

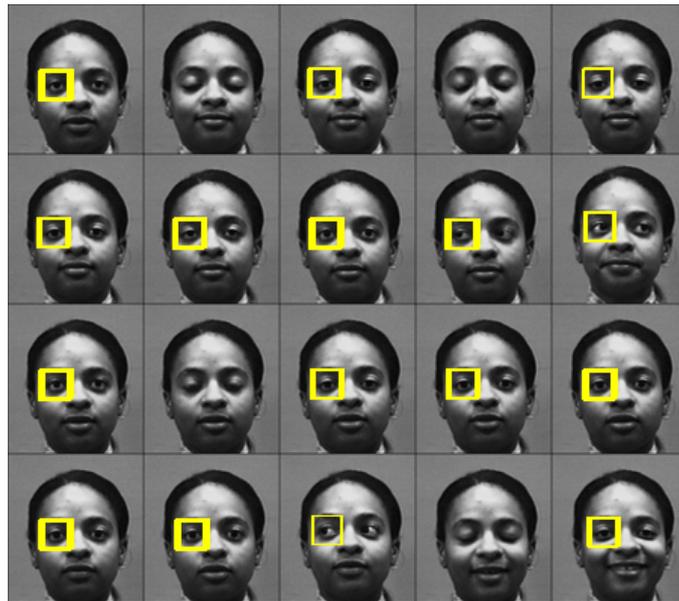


Figura 4.9 — Resultado da detecção de olhos para um conjunto de treinamento com 1296 imagens.

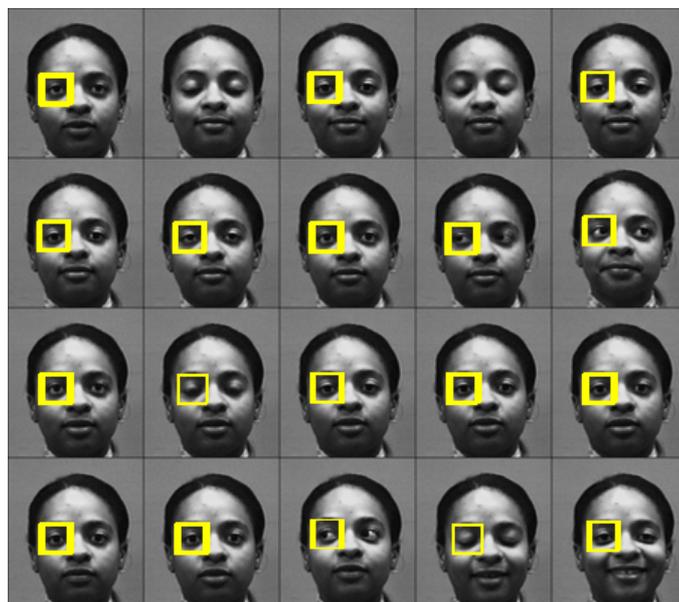


Figura 4.10 — Resultado da detecção de olhos para um conjunto de treinamento com 81 imagens.

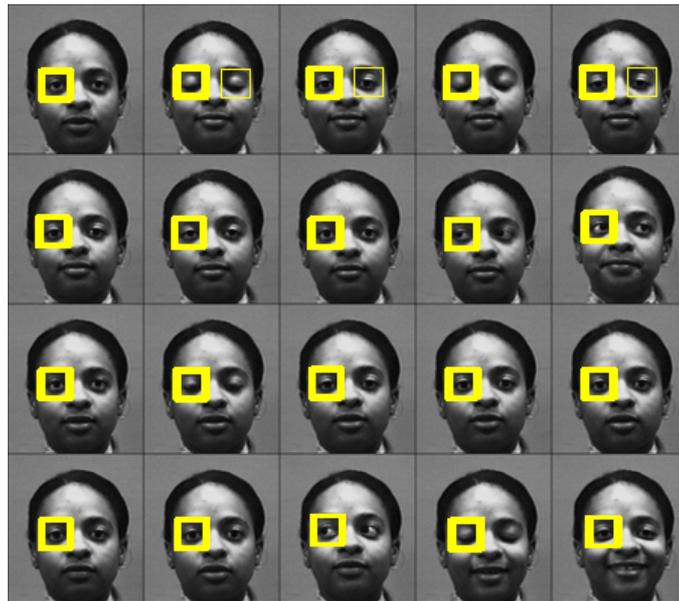


Figura 4.11 — Resultado da detecção de olhos para conjunto de treinamento com 10 imagens de olhos (olho esquerdo das primeiras 10 imagens).

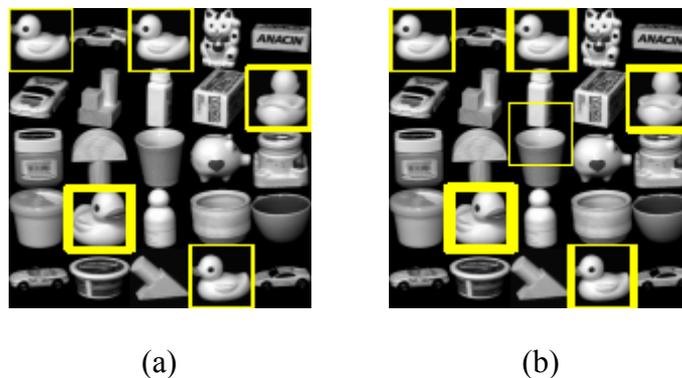


Figura 4.12 — Detecção, pelo método de Autofaces, de objetos tridimensionais: (a) exata e (b) com alarme falso-positivo, gerado por um limiar de detecção menos rigoroso.

4.2.3 Detecção de objetos tridimensionais

Na avaliação da detecção de objetos tridimensionais, a detecção sem alarmes falsos é alcançada (Figura 4.12(a)). Diferente do caso de detecção facial, o objeto de interesse à detecção é bastante diferente dos demais objetos que compõem a imagem de teste. Essa característica facilita o processo de detecção. Entretanto, em razão da alta complexidade (diferenças morfológicas) do conjunto de treinamento, pequenas variações no limiar de

detecção (DFFS) podem acarretar em alarmes falso-positivos (Figura 4.12(b)). Ainda assim, é importante notar que o único alarme falso ocorre em uma região da imagem cuja forma se assemelha ao objeto procurado. Nesse caso, o objeto detectado apresenta base larga (correspondendo ao corpo) e parte superior alongada (correspondendo ao pescoço e à cabeça).

4.2.4 Análise de robustez

A inserção de ruído branco na imagem de teste acarreta a ocorrência de alarmes falso-negativos. Os alarmes falsos (sobre olhos originalmente detectados, vide Figura 4.10) iniciam para níveis de PSNR inferiores a 28 dB (Figura 4.13). Para níveis de PSNR até 22 dB, o método ainda apresenta resultados satisfatórios (Figura 4.14). Abaixo desse valor, padrões semelhantes ao utilizado no treinamento — olho aberto e direcionado para frente — não são detectados (Figura 4.15).

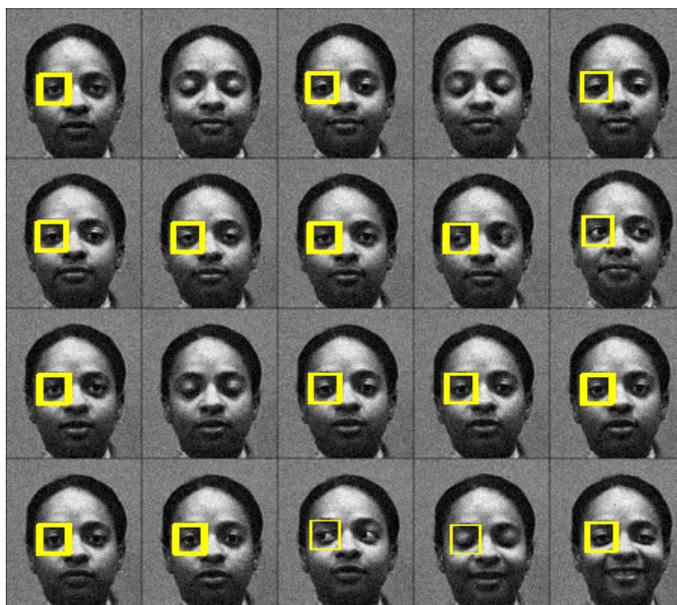


Figura 4.13 — Ocorrência de alarme falso-negativo na detecção. Imagem submetida a ruído branco aditivo; PSNR = 28 dB.

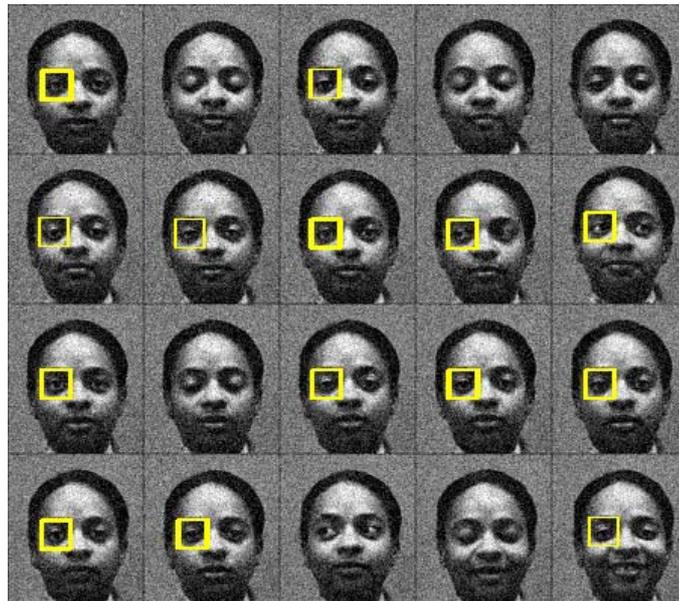


Figura 4.14 — Resultado da detecção para uma imagem submetida a ruído branco aditivo;
 $PNSR = 22$ dB.

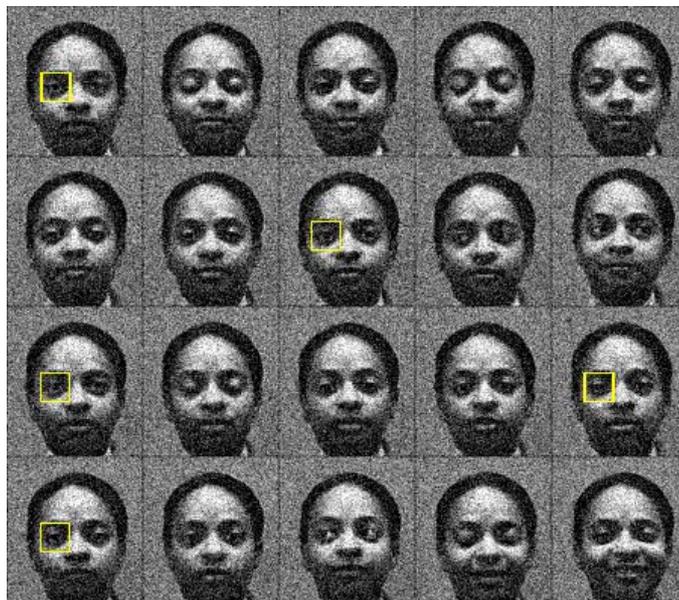


Figura 4.15 — Resultado da detecção para uma imagem submetida a ruído branco aditivo;
 $PNSR = 19$ dB.

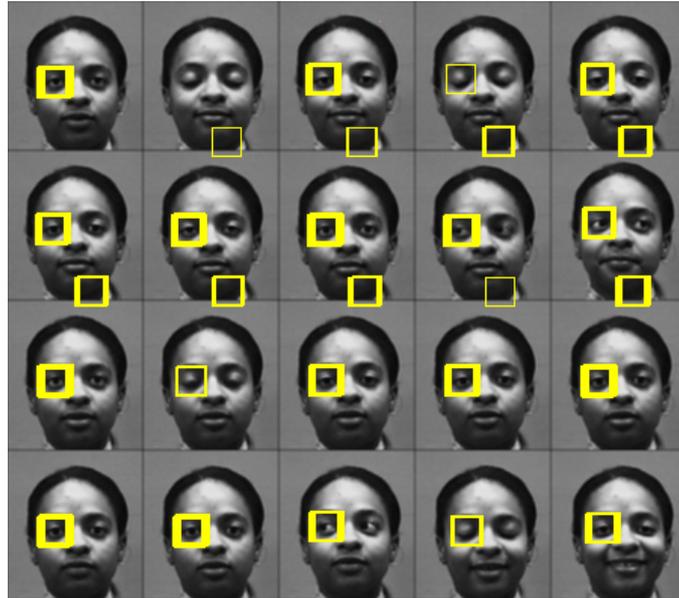


Figura 4.16 — Resultado da detecção em que alarmes falso-positivos são gerados em razão da distorção no foco. Imagem submetida a um filtro PB com resposta ao impulso de dimensão (2×2) .

As distorções no foco ocasionam alarmes falso-positivos sobre a imagem. Um exemplo desse resultado pode ser visto na Figura 4.16, onde a imagem de teste é submetida a um filtro PB média-móvel com resposta ao impulso de dimensão (2×2) .

Por fim, é feita a avaliação da sensibilidade do algoritmo frente a variações do fundo da imagem. Ambos os padrões são detectados corretamente (Figura 4.17), com o uso de apenas duas Autofaces no espaço de faces. Verifica-se a robustez do método mesmo frente às restrições de baixo desempenho para conjuntos de treinamento compostos por imagens morfologicamente bastante diferentes e frente às observações ao redor da influência da cena sobre o processo de detecção, descritas em [41].

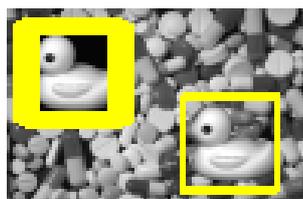


Figura 4.17 — Resultado da detecção pelo método de Autofaces.

4.2.5 Considerações gerais sobre o método de Autofaces

Comparando o desempenho das Autofaces nas avaliações de detecção de faces frontais e de características faciais, é possível notar uma considerável melhoria quando utilizados padrões de treinamento homogêneos (segundo caso). A homogeneidade dos padrões pode ser avaliada através da energia associada aos autovalores da matriz de covariância do conjunto de treinamento. Quando grande parte da energia está concentrada em poucos autovalores, o método apresenta um melhor desempenho. Assim, melhores resultados da detecção de faces frontais podem ser obtidos ao desconsiderar-se a possibilidade de rotação das faces no conjunto de treinamento e nas imagens de teste.

4.3 Resultados obtidos pelo método de Antifaces

Nesta seção, são apresentados os resultados obtidos através do uso do método de Antifaces, seguindo as condições de teste definidas na Seção 4.1.

4.3.1 Detecção de faces frontais

Cinco avaliações da detecção de faces frontais são consideradas. Como no caso das Autofaces, apenas nos três primeiros testes a face a ser detectada pertence ao mesmo banco de imagens das demais faces. Nas duas primeiras avaliações, a detecção exata é alcançada com o uso de 15 e 16 detectores, respectivamente (Figura 4.18). Na terceira avaliação, a detecção exata não é alcançada mesmo para o uso de 100 detectores (Figura 4.19). Diminuindo-se o limiar M , é possível eliminar o alarme falso-positivo presente. Porém, nesse caso, alarmes falso-negativos são também gerados. Nas duas últimas avaliações, o resultado sem alarmes falsos é alcançado com 13 e 7 detectores, respectivamente (Figura 4.20). A presença de apenas um alarme falso na Figura 4.19 demonstra que, apesar da detecção inexata, o algoritmo apresenta um comportamento consistente.

Tratando-se da avaliação de detecção de uma face específica em uma imagem natural, o método de Antifaces não foi capaz de alcançar um resultado satisfatório. A face de interesse não foi detectada.

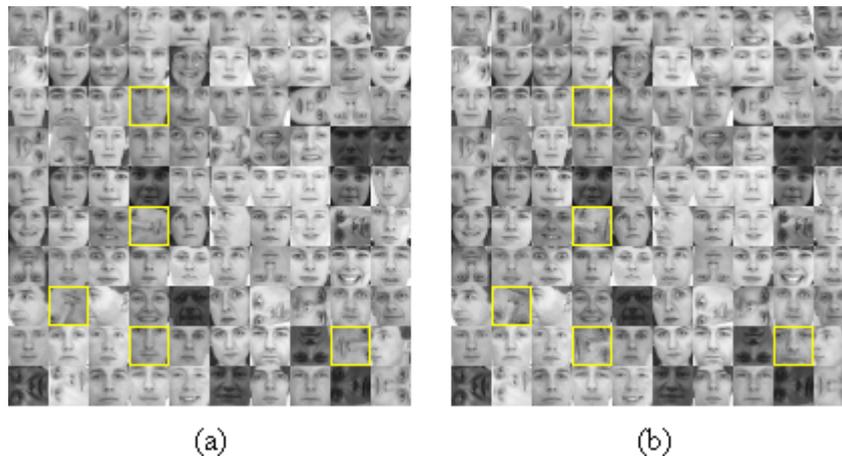


Figura 4.18 — Resultados de detecção de uma face frontal específica. Para uma detecção sem alarmes falsos, foram necessários: (a) 15 detectores e (b) 16 detectores.



Figura 4.19 — Resultado de detecção inexata (100 detectores foram aplicados).

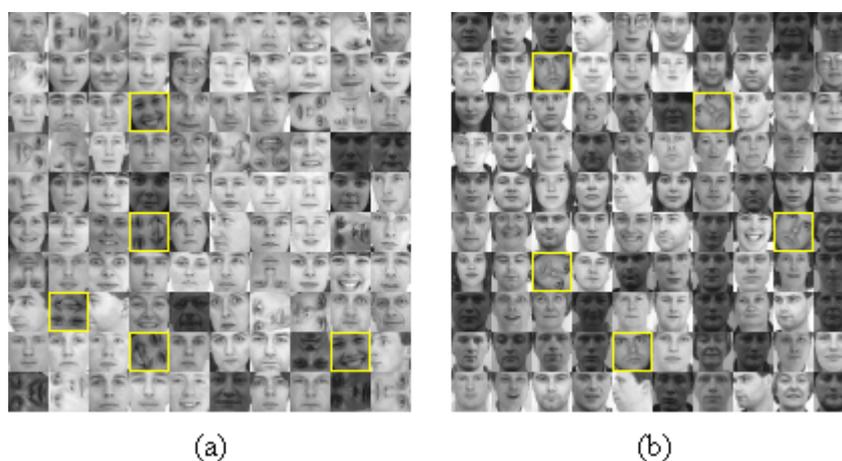


Figura 4.20 — Resultados de detecção de uma face frontal específica. Para uma detecção sem alarmes falsos, foram necessários: (a) 13 detectores e (b) 7 detectores.

4.3.2 Detecção de características faciais

Três avaliações referentes à detecção de características faciais são apresentadas. Os conjuntos de treinamento são compostos por 1296, 81 e 10 imagens, respectivamente.

Utilizando-se um conjunto de 1296 imagens de treinamento, o melhor resultado alcançado é apresentado na Figura 4.21(a). A partir desse resultado, para uma pequena variação no limiar M , na intenção de obter uma detecção menos rigorosa, vários alarmes falsos são gerados sem que todos os padrões de interesse tenham sido detectados (Figura 4.21(b)).

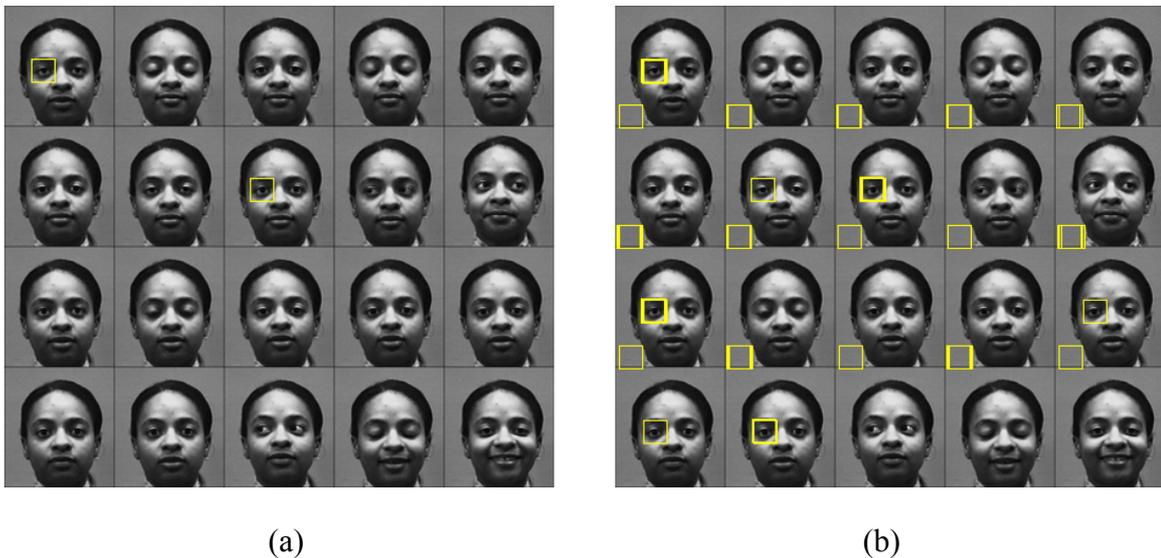


Figura 4.21 — Resultados da detecção de olhos para um conjunto de treinamento com 1296 imagens: (a) melhor resultado obtido sem alarmes falso-positivo; (b) presença de alarmes falso-positivo sem a detecção de todos os olhos.

Para um conjunto de treinamento composto por 81 imagens, uma maior capacidade de generalização é observada. Onze padrões são detectados corretamente (Figura 4.22), lembrando de que o primeiro deles (pertencente à face na posição superior esquerda da imagem) compõe o conjunto de treinamento.

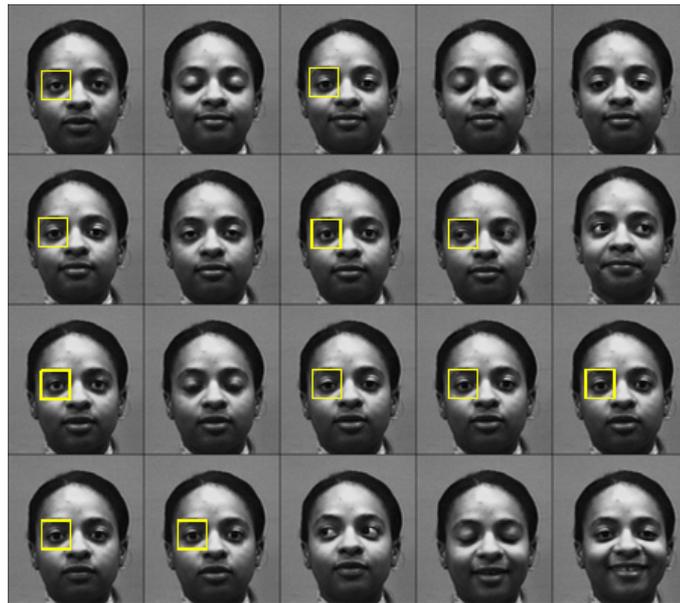


Figura 4.22 — Resultado da detecção de olhos para um conjunto de treinamento composto por 81 imagens.

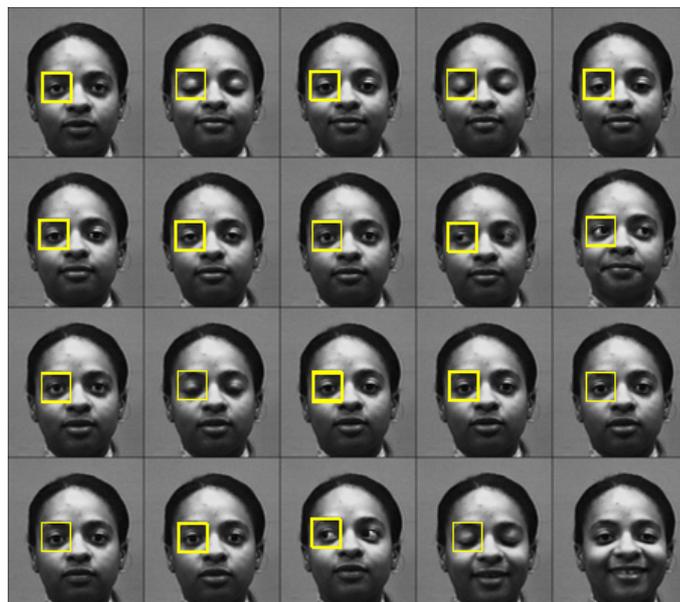


Figura 4.23 — Resultado da detecção de olhos, usando Antifaces, para um conjunto de treinamento composto por 10 imagens naturais (10 primeiros olhos esquerdos).

Para o último caso, em que o conjunto de treinamento é composto por 10 imagens (olhos esquerdos das 10 primeiras faces da imagem de teste), os resultados podem ser

considerados menos satisfatórios do que os da avaliação anterior. Tal consideração é feita uma vez que o alarme falso-negativo gerado é referente a uma imagem bastante semelhante a imagens usadas no treinamento. Além dos 10 olhos pertencentes ao conjunto de treinamento, nove outros são detectados (Figura 4.23).

4.3.3 Detecção de objetos tridimensionais

Na avaliação da detecção de objetos tridimensionais, o uso de apenas três detectores se faz necessário para alcançar a detecção sem alarmes falsos (Figura 4.24(a)). Na imagem da Figura 4.24(b), é observado outro caso em que apenas a primeira ocorrência do objeto é detectada, sendo essa a única pertencente ao conjunto de imagens de treinamento. Esse fato decorre da sensibilidade do algoritmo ao limiar M escolhido (máximo valor absoluto do produto interno entre o detector e as imagens do conjunto de treinamento). Na Tabela 4.1, são apresentados os valores das grandezas que envolvem as detecções realizadas para os dois casos. Diretamente dessa tabela, é possível verificar a sensibilidade do algoritmo aos valores de projeto. No caso de detecção com falso-negativos, Figura 4.24(b), as condições de alisamento e do limiar M especificadas em [43] não são suficientes para a detecção exata.

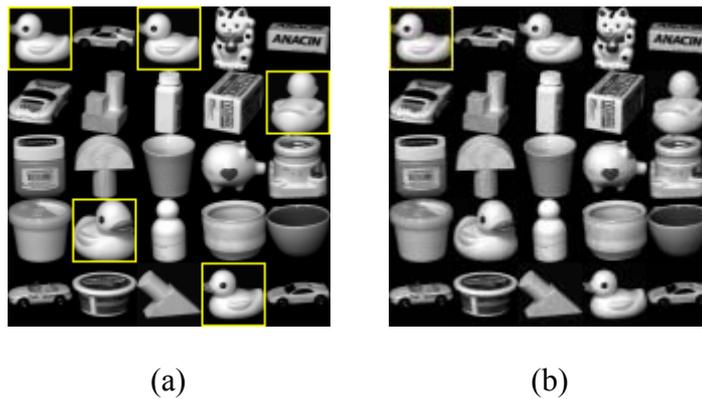


Figura 4.24 — Resultados da detecção de objetos tridimensionais: (a) detecção exata; (b) detecção inexata.

Tabela 4.1 — Valores do alisamento e do limiar M para os casos de (a) detecção exata e (b) detecção com alarmes falso-negativos

Detector	Alisamento	M	
1	5,1	$6,0 \cdot 10^{-3}$	Caso (a)
2	8,7	$5,8 \cdot 10^{-3}$	
3	9,5	$8,8 \cdot 10^{-3}$	
1	6,8	$3,5 \cdot 10^{-4}$	Caso (b)
2	8,4	$3,2 \cdot 10^{-4}$	

Através da detecção exata, apresentada na Figura 4.24(a), para a qual apenas um alarme é gerado sobre cada imagem do objeto de interesse, é possível constatar a baixa capacidade de generalização do método de Antifaces. Nesse caso, pode ser observado que, deslocando-se a janela de teste em apenas um *pixel* para algum dos lados a partir do centro de qualquer um dos objetos de interesse, o sistema já passa a classificá-lo como desconhecido.

4.3.4 Análise de robustez

Como é possível observar na Figura 4.25, a inserção de ruído branco acarreta ocorrência de alarmes falso-positivos. Tais alarmes são gerados para valores de PSNR inferiores a 38 dB. Para PSNR = 28 dB, além de diversos alarmes falso-positivos, a ocorrência de um alarme falso-negativo é observada (Figura 4.26).

Submetendo-se a imagem a distorções no foco (via filtro PB), os resultados da detecção, obtidos pelo método de Antifaces, são bastante diversos. Em função do maior alisamento da imagem causado por esse tipo de filtro, o número de alarmes, na média das avaliações, diminui. Esse resultado concorda com a teoria descrita na Seção 3.2.1. Nas Figuras 4.27 e 4.28, são apresentados os resultados das aplicações de filtros usando respostas ao impulso de dimensões (2×2) e (8×8) , respectivamente.

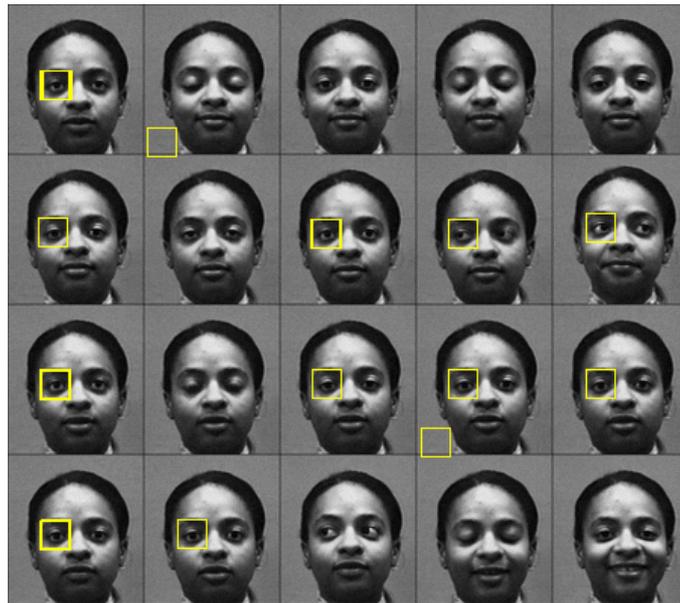


Figura 4.25 — Presença de alarmes falsos na detecção por Antifaces. Imagem de teste submetida a ruído branco aditivo; PSNR = 38 dB.

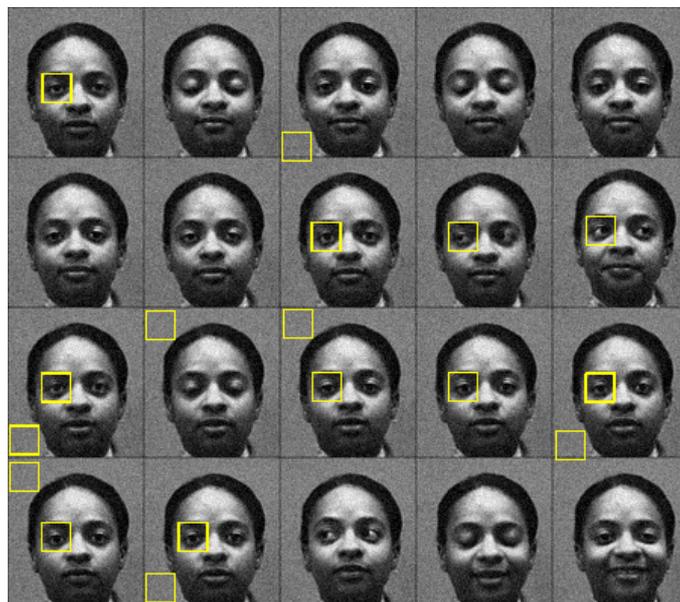


Figura 4.26 — Presença de alarmes falsos na detecção por Antifaces. Imagem de teste submetida a ruído branco aditivo; PSNR = 28 dB.

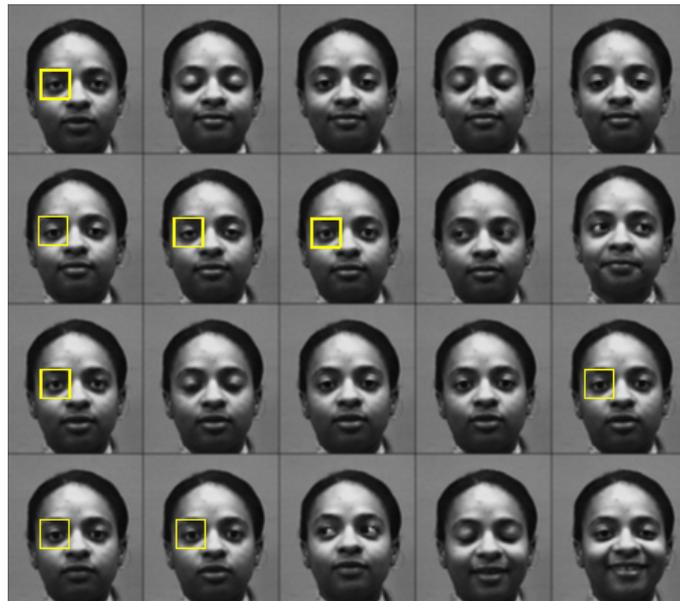


Figura 4.27 — Resultado da detecção por Antifaces em que houve uma redução no número de alarmes; imagem submetida a um filtro PB, com resposta ao impulso de dimensão (2×2) .

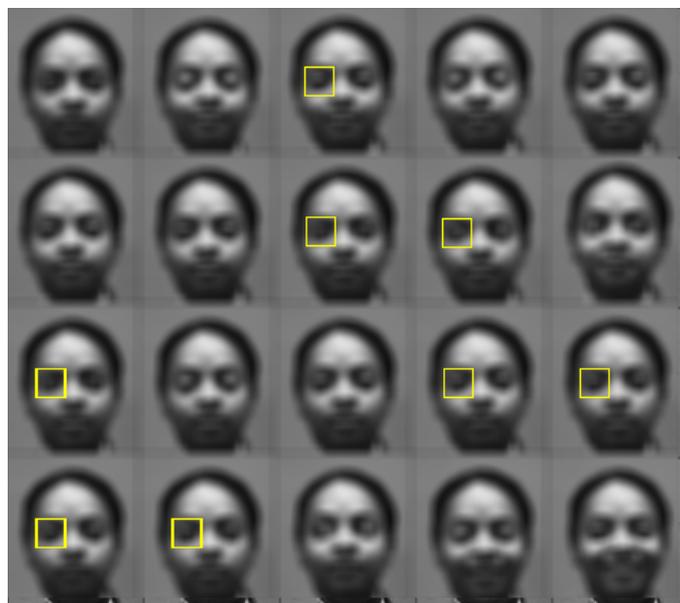


Figura 4.28 — Resultado da detecção por Antifaces; imagem submetida a um filtro PB, com resposta ao impulso de dimensão (8×8) .

Frente à avaliação de sensibilidade do algoritmo com relação às variações do fundo da imagem, as Antifaces são capazes de detectar apenas o padrão com fundo preto (Figura 4.29(a)). É importante observar que uma pequena modificação do limiar de detecção acarreta ocorrência de alarmes falso-positivos (Figura 4.29(b)). A presença de alarmes falso-positivos e falso-negativos na mesma imagem demonstra que a escolha de um novo limiar não é suficiente para atingir uma detecção sem alarmes falsos.

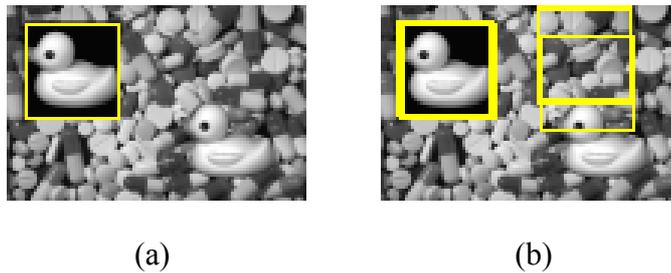


Figura 4.29 — Resultados da detecção pelo método de Antifaces: (a) detecção com alarme falso-negativo; (b) detecção com alarmes falso-positivos.

4.3.5 Considerações gerais sobre o método de Antifaces

Para as avaliações da detecção de faces frontais, o menor número de detectores necessários à detecção exata foi sete. Esse caso refere-se ao teste que envolve a detecção de imagens de uma face não pertencente ao mesmo banco de imagens das demais presentes na cena. Em uma situação de teste similar, a detecção sem alarmes falsos é alcançada com apenas quatro detectores em [43]. Com isso, é possível notar a influência das condições de aquisição das imagens utilizadas frente ao desempenho do método, ou seja, a relação entre o número de detectores necessários para se obter a detecção exata e a semelhança entre as imagens.

O cálculo dos detectores considera, como foi visto no Capítulo 3, um compromisso entre o alisamento (que permite rejeitar as imagens que não se deseja detectar), a minimização do produto interno entre o detector e as imagens do padrão de interesse, e a independência entre os detectores. Dessa forma, superestimando algum desses parâmetros, os demais são prejudicados. Como discutido em [43], o alisamento dos detectores deve ser tão pequeno quanto possível. Assim, é necessário inicialmente que sejam definidos os

valores dos demais parâmetros, sendo o alisamento o último a ser determinado. Durante este trabalho, foi constatado que o grau de independência necessário entre os detectores não é bem definido. Em alguns casos, um determinado grau de independência é suficiente para atingir a detecção sem alarmes falsos, fazendo-se uso de poucos detectores. Em outros, entretanto, a mesma independência acarreta necessidade do uso de um número de detectores quatro vezes maior, para alcançar um resultado sem alarmes falsos. Geralmente, havendo necessidade do uso de muitos detectores, o aumento da independência é suficiente para reduzir esse número em até quatro vezes. Quantitativamente, a independência necessária entre os detectores, definida por (3.7), tem variado durante os testes na faixa de ordens de grandeza entre 10^{-6} e 10^{-15} . Quanto ao valor absoluto (máximo) entre o detector e as imagens do conjunto de treinamento, diferente do que foi sugerido em [43], é constatado que se deve manter na ordem de 10^{-2} , não importando o tamanho ou complexidade do conjunto de treinamento.

4.4 Considerações gerais

Em uma imagem, a procura por características faciais, no lugar da procura por faces, trouxe um ganho no sentido de que padrões menos complexos e mais uniformes contribuem para a capacidade de abstração de ambos os métodos. Os resultados obtidos através do método de Autofaces e de Antifaces são comparados, para esse caso, na Tabela 4.2.

Tabela 4.2 — Resultados da detecção de características faciais

	Autofaces			Antifaces		
Imagens de treinamento	1296	81	10	1296	81	10
Autofaces / Antifaces	2	2	2	15	20	20
Taxa de alarmes falso-negativos	0,2	0,1	0,0	0,95	0,5	0,1
Alarmes falso-positivos	0	0	0	0	0	0

De forma geral, as Autofaces alcançaram resultados bastante superiores, em termos de generalização e robustez (Tabela 4.3).

Tabela 4.3 — Resultados da avaliação de robustez

	Tipo de distorção	Elevação da taxa de alarmes falso-negativos	Alarmes falso-positivos	PSNR
Autofaces (duas)	Ruído branco	0,05	0	28 dB
		0,2	0	22 dB
		0,65	0	19 dB
	PB (2×2)	-0,05	9	-
Antifaces (vinte)	Ruído branco	0,0	2	38 dB
		0,05	7	28 dB
	PB (2×2)	0,15	0	-
	PB (8×8)	0,15	0	-

Discussões e Conclusões

Neste trabalho, foi apresentado um estudo detalhado sobre duas técnicas de detecção e reconhecimento facial baseadas em projeções vetoriais: Autofaces e Antifaces. A seguir, são discutidos os resultados obtidos dos testes realizados com ambas as técnicas, assim como são apresentadas as contribuições alcançadas durante o desenvolvimento desta dissertação. Por fim, algumas possibilidades de trabalhos futuros serão sugeridas.

5.1 Discussão dos resultados

Algumas condições de treinamento foram especificadas no Capítulo 4 visando a alcançar um conjunto de regras que trouxesse resultados satisfatórios para ambos os métodos aqui estudados. Entretanto, frente a aplicações específicas, possivelmente há situações de treinamento mais apropriadas para cada uma das abordagens. Um exemplo de tais aplicações é o uso de Autofaces para detecção de faces com rotação ou de objetos tridimensionais. Nesse caso, em que os padrões são morfologicamente bastante diferentes, é usual a construção de mais de um espaço de faces. Sendo um conjunto de imagens adequado para treinar um dos métodos, frente a uma determinada aplicação, esse conjunto não necessariamente será adequado para treinar o outro método. Uma vez que a justa comparação entre o desempenho dos métodos deve ser realizada mediante situações idênticas de treinamento e testes, a procura pelas melhores condições de treinamento para cada método não foi analisada durante este trabalho.

Faces de diversos bancos de imagens foram utilizadas durante os testes, inclusive misturando imagens de diferentes bancos. Os testes de detecção seguem o mesmo padrão apresentado em [43].

O método de Antifaces apresenta um bom desempenho com relação ao treinamento com elevado número de imagens. Entretanto, apresenta um custo computacional razoavelmente elevado para a fase de treinamento, quando as imagens possuem dimensões relativamente altas. Esse fato é agravado pela sensibilidade do algoritmo às variáveis de projeto, assim como pela carência de regras bem definidas para a determinação do valor dessas variáveis, o que faz com que alguns detectores tenham de ser calculados e recalculados mais de uma vez à procura de um resultado consistente. Uma solução para confrontar a qualidade (número de detalhes) e o tamanho das imagens durante uma aplicação de detecção facial é a procura por características faciais. Assim, o número de padrões aumenta, porém as imagens de treinamento têm suas dimensões reduzidas.

Pela análise dos resultados do método de Antifaces, uma baixa capacidade de generalização é constatada no momento da detecção dos padrões desejados. Entretanto, essa característica pode ser traduzida como uma alta seletividade. Por esse motivo, essa abordagem pode apresentar-se como uma ferramenta eficiente em aplicações de reconhecimento facial.

Em razão da alta sensibilidade ao ruído e pouca capacidade de generalização, no lugar de detecção facial, uma aplicação típica para Antifaces seria a detecção de objetos em ambiente com fundo e iluminação controlados. Um exemplo desse tipo de aplicação é o controle de qualidade em processos industriais nos quais as peças inspecionadas seguem um padrão morfológico bem definido e são sempre apresentadas sobre um mesmo cenário.

Tratando-se da comparação dos dois métodos estudados, o método de Autofaces apresentou maior capacidade de generalização e menor sensibilidade ao ruído.

5.2 Contribuições

O trabalho desenvolvido consiste no primeiro estudo comparativo entre os métodos de Autofaces e Antifaces, tanto em nível nacional como internacional, de que se tem conhecimento. Considerações, aplicações e limitações do método de Antifaces foram apresentadas, formando uma base teórica que serve de subsídio para futuros trabalhos na área.

Uma nova abordagem para a determinação dos detectores subótimos, para o método de Antifaces, é apresentada no Apêndice 5.

5.3 Propostas para trabalhos futuros

Em razão de ser uma técnica ainda pouco explorada, as Antifaces sugerem uma série de estudos a serem desenvolvidos ao seu redor. Alguns exemplos são:

- Estudo dos efeitos causados pela substituição da DCT por outra transformada, que também forneça coeficientes reais, na avaliação do alisamento das Antifaces.
- Estudo específico sobre as características do conjunto de treinamento mais adequadas para cada tipo de aplicação.
- Estudo da eficiência (segurança) frente a aplicações de reconhecimento facial.
- Busca por maior capacidade de generalização, de forma que se possam realizar tarefas de detecção de faces genéricas.

Neste apêndice, é feita uma breve revisão, em sua maior parte baseada em [35],[45]-[49], sobre os conceitos matemáticos envolvidos nos métodos apresentados nos Capítulos 2 e 3.

Norma de um vetor

A norma de um vetor é análoga ao valor absoluto de escalares reais. A norma consiste em uma medida de distância para o espaço vetorial. Existem três tipos principais de norma de um vetor \mathbf{v} :

- A norma L_1 , definida como

$$\|\mathbf{v}\|_1 = |v_1| + |v_2| + \cdots + |v_n| = \sum_{i=1}^n |v_i|. \quad (\text{A1.1})$$

- A norma L_2 (norma Euclidiana), definida como

$$\|\mathbf{v}\|_2 = (|v_1|^2 + |v_2|^2 + \cdots + |v_n|^2)^{1/2} = \left(\sum_{i=1}^n |v_i|^2\right)^{1/2}. \quad (\text{A1.2})$$

- A norma L_∞ , definida como

$$\|\mathbf{v}\|_\infty = \max_{1 \leq i \leq n} |v_i|. \quad (\text{A1.3})$$

Na ausência de um subscrito, assume-se a norma $\|\cdot\|$ como sendo L_2 .

Distância Euclidiana

A distância Euclidiana entre dois vetores \mathbf{v}_1 e \mathbf{v}_2 é dada por

$$d_E(\mathbf{v}_1, \mathbf{v}_2) = \|\mathbf{v}_1 - \mathbf{v}_2\| = \sqrt{(\mathbf{v}_1 - \mathbf{v}_2)^T (\mathbf{v}_1 - \mathbf{v}_2)} \quad (\text{A1.4})$$

Singularidade e posto de uma matriz

Uma matriz quadrada \mathbf{A} é dita singular se existe algum vetor $\mathbf{x} \neq \mathbf{0}$, tal que $\mathbf{Ax} = \mathbf{0}$ ou $\mathbf{A}^T \mathbf{x} = \mathbf{0}$, onde $\mathbf{x} = \mathbf{0}$ representa um vetor com todos os componentes nulos.

Uma vez que, para qualquer matriz quadrada, o número de linhas linearmente independentes é igual ao número de colunas linearmente independentes, o posto de uma matriz é definido como o número de linhas ou colunas linearmente independentes. No caso de uma matriz \mathbf{A} de dimensão $(N \times N)$ possuir posto N , ela é dita matriz de “posto completo”.

Conceito de espaço e espaço nulo

Considere um conjunto de equações lineares representado por

$$\mathbf{Ax} = \mathbf{b}, \quad (\text{A1.5})$$

onde \mathbf{A} é uma matriz quadrada de dimensão $(N \times N)$; \mathbf{b} e \mathbf{x} são vetores de dimensão $(N \times 1)$. A Equação (A1.5) define \mathbf{A} como um mapeamento linear do espaço vetorial \mathbf{x} para o espaço vetorial \mathbf{b} . Se \mathbf{A} é singular, existe algum subespaço de \mathbf{x} , chamado “espaço nulo”, que é mapeado através de \mathbf{A} para zero ($\mathbf{Ax} = \mathbf{0}$). A dimensão do espaço nulo (número de vetores \mathbf{x} linearmente independentes que podem ser levados para zero) é chamada de nulidade de \mathbf{A} . Existe também algum subespaço de \mathbf{b} em que \mathbf{x} pode ser mapeado através de \mathbf{A} . Esse subespaço de \mathbf{b} é chamado de espaço de \mathbf{A} . A dimensão desse espaço é equivalente ao posto de \mathbf{A} . Se \mathbf{A} é não-singular, então seu espaço será todo o espaço vetorial \mathbf{b} , logo seu posto será completo (N). Por outro lado, se \mathbf{A} for singular, seu posto não será completo ($< N$).

A relação entre o posto e a nulidade é dada por

$$\text{posto} + \text{nulidade} = N \quad (\text{A1.6})$$

A matriz de covariância

A matriz de covariância, \mathbf{C}_x , da variável \mathbf{x} pode ser definida como

$$\mathbf{C}_x = E\{[\mathbf{x} - E(\mathbf{x})][\mathbf{x} - E(\mathbf{x})]^T\}. \quad (\text{A1.7})$$

Segue direto de (A1.7) que a matriz de covariância é a autocorrelação de um conjunto de variáveis “centralizadas”, ou seja, considerando-se que as variáveis tenham média zero.

O problema dos autovalores

Um escalar λ e um vetor \mathbf{u} de dimensão $(N \times 1)$ são chamados, respectivamente, de autovalor e autovetor de uma matriz \mathbf{A} , de dimensão $(N \times N)$, se

$$\mathbf{A}\mathbf{u} = \lambda\mathbf{u}. \quad (\text{A1.8})$$

Cálculo dos autovalores e autovetores

Para uma matriz \mathbf{A} de dimensão $(N \times N)$, o polinômio de ordem N definido por

$$\det(\mathbf{A} - \lambda\mathbf{I}) = 0 \quad (\text{A1.9})$$

é chamado de polinômio característico, ou equação característica de \mathbf{A} . As N raízes (não necessariamente distintas) do polinômio característico são os autovalores (ou raízes características) de \mathbf{A} . Cada autovalor possui um autovetor \mathbf{u} correspondente (diferente de $\mathbf{0}$) que pode ser encontrado resolvendo-se o sistema descrito por (A1.8) ou, isolando-se o termo \mathbf{u} , por

$$(\mathbf{A} - \lambda\mathbf{I})\mathbf{u} = \mathbf{0}. \quad (\text{A1.10})$$

Principais características e propriedades associadas ao problema dos autovalores

O traço de uma matriz \mathbf{A} de dimensão $(N \times N)$, normalmente descrito por $\text{tr}(\mathbf{A})$, é definido como a soma dos N componentes da diagonal principal de \mathbf{A} . Em termos dos autovalores de \mathbf{A} , o $\text{tr}(\mathbf{A})$ pode ser definido como

$$\text{tr}(\mathbf{A}) = \sum_{i=1}^N \lambda_i . \quad (\text{A1.11})$$

O determinante de uma matriz, por sua vez, pode ser definido sob os mesmos termos como

$$\det(\mathbf{A}) = \prod_{i=1}^N \lambda_i . \quad (\text{A1.12})$$

Os autovalores de uma matriz triangular ou diagonal são os componentes da diagonal da própria matriz.

Condicionamento de uma matriz

Formalmente, o condicionamento de uma matriz é definido pela razão entre o maior e o menor autovalor, λ_{\max} e λ_{\min} , respectivamente. Uma matriz é singular se essa razão for infinita e é mal-condicionada se essa razão for muito grande, i.e., se o complemento da razão atingir a ordem da precisão de ponto flutuante da máquina (por exemplo, menor do que 10^{-6} para precisão simples ou menor do que 10^{-12} para precisão dupla). Quanto maior o número de autovalores nulos, maior a dimensão do espaço nulo.

Demonstrações Envolvendo Autovalores

Demonstração 1

Considere uma matriz \mathbf{A} , de dimensão $(N \times M)$, onde $M < N$. A matriz $\mathbf{A}\mathbf{A}^T$ possui no máximo M autovalores não nulos.

Demonstração: o posto de uma matriz é menor ou igual ao menor dos seus números de linhas ou colunas [46]. Logo, sendo $M < N$, a matriz \mathbf{A} possui no máximo M linhas linearmente independentes. Da mesma forma, a matriz transposta \mathbf{A}^T possui no máximo M colunas linearmente independentes. Como o posto de um produto de duas matrizes não pode exceder o posto de cada uma das matrizes separadamente, o posto de $\mathbf{A}\mathbf{A}^T$ é no máximo igual a M .

Demonstração 2

Uma matriz \mathbf{A} , de dimensões $(M \times M)$ e posto completo, possui M autovalores não-nulos. Fazendo-se com que \mathbf{A} tenha linhas com média zero, apenas $(M - 1)$ autovalores serão não-nulos, visto que essa operação reduz o posto da matriz.

Demonstração: considerando a matriz de posto completo

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1M} \\ a_{21} & a_{22} & \cdots & a_{2M} \\ \vdots & \vdots & \ddots & \vdots \\ a_{M1} & a_{M2} & \cdots & a_{MM} \end{bmatrix}, \quad (\text{A2.1})$$

a operação de extração da média resulta em

$$\mathbf{A}' = \begin{bmatrix} a_{11} - \mu_1 & a_{12} - \mu_1 & \cdots & a_{1M} - \mu_1 \\ a_{21} - \mu_2 & a_{22} - \mu_2 & \cdots & a_{2M} - \mu_2 \\ \vdots & \vdots & \ddots & \vdots \\ a_{M1} - \mu_M & a_{M2} - \mu_M & \cdots & a_{MM} - \mu_M \end{bmatrix}, \quad (\text{A2.2})$$

onde

$$\mu_i = \frac{1}{M} \sum_{j=1}^M a_{ij} = \frac{a_{i1} + a_{i2} + \cdots + a_{iM}}{M}. \quad (\text{A2.3})$$

(A2.3) pode ser reescrita de forma vetorial, assim

$$\boldsymbol{\mu} = \frac{1}{M} \sum_{i=1}^M \mathbf{a}_i = \frac{\mathbf{a}_1 + \mathbf{a}_2 + \cdots + \mathbf{a}_M}{M}, \quad (\text{A2.4})$$

onde $\boldsymbol{\mu}$ é o vetor (coluna) de média e \mathbf{a}_i são vetores provenientes das colunas de \mathbf{A} .

Desenvolvendo-se (A2.4), tem-se

$$M\boldsymbol{\mu} = \mathbf{a}_1 + \mathbf{a}_2 + \cdots + \mathbf{a}_M \quad (\text{A2.5})$$

$$\mathbf{0} = (\mathbf{a}_1 + \mathbf{a}_2 + \cdots + \mathbf{a}_M) - M\boldsymbol{\mu} \quad (\text{A2.6})$$

$$\mathbf{0} = (\mathbf{a}_1 - \boldsymbol{\mu}) + (\mathbf{a}_2 - \boldsymbol{\mu}) + \cdots + (\mathbf{a}_M - \boldsymbol{\mu}) \quad (\text{A2.7})$$

$$- (\mathbf{a}_M - \boldsymbol{\mu}) = (\mathbf{a}_1 - \boldsymbol{\mu}) + (\mathbf{a}_2 - \boldsymbol{\mu}) + \cdots + (\mathbf{a}_{M-1} - \boldsymbol{\mu}) \quad (\text{A2.8})$$

$$(\mathbf{a}_M - \boldsymbol{\mu}) = - [(\mathbf{a}_1 - \boldsymbol{\mu}) + (\mathbf{a}_2 - \boldsymbol{\mu}) + \cdots + (\mathbf{a}_{M-1} - \boldsymbol{\mu})] \quad (\text{A2.9})$$

A partir de (A2.9), é possível observar que uma das colunas de \mathbf{A}' pode ser escrita como função das demais, ou seja, apenas $(M-1)$ colunas de \mathbf{A}' são linearmente independentes.

Fundamentos da Análise de Componentes Principais (PCA)

A PCA baseia-se no conceito de que um conjunto de P variáveis dependentes pode ser expresso através de um conjunto de Q variáveis independentes, onde $Q < P$. Considerando um conjunto infinito de vetores Q -dimensionais \mathbf{x} , é sempre possível gerar uma combinação linear que mapeie \mathbf{x} em um novo ponto \mathbf{y} , dentro de um espaço definido pelas variáveis ortonormais \mathbf{u}_j , para $j = 1, 2, \dots, Q$, sem perda de informação. Ou seja, \mathbf{x} pode ser expresso como

$$\mathbf{x} = \sum_{j=1}^Q y_j \mathbf{u}_j . \quad (\text{A3.1})$$

Por serem ortonormais, os vetores \mathbf{u}_j devem satisfazer a condição

$$\mathbf{u}_j^T \mathbf{u}_i = \delta(i - j) , \quad (\text{A3.2})$$

onde $\delta(i - j)$ é a função Delta de Kronecker, ou seja, vale 1 (um) se $i = j$ e zero se $i \neq j$. Se apenas P dimensões, no lugar de Q , são usadas, alguma informação é perdida e a estimativa de \mathbf{x} será

$$\hat{\mathbf{x}} = \sum_{j=1}^P y_j \mathbf{u}_j . \quad (\text{A3.3})$$

O problema agora consiste em determinar \mathbf{u}_j , para $j = 1, 2, \dots, P$, de forma que \mathbf{x} e $\hat{\mathbf{x}}$ sejam tão próximos um do outro quanto possível. Uma medida dessa proximidade é o erro quadrático, definido por

$$e^2 = (\mathbf{x} - \hat{\mathbf{x}})^T (\mathbf{x} - \hat{\mathbf{x}}). \quad (\text{A3.4})$$

Portanto, para se obter o erro mínimo (solução ótima), (A3.4) deve ser minimizada. Considerando-se \mathbf{x} como sendo uma variável aleatória, a minimização terá sentido somente se for feita em termos probabilísticos ou em termos médios. Assim, a função a ser minimizada passa a ser

$$E(e^2) = E[(\mathbf{x} - \hat{\mathbf{x}})^T (\mathbf{x} - \hat{\mathbf{x}})]. \quad (\text{A3.5})$$

Substituindo-se (A3.1) e (A3.3) em (A3.5), tem-se

$$\begin{aligned} E(e^2) &= E\left[\left(\sum_{j=1}^Q y_j \mathbf{u}_j - \sum_{j=1}^P y_j \mathbf{u}_j\right)^T \left(\sum_{j=1}^Q y_j \mathbf{u}_j - \sum_{j=1}^P y_j \mathbf{u}_j\right)\right] \\ &= E\left[\left(\sum_{j=P+1}^Q y_j \mathbf{u}_j\right)^T \left(\sum_{j=P+1}^Q y_j \mathbf{u}_j\right)\right]. \end{aligned} \quad (\text{A3.6})$$

Agora, aplicando-se (A3.2) em (A3.6), obtém-se

$$E(e^2) = E\left(\sum_{j=P+1}^Q y_j^2\right). \quad (\text{A3.7})$$

Multiplicando ambos os lados de (A3.1) por \mathbf{u}_j^T , tem-se

$$\mathbf{u}_j^T \mathbf{x} = \mathbf{u}_j^T \sum_{i=1}^Q y_i \mathbf{u}_i \quad (\text{A3.8})$$

$$\mathbf{u}_j^T \mathbf{x} = y_j. \quad (\text{A3.9})$$

Substituindo-se (A3.9) em (A3.7), obtém-se

$$E(e^2) = E\left(\sum_{j=P+1}^Q \mathbf{u}_j^T \mathbf{x} \mathbf{x}^T \mathbf{u}_j\right). \quad (\text{A3.10})$$

Uma vez que \mathbf{u}_j é determinístico, a ordem das operações de soma e valor esperado pode ser comutada. Assim,

$$E(e^2) = \sum_{j=P+1}^Q \mathbf{u}_j^T [E(\mathbf{x}\mathbf{x}^T)] \mathbf{u}_j. \quad (\text{A3.11})$$

O termo dentro dos colchetes é definido como a matriz de autocorrelação de \mathbf{x} , \mathbf{R}_x . Assim, a expressão final do erro quadrático é

$$E(e^2) = \sum_{j=P+1}^Q \mathbf{u}_j^T \mathbf{R}_x \mathbf{u}_j. \quad (\text{A3.12})$$

Para casos em que $E(\mathbf{x}) = \mathbf{0}$, a matriz de correlação \mathbf{R}_x é equivalente à matriz de covariância \mathbf{C}_x .

A função de erro quadrático deve agora ser minimizada, de forma a encontrar os \mathbf{u}_j ótimos. Isso pode ser feito igualando-se a derivada da função de erro a zero. Entretanto, a diferenciação deve ser feita de forma que a condição imposta por (A3.2) seja preservada. Esse problema é resolvido através do uso de uma função $g(\mathbf{u}_j)$, com restrição, e da técnica de multiplicadores de Lagrange, formulada como segue:

$$g(\mathbf{u}_j) = \sum_{j=P+1}^Q \mathbf{u}_j^T \mathbf{R}_x \mathbf{u}_j - \sum_{j=P+1}^Q \lambda_j [\mathbf{u}_j^T \mathbf{u}_j - 1]. \quad (\text{A3.13})$$

Derivando-se (A3.13) e igualando a zero, tem-se

$$\frac{\partial g(\mathbf{u}_j)}{\partial \mathbf{u}_j} = \mathbf{R}_x \mathbf{u}_j - \lambda_j \mathbf{u}_j \quad (\text{A3.14})$$

$$\frac{\partial g(\mathbf{u}_j)}{\partial \mathbf{u}_j} = (\mathbf{R}_x - \lambda_j \mathbf{I}) \mathbf{u}_j = 0, \quad (\text{A3.15})$$

onde \mathbf{I} é a matriz identidade. (A3.15) representa um problema de autovalores, podendo ser facilmente resolvido através de métodos numéricos.

O j -ésimo autovalor calculado é, por definição, igual à variância do j -ésimo autovetor. Considerando que \mathbf{x} assumia apenas poucos valores distintos, dentro de um

universo maior de possibilidades, existirão dimensões para as quais os autovalores associados serão nulos, indicando que o conjunto de dados ocupa apenas um subespaço do espaço original. Em outras palavras, o número de autovalores não-zero é igual ao posto da matriz \mathbf{R}_x , ou seja, ao número de linhas linearmente independentes.

Resolvido o problema dos autovalores, a escolha dos autovetores que irão compor o subespaço de interesse depende do tipo de aplicação. Escolhidos os autovetores a serem utilizados, uma matriz de transformação \mathbf{U} é montada, colocando-se os autovetores escolhidos organizados como vetores coluna. O procedimento de transformação (ou projeção dentro desse subespaço) é efetuado diretamente por (ver (A3.9))

$$\mathbf{y} = \mathbf{U}^T \mathbf{x} \quad (\text{A3.16})$$

ou por

$$\mathbf{y} = \mathbf{U}^T (\mathbf{x} - \boldsymbol{\mu}_x) \quad (\text{A3.17})$$

no caso de \mathbf{x} possuir média diferente de zero.

Decomposição em Valores Singulares (SVD)

A Decomposição em Valores Singulares é baseada no seguinte teorema de álgebra linear: qualquer matriz \mathbf{A} de dimensão $(M \times N)$, onde $M \geq N$, pode ser escrita como o produto de três matrizes (A4.1): uma matriz \mathbf{U} $(M \times N)$ com colunas ortogonais; uma matriz diagonal $\mathbf{\Omega}$ $(N \times N)$ com elementos positivos ou nulos (valores singulares); uma matriz ortogonal \mathbf{V}^T $(N \times N)$.

$$[\mathbf{A}]_{(M \times N)} = [\mathbf{U}]_{(M \times N)} \begin{bmatrix} \omega_1 & 0 & \cdots & 0 \\ 0 & \omega_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \omega_N \end{bmatrix}_{(N \times N)} [\mathbf{V}]_{(N \times N)}^T. \quad (\text{A4.1})$$

Tanto \mathbf{U} quanto \mathbf{V} são matrizes ortogonais. Portanto,

$$\mathbf{U}^T \mathbf{U} = \mathbf{V}^T \mathbf{V} = \mathbf{I}, \quad (\text{A4.2})$$

onde \mathbf{I} é a matriz identidade. Visto que \mathbf{V} é uma matriz quadrada, e também que suas linhas são ortonormais, $\mathbf{V}\mathbf{V}^T = \mathbf{I}$.

Por definição, as colunas de \mathbf{U} correspondem aos autovetores de $\mathbf{A}\mathbf{A}^T$, e as colunas de \mathbf{V} correspondem aos autovetores de $\mathbf{A}^T\mathbf{A}$. A matriz $\mathbf{\Omega}$, por sua vez, é composta por coeficientes ω_j que correspondem à raiz quadrada dos autovalores de $\mathbf{A}^T\mathbf{A}$.

A SVD pode também ser obtida no caso em que $M < N$. Nesse caso, os valores singulares ω_j para $j = M + 1, \dots, N$ são todos nulos. A SVD pode ainda ser obtida mesmo que a matriz \mathbf{A} seja singular.

Avaliação do condicionamento e singularidade de uma matriz via SVD

Em decorrência da equivalência entre os valores singulares e os autovalores, através da decomposição em valores singulares, é possível avaliar o condicionamento e a singularidade de uma matriz.

A inversa de uma matriz calculada por SVD

Se a matriz \mathbf{A} é quadrada de dimensão $(N \times N)$, então \mathbf{U} , \mathbf{V} e $\mathbf{\Omega}$ são todas matrizes quadradas com a mesma dimensão. Suas inversas são também triviais de calcular: como \mathbf{U} e \mathbf{V} são matrizes ortogonais, então suas inversas são iguais às suas transpostas; $\mathbf{\Omega}$ é diagonal, logo sua inversa é a matriz diagonal cujos elementos são os inversos dos elementos ω_j . Segue diretamente de (A4.1) que a inversa de \mathbf{A} é

$$\mathbf{A}^{-1} = \mathbf{V}[\text{diag}(1/\omega_j)]\mathbf{U}^T. \quad (\text{A4.3})$$

O único problema nesse cálculo é quando um dos ω_j é nulo, ou (numericamente) muito pequeno, o que ocorre quando a matriz \mathbf{A} é singular ou mal condicionada, respectivamente.

Solução de sistemas de equações lineares via SVD

A SVD dá origem a bases ortogonais para o espaço e para o espaço nulo de uma matriz. Especificamente, as colunas de \mathbf{U} cujos respectivos elementos ω_j são não-nulos formam um conjunto de vetores de base ortonormais para o espaço de \mathbf{A} ; as colunas de \mathbf{V} cujos respectivos elementos ω_j são nulos formam uma base ortonormal para o espaço nulo.

Considerando então \mathbf{A} como sendo uma matriz singular e $\mathbf{b} = \mathbf{0}$, o conjunto de equações descrito por

$$\mathbf{Ax} = \mathbf{b} \quad (\text{A4.4})$$

(chamado sistema de equações homogêneas), possui solução trivial de norma mínima $\mathbf{x} = \mathbf{0}$. Entretanto, muitas aplicações em engenharia requerem uma solução não trivial, buscando resolver o problema sob uma restrição de $\|\mathbf{x}\|=1$. Esse tipo de problema é

solucionado imediatamente por SVD. Assim, qualquer coluna de \mathbf{V} que possua um ω_j correspondente igual a zero fornece uma solução.

Quando o vetor \mathbf{b} , por outro lado, não é nulo, a questão é se ele pode ser mapeado no espaço de \mathbf{A} ou não. Caso seja possível, o conjunto de equações não-homogêneas possui uma solução \mathbf{x} . Caso \mathbf{A} seja singular, qualquer combinação linear entre a solução \mathbf{x} e qualquer vetor no espaço nulo (qualquer coluna de \mathbf{V} com o respectivo elemento $\omega_j = 0$) também consiste em uma solução.

Para selecionar apenas um membro particular desse conjunto de soluções, a SVD fornece o vetor solução com a menor $\|\mathbf{x}\|$ desse conjunto. Para tal, os termos $1/\omega_j$ devem ser substituídos por zero no caso de $\omega_j = 0$ e, em seguida, calcula-se

$$\mathbf{x} = \mathbf{V}[\text{diag}(1/\omega_j)]\mathbf{U}^T \mathbf{b}. \quad (\text{A4.5})$$

Em [47], é apresentada uma prova de que esse é o vetor solução de norma mínima. As colunas de \mathbf{V} que estão no espaço nulo completam o conjunto solução.

Caso \mathbf{b} não esteja no espaço da matriz singular \mathbf{A} , então o conjunto de equações descrito por (A4.4) não possui solução. Porém, (A4.5) ainda pode ser utilizada para construir um vetor solução \mathbf{x} . Esse vetor não fornecerá uma solução para $\mathbf{Ax} = \mathbf{b}$, pois, como foi visto, ela não existe. Entretanto, a solução fornecida por (A4.5) será a solução com o menor erro, no sentido dos mínimos quadrados. Em outras palavras, a Equação (A4.5) encontra o vetor \mathbf{x} que minimiza

$$r = \|\mathbf{Ax} - \mathbf{b}\|, \quad (\text{A4.6})$$

onde r é chamado resíduo da solução. Um exemplo simplificado de decomposição em valores singulares é ilustrado na Figura D.1.

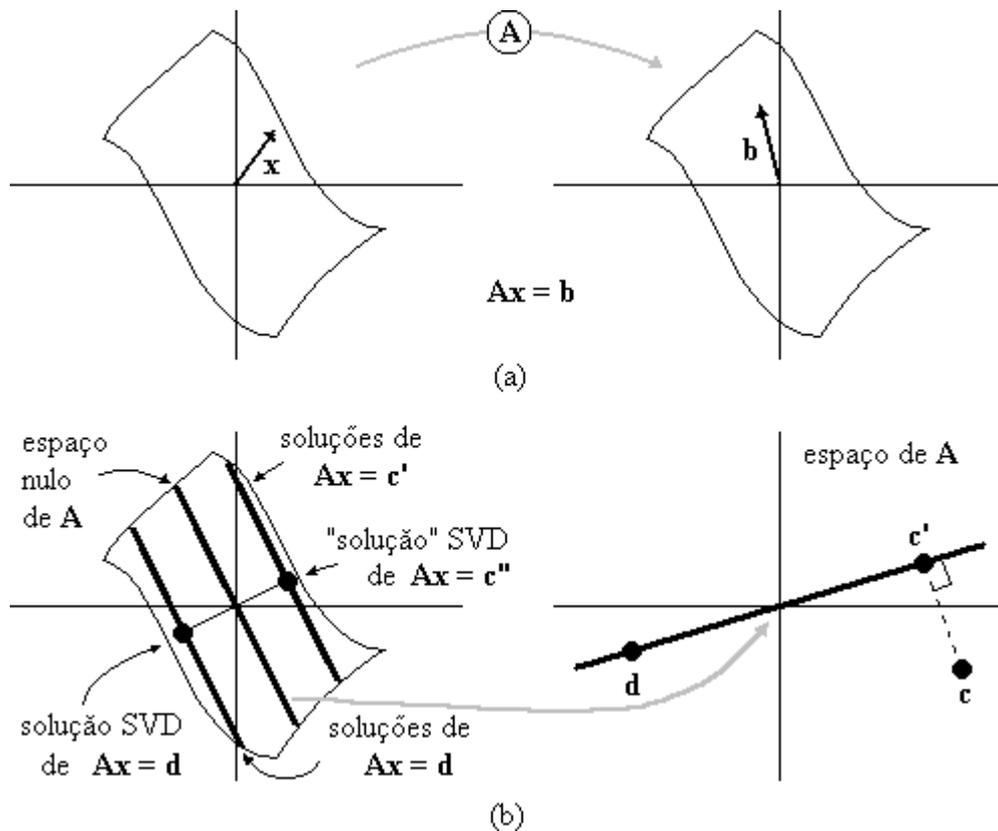


Figura D.1 — Exemplo de SVD em que: (a) a matriz não-singular A mapeia o vetor x no espaço de b ; (b) a matriz singular A mapeia um espaço vetorial em outro de menor dimensão. Nesse caso, o espaço nulo de A é mapeado para o zero, a solução de $Ax = d$ consiste em qualquer solução particular somada a qualquer vetor do espaço nulo e $Ax = c''$ é a solução com o menor erro, no sentido dos mínimos quadrados.

Determinação dos Detectores Subótimos

Este apêndice apresenta a determinação dos detectores subótimos, fazendo uso da decomposição em valores singulares (Apêndice 4)

Conforme visto na Seção 3.3, o problema consiste em:

Dado um conjunto de M padrões, $\mathbf{T} = [\mathbf{t}_1, \mathbf{t}_2, \dots, \mathbf{t}_M]$, onde \mathbf{t}_j consiste em um padrão \mathbf{F}_j de dimensão $(N \times N)$ vetorizado, encontrar um detector \mathbf{d}_1 de dimensão $(N^2 \times 1)$ que minimize

$$\sum_{j=1}^M \langle \mathbf{d}_1, \mathbf{t}_j \rangle^2 + \lambda S(\mathbf{d}_1), \quad (\text{A5.1})$$

sujeito a

$$\|\mathbf{d}_1\| = 1.$$

Através do processo inverso da vetorização feita para transformar \mathbf{F}_j em \mathbf{t}_j , pode-se escrever o vetor \mathbf{d}_1 na forma matricial \mathbf{D}_1 , e a função de alisamento $S(\mathbf{d}_1)$ pode então ser representada por

$$S(\mathbf{d}_1) = \sum_{k=2}^N \sum_{l=2}^N (k^2 + l^2) \mathcal{D}_1(k, l)^2, \quad (\text{A5.2})$$

onde $\mathcal{D}_1(k, l)$ denota os coeficientes da DCT (transformada cosseno discreta) bidimensional de \mathbf{D}_1 .

Em razão de (A5.2) ser descrita no domínio DCT, é conveniente que todas as variáveis sejam transformadas para o domínio da frequência. Da mesma maneira que para \mathcal{D}_1 , \mathcal{F}_j é a DCT bidimensional do padrão \mathbf{F}_j . A vetorização de \mathcal{D}_1 e de \mathcal{F}_j resulta em $\mathbf{t}_j = \text{vec}(\mathcal{F}_j)$ e $\mathbf{d}_1 = \text{vec}(\mathcal{D}_1)$.

Substituindo-se (A5.2) em (A5.1) e transformando-se todas as variáveis para o domínio DCT, segue que

$$\sum_{j=1}^M \mathbf{d}_1^T \mathbf{t}_j \mathbf{t}_j^T \mathbf{d}_1 + \lambda \sum_{k=2}^N \sum_{l=2}^N (k^2 + l^2) \mathcal{D}_1(k, l)^2. \quad (\text{A5.3})$$

Sendo \mathbf{P} uma matriz de pesos descrita por

$$\mathbf{P}(k, l) = k^2 + l^2, \quad (\text{A5.4})$$

e

$$\mathbf{P}_1 = \text{diag}(\text{vet}(\mathbf{P})), \quad (\text{A5.5})$$

(A5.3) pode ser reescrita de forma matricial, como

$$\begin{aligned} \sum_{j=1}^M \mathbf{d}_1^T \mathbf{t}_j \mathbf{t}_j^T \mathbf{d}_1 + \lambda \mathbf{d}_1^T \mathbf{P}_1 \mathbf{d}_1 &= \mathbf{d}_1^T \left(\sum_{j=1}^M \mathbf{t}_j \mathbf{t}_j^T + \lambda \mathbf{P}_1 \right) \mathbf{d}_1 \\ &= \mathbf{d}_1^T (\mathcal{T} \mathcal{T}^T + \lambda \mathbf{P}_1) \mathbf{d}_1 \\ &= \mathbf{d}_1^T \mathbf{M} \mathbf{M}^T \mathbf{d}_1, \end{aligned} \quad (\text{A5.6})$$

onde \mathcal{T} é a DCT bidimensional de \mathbf{T} e

$$\mathbf{M} \mathbf{M}^T = \mathcal{T} \mathcal{T}^T + \lambda \mathbf{P}_1. \quad (\text{A5.7})$$

Visto que $\mathcal{T} \mathcal{T}^T$ é uma matriz simétrica e \mathbf{P}_1 uma matriz diagonal, $\mathbf{M} \mathbf{M}^T$ mantém-se simétrica. Realizando-se a SVD de \mathbf{M} , obtém-se

$$\mathbf{M} = \mathbf{U} \mathbf{S} \mathbf{V}^T, \quad (\text{A5.8})$$

onde \mathbf{U} é a matriz de autovetores de $\mathbf{M}\mathbf{M}^T$, \mathbf{S} é a matriz diagonal de valores singulares ($\mathbf{S} = \text{diag}([\sigma_1, \sigma_2, \dots, \sigma_N])$, onde $\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_N$ e σ_n^2 são os autovalores de $\mathbf{M}^T\mathbf{M}$) e \mathbf{V} é a matriz de autovetores de $\mathbf{M}^T\mathbf{M}$. Uma vez que \mathbf{M} é simétrica, $\mathbf{M}^T = \mathbf{M}$ e $\mathbf{U} = \mathbf{V}$. Logo,

$$\mathbf{d}_1^T \mathbf{M}\mathbf{M}^T \mathbf{d}_1 = \mathbf{d}_1^T \mathbf{U}\mathbf{S}\mathbf{V}^T \mathbf{V}\mathbf{S}^T \mathbf{U}^T \mathbf{d}_1. \quad (\text{A5.9})$$

Sendo, ainda, \mathbf{V} uma matriz ortonormal, tem-se que $\mathbf{V}^{-1} = \mathbf{V}^T$, logo $\mathbf{V}^T \mathbf{V} = \mathbf{I}$. Assim,

$$\mathbf{d}_1^T \mathbf{M}\mathbf{M}^T \mathbf{d}_1 = \mathbf{d}_1^T \mathbf{U}\mathbf{S}\mathbf{S}^T \mathbf{U}^T \mathbf{d}_1. \quad (\text{A5.10})$$

Definindo-se

$$\mathbf{y} = \mathbf{U}^T \mathbf{d}_1, \quad (\text{A5.11})$$

o problema reduz-se a minimizar

$$\mathbf{y}^T \mathbf{S}\mathbf{S}^T \mathbf{y} \quad (\text{A5.12})$$

sujeito a

$$\|\mathbf{y}\| = 1,$$

onde a matriz $\mathbf{S}\mathbf{S}^T$ é diagonal com termos s_n^2 . Dessa forma, a função objetivo passa a ser uma soma ponderada do quadrado dos componentes de \mathbf{y} . Um limite inferior para essa soma pode ser obtido substituindo-se cada um dos valores singulares pelo menos significativo. Logo,

$$\begin{aligned} \mathbf{y}^T \mathbf{S}\mathbf{S}^T \mathbf{y} &= \sum_{n=1}^N s_n^2 y_n^2 \\ &\geq \sum_{n=1}^N s_N^2 y_n^2 \\ &= s_N^2 \sum_{n=1}^N y_n^2 \\ &= s_N^2 \|\mathbf{y}\|^2 \\ &= s_N^2, \end{aligned} \quad (\text{A5.13})$$

onde, no último passo de (A5.13), foi empregada a restrição de que \mathbf{y} seja um vetor de norma unitária. A inequação torna-se uma igualdade no momento em que $\mathbf{y} = [0, 0, \dots, 0, 1]^T$, sendo nesse caso $\mathbf{y} = \mathbf{y}_{\text{ótimo}}$.

Substituindo-se $\mathbf{y}_{\text{ótimo}}$ em (A5.11), temos que

$$\mathbf{d}_{\text{1ótimo}} = \mathbf{U}\mathbf{y}_{\text{ótimo}} = \mathbf{u}_N, \quad (\text{A5.14})$$

onde \mathbf{u}_N é a N -ésima coluna de \mathbf{U} (autovetor associado ao menor autovalor) [50].

Para o segundo detector, agora, basta incluímos a restrição de que

$$\mathbf{d}_2^T \mathbf{P}_2 \mathbf{d}_1 = 0, \quad (\text{A5.15})$$

onde \mathbf{P}_2 é definida como

$$\mathbf{P}_2(k, l) = \mathbf{P}_1(k, l)^{-3/2}. \quad (\text{A5.16})$$

Isto pode ser feito da seguinte forma:

$$\sum_{j=1}^M \langle \mathbf{d}_2, \mathbf{t}_j \rangle^2 + \lambda \sum_{k=2}^N \sum_{l=2}^N (k^2 + l^2) \mathcal{D}_2(k, l)^2 + \lambda_2 (\mathbf{d}_2^T \mathbf{P}_2 \mathbf{d}_1 \mathbf{d}_1^T \mathbf{P}_2 \mathbf{d}_2). \quad (\text{A5.17})$$

Os dois primeiros termos de (A5.17) correspondem ao mesmo processo de otimização realizado para o detector \mathbf{d}_1 . O terceiro termo inclui a restrição de independência entre os detectores na forma quadrática como uma função de penalidade [51]. O fator de penalidade λ_2 garante que a restrição mantenha-se tão próxima de zero quanto se queira.

Procedendo-se da mesma forma que em (A5.6), pode-se definir uma matriz $\mathbf{M}_2 \mathbf{M}_2^T$ como

$$\mathbf{M}_2 \mathbf{M}_2^T = \mathcal{T} \mathcal{T}^T + \lambda \mathbf{P}_1 + \lambda_2 (\mathbf{P}_2 \mathbf{d}_1 \mathbf{d}_1^T \mathbf{P}_2). \quad (\text{A5.18})$$

O segundo detector é então determinado da mesma forma que o primeiro. O mesmo procedimento é utilizado para os demais detectores.

Determinados os detectores \mathcal{d}_j (domínio da frequência), é realizado o processo inverso de transformação DCT (bidimensional) para que se obtenha os detectores \mathbf{d}_j (domínio da seqüência (imagem)).

Referências Bibliográficas

- [1] YANG, M. H., KRIEGMAN, D. J., AHUJA, N. **Detecting Faces in Images: A Survey**. In: IEEE Transactions on Pattern Analysis and Machine Intelligence, v.24, p-34-58, Jan. 2002.
- [2] DENG. Y., MANJUNATH. B. S. **Unsupervised Segmentation of Color-Texture Regions in Images and Video**. In: IEEE Transactions on Pattern Analysis and Machine Intelligence, v.23, p-800-810, Aug. 2001.
- [3] IMAGAWA, K., LU, S., IGI, S. Color-Based Hands Tracking System for Sign Language Recognition. In: 3rd IEEE INTERNATIONAL CONFERENCE ON AUTOMATIC FACE AND GESTURE RECOGNITION, Nara (Japan), Apr. 1998. **Proceedings**: Nara, FG'98, 1998, p-462-467.
- [4] HSU, R. L., ABDEL-MOTTALEB, M., JAIN, A. K. **Face Detection in Color Images**. In: IEEE Transactions on Pattern Analysis and Machine Intelligence, v.24, p-800-810, May. 2002.
- [5] WU, H., CHEN, Q., YACHIDA, M. **Face Detection from Color Images Using a Fuzzy Pattern Matching Method**. In: IEEE Transactions on Pattern Analysis and Machine Intelligence, v.21, p-557-563, June 1999.
- [6] ———, ———, ———. Detecting Human Face in Color Images. In: IEEE INTERNATIONAL CONFERENCE ON SYSTEM, MAN, AND CYBERNETICS, Beijing (China), Oct. 1996. **Proceedings**: Beijing, 1996, v.3, p-2232-2237.
- [7] ———, ———, ———. A Fuzzy-Theory-Based Face Detector. In: IEEE INTERNATIONAL CONFERENCE PATTERN RECOGNITION, Vienna (Austria), Aug. 1996. **Proceedings**: Vienna, ICPR 1996, v.3, p-406-410.
- [8] ———, ———, ———. Face Detection by Fuzzy Pattern Matching. In: 5th INTERNATIONAL CONFERENCE ON COMPUTER VISION, Massachusetts (USA), June 1995. **Proceedings**: Massachusetts, 1995, p-591-596.
- [9] ———, YOKOYAMA, T., PRAMADIHANTO, D., et al. Face and Facial Feature Extraction from Color Image. In: 2nd INTERNATIONAL CONFERENCE ON AUTOMATIC FACE AND GESTURE RECOGNITION, Vermont (USA), Oct. 1996. **Proceedings**: Vermont, 1996, p-345-350.

-
- [10] ALBIOL, A., TORRES, L., BOUMAN, C. A., *et al.* A Simple and Efficient Face Detection Algorithm for Video Database Applications. In: INTERNATIONAL CONFERENCE ON IMAGE PROCESSING, Vancouver (Canada), Sept. 2000. **Proceedings**: Vancouver, ICIP 2000, v.2, p-239-242.
- [11] ———, ———, DELP, E. J. An Unsupervised Color Image Segmentation Algorithm for Face Detection Application. In: INTERNATIONAL CONFERENCE ON IMAGE PROCESSING, Thessaloniki (Greece), Oct. 2001. **Proceedings**: Thessaloniki, ICIP 2001, v.2, p-681-684.
- [12] LUCEY, S., SRIDHARAN, S., CHANDRAN, V. A Suitability Metric for Mouth Tracking Through Chromatic Segmentation. In: INTERNATIONAL CONFERENCE ON IMAGE PROCESSING, Thessaloniki (Greece), Oct. 2001. **Proceedings**: Thessaloniki, ICIP 2001, v.3, p-258-261.
- [13] FERIS, R. S., CAMPOS, T. E., CESAR JUNIOR, R. M. Detection and Tracking of Facial Features in Video Sequences. In: MEXICAN INTERNATIONAL CONFERENCE ON ARTIFICIAL INTELIGENCE, Acapulco (Mexico), Apr. 2000. **Proceedings**: Acapulco, MICAI 2000.
- [14] NOVAK, V. **Detection of Humans Using Color Information**, 2001. Dissertação (Mestrado). Center of Machine Perception, Faculty of Electrical Engineering, Czech Technical University. Praha, Czech Republic.
- [15] MENSER, B., MÜLLER, F. Face Detection in Color Images Using Principal Component Analysis. In: 7th INTERNATIONAL CONFERENCE ON IMAGE PROCESSING AND ITS APPLICATIONS, Manchester (UK), 1999. **Proceedings**: Manchester , 1999, (conf. publ. 465) v.2, p-620-624.
- [16] CHAI, D., NGAN, K. Locating Facial Region of a Head-and-Shoulders Color Image. In: 3rd IEEE INTERNATIONAL CONFERENCE ON AUTOMATIC FACE AND GESTURE RECOGNITION, Nara (Japan), Apr. 1998. **Proceedings**: Nara, FG'98, 1998, p-124-129.
- [17] ANGELOPOULOU, E., MOLANA, R., DANIILIDIS, K. Multispectral Skin Color Modeling. In: IEEE COMPUTER SOCIETY CONFERENCE ON COMPUTER VISION AND PATTERN RECOGNITION, Hawaii (USA), 2001. **Proceedings**: Hawaii, CVPR 2001, v.2, p-635-642.
- [18] KAMPMANN, M. Segmentation of a Head into Face, Ears Neck and Hair for Knowledge-Based Analysis-Synthesis Coding of Videophone Sequences. In: INTERNATIONAL CONFERENCE ON IMAGE PROCESSING, Chicago (USA), 1998. **Proceedings**: Chicago, ICIP 98, 1998, v.2, p-876-880.
- [19] KIM, S. H., KIM, A. K., AHN, S. C., *et al.* Object Oriented Face Detection Using Range and Color Information. In: 3rd IEEE INTERNATIONAL CONFERENCE ON AUTOMATIC FACE AND GESTURE RECOGNITION, Nara (Japan), Apr. 1998. **Proceedings**: Nara, FG'98, 1998, p-76-81.

- [20] SUN, Q. B., HUANG, M. W., WU, J. K. Face Detection Based on Color and Local Symmetry Information. In: 3rd IEEE INTERNATIONAL CONFERENCE ON AUTOMATIC FACE AND GESTURE RECOGNITION, Nara (Japan), Apr. 1998. **Proceedings**: Nara, FG'98, 1998, p-130-135.
- [21] FERIS, R. S. **Detecção e Rastreamento de Faces Utilizando Gabor Wavelet Networks**, 2000. Qualificação (Mestrado). Departamento de Ciências da Computação, IME, USP, São Pulo, Brasil.
- [22] ZARIT, B. D., SUPER, B. J., QUEK, F. K. H. Comparison of Five Color Models in Skin Pixel Classification. In: INTERNATIONAL WORKSHOP ON RECOGNITION ANALYSIS, AND TRACKING OF FACES AND GESTURES IN REAL-TIME SYSTEMS, Corfu (Greece), 1999. **Proceedings**: Corfu, 1999, p-58-63.
- [23] MARTINKAUPPI, J. B., SORIANO, M. N., LAAKSONEN, M. H. Behavior of skin color under varying illumination seen by different cameras at different color spaces. In: MACHINE VISION APPLICATIONS IN INDUSTRIAL INSPECTION IX, Jan. 2001, California (USA). **Proceedings**: California, SPIE 2001, v.4301, p-102-112
- [24] SORIANO, M., MARSZALEC, E., PIETIKÄINEN, M. Color Correction of Face Images Under Different Illuminants by RGB Eigenfaces. In: 2nd AUDIO- AND VIDEO-BASED BIOMETRIC PERSON AUTHENTICATION CONFERENCE (AVBPA99), Mar. 1999, Washington DC (USA). **Proceedings**: Washington, AVBPA99, 1999, p-148-153.
- [25] STÖRRING, M., ANDERSEN, H. J., GRANUM, E. Estimation of the Illuminant Colour from Human Skin Colour. In: 4th IEEE INTERNATIONAL CONFERENCE ON AUTOMATIC FACE AND GESTURE RECOGNITION, Aug. 2000, Grenoble (France). **Proceedings**: Grenoble, 2000, p-64-69.
- [26] ———, GRANUM, E., ANDERSEN, H. J. Estimation of the Illuminant Colour Using Highlights from Human Skin. In: 1st INTERNATIONAL CONFERENCE ON COLOR IN GRAPHICS AND IMAGE PROCESSING, Oct. 2000, Saint-Etienne (France). **Proceedings**: Saint-Etienne, CGIP 2000, p-45-50.
- [27] YOKOYAMA, T., YAGI, Y., YACHIDA, M. Facial Contour Extraction Model. In: 3rd IEEE INTERNATIONAL CONFERENCE ON AUTOMATIC FACE AND GESTURE RECOGNITION, Nara (Japan), Apr. 1998. **Proceedings**: Nara, FG'98, 1998, p-254-259.
- [28] BURGESS, C. A **Tutorial on Support Vector Machines for Pattern Recognition**. In: Data Mining and Knowledge Discovery, number 2, vol. 2, p-121-167, 1998.

- [29] SERRE, T., HEISELE, B., POGGIO, T. et al. Feature Selection for Face Detection. In: Center for Biological and Computational Learning, A. I. memo 1697, MIT, Cambridge, Sept., 2000.
- [30] HEISELE, B., HO, P., POGGIO, T. Face Recognition with Support Vector Machines: Global versus Component-based Approach. In: 8th IEEE INTERNATIONAL CONFERENCE ON COMPUTER VISION, Vancouver (Canada), July, 2001. **Proceedings**: Vancouver, ICCV2001, 2001, p-688-694.
- [31] LI, Y., GONG, S., SHERRAH, J. et al. Multi-View Face Detection Using Support Vector Machines and Eigenspace Modelling. In: 4th IEEE INTERNATIONAL CONFERENCE ON KNOWLEDGE-BASED INTELLIGENT ENGINEERING & ALLIED TECHNOLOGIES, Brighton (UK), Aug. 2000. **Proceedings**: Brighton, 2000, p-241-244.
- [32] PONTIL, M., VERRI, A. **Support Vector Machines for 3D Object Recognition**. In: IEEE Transactions on Pattern Analysis and Machine Intelligence, v.20, p-637-646, June, 1998.
- [33] OSUNA, E., FREUND, R., GIROSI, F. Training Support Vector Machines: an Application to Face Detection. In: IEEE COMPUTER SOCIETY CONFERENCE ON COMPUTER VISION AND PATTERN RECOGNITION, Puerto Rico, 1997. **Proceedings**: Puerto Rico, CVPR 1997, p-130-136.
- [34] BAKER, S., NAYAR, S. Pattern Rejection. In: IEEE COMPUTER SOCIETY CONFERENCE ON COMPUTER VISION AND PATTERN RECOGNITION, San Francisco (USA), 1996. **Proceedings**: San Francisco, CVPR 1996, p-544-549.
- [35] ZWILLINGER, D. *Standard Mathematical Tables and Formulae*. 30th Edition: New York, CRC Press, 1996.
- [36] CAMPOS, T. E. **Técnicas de Seleção de Atributos e de Classificação para Reconhecimento de Faces**, 2000. Dissertação (Mestrado). Departamento de Ciências da Computação, IME, USP, São Paulo, Brasil.
- [37] MOGHADDAM, B., WAHID, W., PENTLAND, A. Beyond Eigenfaces: Probabilistic Matching for Face Recognition. In: 3rd IEEE INTERNATIONAL CONFERENCE ON AUTOMATIC FACE AND GESTURE RECOGNITION, Nara (Japan), Apr. 1998. **Proceedings**: Nara, FG'98, 1998, p-30-35.
- [38] TURK, M. A., PENTLAND, A. Eigenfaces for Recognition. In: **Journal of Cognitive Neuroscience**, 3(1), p-71-86, Mar. 1991.
- [39] ———, ———. Face Recognition Using Eigenfaces. 1991. In: IEEE COMPUTER SOCIETY CONFERENCE ON COMPUTER VISION AND PATTERN RECOGNITION, 1991. **Proceedings**: 1991, p-586-591.

- [40] MOGHADDAM, B., ————. **Probabilistic Visual Learning for Object Representation**. In: IEEE Transactions on Pattern Analysis and Machine Intelligence, v.19, p-696-710, July 1997.
- [41] ANDRADE NETO, E. **Sistema de Identificação Pessoal Utilizando Técnicas de Reconhecimento e Verificação Facial Automáticas**, 1997. Dissertação (Mestrado). Departamento de Comunicações, Faculdade de Engenharia Elétrica e de Computação, UNICAMP, Campinas – SP, Brasil.
- [42] KING, I., XU, L. Localized Principal Component Analysis Learning for Face Feature Extraction and Recognition. In: WORKSHOP ON 3D COMPUTER VISION '97, 1997, Hong Kong. **Proceedings**: Hong Kong, p-124-128.
- [43] KEREN, D., OSADCHY, M., GOTSMAN, C. **Antifaces: A Novel, Fast Method for Image Detection**. In: IEEE Transactions on Pattern Analysis and Machine Intelligence, v.23, p-747-761, July 2001.
- [44] PONTIL, M., VERRI, A. **Support Vector Machines for 3D Object Recognition**. In: IEEE Transactions on Pattern Analysis and Machine Intelligence, v.20, p-637-646, June 1998.
- [45] HEL-OR, Y., HEL-OR, H. Real Time Pattern Matching Using Projection Kernels. In: Manuscript, (www.faculty.idc.ac.il/toky/Publications/publications.htm), Sept. 2002.
- [46] POULARIKAS, A. *The Handbook of Formulas and Tables for Signal Processing*. Boca Raton, CRC Press, 1999.
- [47] PRESS, W., TEUKOLSKY, S., VETERLLING, W., *at al. Numerical Recipes in C*. Second edition. New York: Cambridge University Press, 1992.
- [48] GRAY, R., DAVISSON, L. *An Introduction to Statistical Signal Processing* (www-ee.Stanford.edu/~gray/sp.html).
- [49] BIRK, H. **Recognizing Gestures From the Hand Alphabet Using Principal Component Analysis**, 1996. Dissertação (Mestrado). Laboratory of Image Analysis, Aalborg University, Denmark.
- [50] SIMONCELLI, E. Least Square Optimization. In: Manuscript, (<http://www.cns.nyu.edu/~eero/math-tools/>) Oct. 2001.
- [51] BAZARAA, M. *Nonlinear Programming*. New York : John Wiley & Sons, 1979.