

UNIVERSIDADE FEDERAL DE CAMPINA GRANDE
CENTRO DE ENGENHARIA ELÉTRICA E INFORMÁTICA
COORDENAÇÃO DE PÓS-GRADUAÇÃO EM INFORMÁTICA

Dissertação de Mestrado

Detecção de Olhos em Imagens com Faces Humanas

Bruno de Brito Leite

Campina Grande

Maio de 2008

UNIVERSIDADE FEDERAL DE CAMPINA GRANDE
CENTRO DE ENGENHARIA ELÉTRICA E INFORMÁTICA
COORDENAÇÃO DE PÓS-GRADUAÇÃO EM INFORMÁTICA

Detecção de Olhos em Imagens com Faces Humanas

Bruno de Brito Leite

Dissertação submetida à Coordenação do Curso de Pós-Graduação em Informática do Centro de Engenharia Elétrica e Informática da Universidade Federal de Campina Grande – Campus I como parte dos requisitos necessários para obtenção do grau de Mestre em Ciência da Computação (MSc).

Área de Concentração: Ciência da Computação

Linha de Pesquisa: Modelos Computacionais e Cognitivos

Herman Martins Gomes
João Marques de Carvalho

Orientadores

Campina Grande

Maio de 2008

FICHA CATALOGRÁFICA PREPARADA PELA BIBLIOTECA CENTRAL DA UFCG.

L533d Leite, Bruno de Brito.

Detecção de olhos em imagens com faces humanas / Bruno de Brito Leite. -
Campina Grande, 2008.

109f. : il. Col.

Dissertação (Mestrado em Ciência da Computação) - Universidade Federal
de Campina Grande, Centro de Engenharia Elétrica e Informática.

Referências.

Orientadores: Ph.D Herman Martins Gomes, Ph.D João Marques de Carvalho.

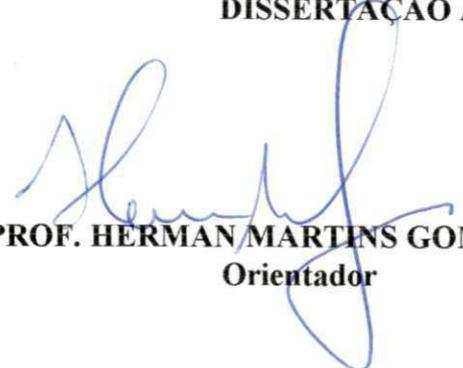
1. Detecção de olhos. 2. Processamento de imagens.
3. Visão computacional. 4. Detecção de padrões . I. Título.

CDU 004.81:159.953.5 (043)

“DETECÇÃO DE OLHOS EM IMAGENS COM FACES HUMANAS”

BRUNO DE BRITO LEITE

DISSERTAÇÃO APROVADA EM 30.05.2008



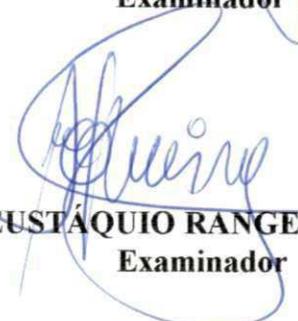
PROF. HERMAN MARTINS GOMES, Ph.D
Orientador



PROF. JOÃO MARQUES DE CARVALHO, Ph.D
Orientador



PROF. CLAUDIO ROSITO JUNG, Dr.
Examinador



PROF. JOSÉ EUSTÁQUIO RANGEL DE QUEIROZ, D.Sc
Examinador

CAMPINA GRANDE – PB

Agradecimentos

Ao meu pai Luiz Leite que, embora não esteja mais conosco, foi o grande motivador desta conquista.

À minha mãe Alda e a minha irmã Luciana, por todo o apoio e a força dada para que esse objetivo de vida pudesse ser realizado.

Aos orientadores Herman Martins Gomes e João Marques de Carvalho, pela dedicação e paciência durante a orientação.

À professora Patrícia Machado, que indicou o caminho que me trouxe até aqui.

Aos meus amigos do projeto Iphotobot (Claúdio, Luana, Luciana, Eanes, Walter, Rodrigo, Thiago, Felipe, Einstein, Paulo, Eduardo, Odilon, Xycho e Vinicius), pelos momentos de descontração durante a realização deste trabalho.

Aos membros da banca examinadora, pelas críticas e sugestões que contribuiram para o enriquecimento desta dissertação.

A equipe da COPIN (professores e funcionários), por garantir a infra-estrutura necessária para os mestrandos em Ciência da Computação da UFCG.

Aos meus “irmãozinhos” (Bruno Alexandre, Saulo e Fernando), pelos auxílios e dicas tão importantes na etapa final.

A Antônio Jaime, Emanuela Gadelha e Laisa Helena, pela amizade fortalecedora nos momentos difíceis ao longo desta caminhada.

Este trabalho foi desenvolvido em colaboração com a HP Brasil P&D.

Resumo

Existem diferentes situações em que a detecção de olhos em imagens desempenha um papel fundamental, como por exemplo: interação homem-máquina, determinação de características faciais, monitoramento de atenção de condutores de veículos, sistemas de identificação através de faces ou íris, análise de expressões faciais e sistemas de vídeo conferência. Esta dissertação apresenta uma revisão bibliográfica discutindo trabalhos recentes relacionados ao problema de detecção de olhos. Para cada trabalho, são apresentados: a idéia geral utilizada pelos autores, o desempenho obtido e a base de imagens de face considerada (quando estes dados estão disponíveis). A principal contribuição desta dissertação é a proposição de uma nova abordagem para determinação da posição de olhos em imagens de faces. A etapa de pré-processamento desta abordagem inclui compensação de iluminação com filtragem homomórfica e operações pontuais de expansão e equalização de histograma para melhoria de brilho e contraste. A etapa de detecção utiliza: abordagens passivas, baseadas na (i) análise da aparência, que consideram características de cor de pele; (ii) em aprendizagem, que utilizam uma rede neural treinada com características extraídas de exemplos e contra exemplos de olhos; e (iii) em modelos, que produzem uma resposta derivada da avaliação de uma função de casamento de modelos. Estas abordagens são integradas através de três regras de combinação de classificadores (produto, média e *ranking*). A viabilidade técnica da abordagem proposta é validada através da realização de experimentos. Estes experimentos demonstram que a abordagem proposta apresenta desempenho superior aos trabalhos existentes que foram incluídos na revisão bibliográfica.

Abstract

There is a number of application scenarios in which eye detection plays a major role, such as: human-computer interaction, facial feature tracking, driver vigilance monitoring, face and iris recognition, facial expression analysis, and video conferencing systems. This dissertation presents a bibliographic review of recent eye detection related works. For each reviewed work, a description is presented, containing the general approach adopted by the authors, the achieved performance, and the image face database considered in the experiments (when information on this data is available). The main contribution of this dissertation is the proposition of a new approach for eye detection in face images. The pre-processing step includes illumination compensation using homomorphic filtering and histogram operations such as stretching and equalization for brightness and contrast improvement. The detection step uses: passive approaches based on (i) appearance analysis, regarding skin tone features; (ii) learning, which uses a neural network trained using features extracted from examples and counter examples of eyes; and (iii) template matching, which produces response from evaluation of template matching function. These approaches are integrated by three classifier combination rules (product, mean, and ranking). Experimental results are the proof of concept for the proposed approach. In these experiments, the performance of the proposed approach overcome the performance of existing approaches discussed in the literature review.

Conteúdo

1	Introdução	1
1.1	Motivações	1
1.2	Objetivos e relevância	3
1.3	Estrutura da dissertação	4
1.4	Notação adotada	5
2	Revisão Bibliográfica	6
2.1	Métodos ativos	6
2.2	Métodos passivos	7
2.2.1	Abordagens baseadas na análise da aparência	8
2.2.2	Abordagens baseadas em aprendizagem	13
2.2.3	Abordagens baseadas em modelos	16
2.3	Sumário	18
3	Abordagem proposta	21
3.1	Detecção de faces	22
3.2	Compensação de iluminação	24
3.3	Pré-processamento com operações pontuais	28
3.3.1	Expansão de histograma	29
3.3.2	Equalização de histograma	33
3.4	Limitação do espaço de busca	35
3.5	Geração de mapa de evidência baseado na análise da aparência	40
3.6	Geração de mapa de evidência baseado em aprendizagem	45
3.7	Geração de mapa de evidência baseado em modelos	53

3.8	Combinação de mapas de evidência	56
3.9	Sumário	58
4	Implementação da abordagem proposta	60
4.1	Organização do sistema	61
4.2	Funcionamento do sistema	63
4.3	Sumário	68
5	Experimentos	71
5.1	Bases de imagens de faces	71
5.2	Tolerância a erro	74
5.3	Avaliação do desempenho com uso da compensação de iluminação	75
5.4	Comparativo entre operações pontuais para melhoria de brilho e contraste .	78
5.5	Comparativo entre a abordagem proposta e sistemas existentes	81
5.6	Comentários finais	84
5.7	Sumário	87
6	Considerações finais	88
6.1	Sumário da dissertação	88
6.2	Considerações e contribuições do trabalho realizado	89
6.3	Trabalhos futuros	90
A	Resultados da avaliação do desempenho com uso da compensação de iluminação	100
B	Resultados do comparativo entre operações pontuais para melhoria de brilho e contraste	103
C	Resultados apresentados pela abordagem proposta utilizando as bases de faces Caltech e JAFFE	106

Lista de Tabelas

2.1	Sumário de abordagens para detecção de olhos baseadas na análise da aparência.	13
2.2	Sumário de abordagens para detecção de olhos baseadas em aprendizagem.	17
2.3	Sumário de abordagens para detecção de olhos baseadas em modelo.	18
2.4	Base de imagens utilizadas e desempenho dos artigos revisados.	19
3.1	Estatísticas das coordenadas normalizadas	38
3.2	Distribuição da base de imagens de olhos e não-olhos.	49
3.3	Distribuição da base de imagens de olhos e não-olhos.	52
3.4	Funções de combinação utilizadas no módulo de combinação de mapas de evidência.	58
4.1	Relação entre procedimento proposto e respectivo pacote de implementação.	70
5.1	Características das base de imagens utilizadas nos experimentos.	74
5.2	Resultados obtidos com compensação de iluminação lateral na base IMM. .	78
5.3	Resultados obtidos com pré-processamento para melhoria de brilho e contraste na base IMM.	80
5.4	Tabela comparativa de desempenho entre a abordagem proposta e as existentes.	83
5.5	Tempos de execução e memória requerida para a detecção de olhos em 10 imagens utilizando a abordagem proposta.	85
A.1	Resultados do experimento com uso da compensação de iluminação.	100
B.1	Resultados do experimento comparativo para melhoria de brilho e contraste.	103
C.1	Exemplos de detecções de olhos na base Caltech.	106

C.2 Exemplos de detecções de olhos na base JAFFE. 108

Lista de Figuras

3.1	Diagrama com a arquitetura geral do sistema proposto.	21
3.2	Diagrama com a arquitetura do módulo de detecção de faces.	23
3.3	Formação da imagem a partir da reflectância e iluminação.	25
3.4	Diagrama com a arquitetura do módulo de compensação de iluminação. . .	26
3.5	Representações no domínio do espaço e da frequência.	28
3.6	Função de transferência do filtro passa-alta utilizado na filtragem homomórfica.	29
3.7	Exemplo de compensação de iluminação.	30
3.8	Diagrama do módulo de pré-processamento com operações pontuais.	30
3.9	Diagrama do módulo de expansão de histograma.	30
3.10	Exemplo de expansão linear de histograma em imagem com excesso de iluminação.	32
3.11	Exemplo de expansão linear de histograma em imagem com pouca iluminação.	32
3.12	Expansão linear de histograma em imagem com pouca iluminação.	33
3.13	Diagrama do módulo de equalização de histograma.	34
3.14	Exemplo de equalização de histograma.	35
3.15	Equalização de histograma.	35
3.16	Diagrama do módulo de limitação do espaço de busca.	36
3.17	Exemplos de falsa detecção de faces.	37
3.18	Exemplos de dupla detecção de mesma face.	37
3.19	Diagrama ilustrativo da definição do espaço de busca.	40
3.20	Exemplos de espaço de busca de olhos em imagens de faces.	40
3.21	Diagrama do módulo de geração de mapa de evidência baseado na análise da aparência.	41
3.22	Imagens originais sem rotulamento.	42

3.23	Imagens com regiões rotuladas.	42
3.24	Arquitetura básica de um neurônio artificial.	43
3.25	Organização da rede neural utilizada para detecção de pele.	44
3.26	Mapas de evidência gerados com análise da aparência.	45
3.27	Diagrama do módulo de geração de mapa de evidência baseado em aprendi- zagem.	46
3.28	Organização da grade de filtragem de olhos.	47
3.29	Exemplos de olhos e não-olhos utilizados no treinamento da rede neural. . .	49
3.30	Mapas de evidência baseados em aprendizagem.	53
3.31	Diagrama do módulo de geração de mapa de evidência baseado em modelo. . .	54
3.32	Arranjo para obtenção da correlação.	55
3.33	Padrão de olho utilizado na etapa de casamento.	55
3.34	Mapas de evidência baseados em modelos.	56
3.35	Diagrama do módulo de combinação de mapas de evidência.	57
4.1	Diagrama de <i>deployment</i> do sistema implementado.	62
4.2	Diagrama de classes do sistema implementado.	62
4.3	Diagrama da sequência de execução do programa principal.	64
4.4	Diagrama de sequência do pré-processamento.	65
4.5	Diagrama de sequência da filtragem homomórfica.	66
4.6	Diagrama de sequência de geração e combinação de mapas de evidência. . .	67
4.7	Diagrama de sequência da geração de mapa de evidência baseada na análise de aparência.	67
4.8	Diagrama de sequência da geração de mapa de evidência baseado em apren- dizagem.	68
4.9	Diagrama de sequência da geração de mapa de evidência baseado em modelos. .	69
5.1	Exemplos de imagens contidas na base IMM.	73
5.2	Exemplos de imagens contidas na base Caltech.	73
5.3	Exemplos de imagens contidas na base JAFFE.	74
5.4	Fluxograma da geração de estatísticas e métricas.	76
5.5	Mapas de evidências obtidos com a utilização da compensação de iluminação. .	77

5.6	Mapas de evidências obtidos sem a utilização da compensação de iluminação.	77
5.7	Comparação de desempenho relativa ao uso de compensação de iluminação.	79
5.8	Comparação de desempenho relativa à melhoria de brilho e contraste. . . .	82
5.9	Detecções de olhos na base Caltech produzidas pela abordagem proposta. .	84
5.10	Detecções de olhos na base JAFFE produzidas pela abordagem proposta. . .	85
5.11	Exemplos de erros nas detecções de olhos.	86

Lista de Siglas e Abreviaturas

- AF: *Affine Transformation*
- AMD: *Advanced Micro Devices*
- API: *Application Program Interface*
- DPF: *Dark Pixel Filter*
- FANN: *Fast Artificial Neural Network Library*
- FCF: *Face Circle Fitting*
- FFT: *Fast Fourier Transform*
- GC: *Garbage Collector*
- GMM: *Gaussian Mixture Model*
- GPF: *Generalized Projection Function*
- HMM: *Hidden Markov Models*
- HSI: *Hue, Saturation and Intensity*
- HSV: *Hue, Saturation and Value*
- ICA: *Independent Component Analysis*
- ICR: *Illumination Compensation based on the Multiple Regression Model*
- IDE: *Integrated Development Environment*
- IFFT: *Inverse Fast Fourier Transform*

- IPF: *Integral Projection Function*
- JPEG: *Joint Photographic Experts Group*
- MLP: *Multilayer Layer Perceptron*
- MSE: *Mean Square Error*
- OpenCV: *Intel Open Source Computer Vision Library*
- PCA: *Principal Component Analysis*
- RAM: *Read Access Memory*
- RBF: *Radial Basis Function*
- RGB: *Red, Green and Blue*
- RND: *Recursive Nonparametric Discriminant*
- SML: *Simplified Maximum Likelihood*
- SNNS: *Stuttgart Neural Network Simulator*
- SSE: *Sum Squared Error*
- SVM: *Support Vector Machines*
- TPS: *Topographic Primal Sketch*
- UML: *Unified Modelling Language*
- VPF: *Variance Projection Function*

Capítulo 1

Introdução

A tarefa de detecção de olhos em imagens contendo faces humanas cumpre um papel importante em diversos sistemas computacionais. Como os olhos são uma das características mais salientes em faces humanas, eles têm uma função muito importante na interpretação de desejos, necessidades e estados emocionais em sistemas de análise automática de expressões faciais [JWDF05]. A geometria específica, as características de contraste e o movimento dos olhos provêm informações úteis em diversos sistemas para detecção e reconhecimento de faces e íris, bem como em alguns sistemas de interação homem-máquina e de videoconferência. Exemplos de situações reais que caracterizam a necessidade por mecanismos de detecção de olhos são apresentados na Seção 1.1.

Nesta dissertação, são discutidos os trabalhos investigativos do uso e da combinação de diversas estratégias para determinação da posição dos olhos em uma imagem de face. Descrições detalhadas do problema a ser resolvido e dos objetivos a serem alcançados são apresentados na Seção 1.2. Finalmente, na Seção 1.3 é apresentada uma descrição da estrutura desta dissertação.

1.1 Motivações

Existem diferentes situações em que a detecção de olhos a partir de imagens desempenha um papel fundamental, como por exemplo: interação homem-máquina, determinação de características faciais, monitoramento de atenção de condutores de veículos, sistemas de identificação através de faces ou íris, análise de expressões faciais e sistemas de vídeo conferência.

A seguir, são discutidas em maiores detalhes as principais destas situações.

O computador tem se apresentado como ferramenta essencial de comunicação, educação e entretenimento para pessoas com deficiências físicas e dificuldades motoras. Entretanto, o custo e a metodologia de utilização desencorajam o uso de dispositivo apontador (*mouse*) com adaptação para pessoas com necessidades especiais. Alternativas de baixo custo que permitam a interação do usuário com a interface gráfica de computadores podem ser implementadas através da detecção e acompanhamento do movimento da cabeça e dos olhos [MBZ02].

A localização e o rastreamento dos olhos em tempo real podem servir como base para determinação de outras características faciais (como a posição das sobrancelhas e boca). Estas características podem ser utilizadas na geração sintética de modelos de faces em três dimensões (3D), para interação em ambientes virtuais [GEZMT99].

O número de acidentes nas estradas provocado pela falta de atenção dos motoristas tem causado sérios prejuízos à sociedade. A falta de atenção do condutor provoca a redução da sua percepção e do controle do veículo, colocando em risco a vida do condutor e de outras pessoas. Este fato tem motivado o desenvolvimento de sistemas de monitoramento de atenção de condutores de veículos, que emitem alertas em caso de situações de insegurança. O cansaço, uma das causas da falta de atenção, pode levar as pessoas a apresentar mudanças de comportamento, observáveis através de alterações de características faciais. A partir da observação do comportamento dos olhos é possível determinar tais situações de insegurança [JY01].

As características fisiológicas da face têm sido utilizadas em sistemas de segurança para identificação de indivíduos. A própria face é uma característica biométrica que os humanos utilizam no cotidiano para reconhecimento de pessoas. A geometria específica e o contraste da íris são características biométricas que exibem individualidade a ponto de permitir à aplicação em sistema de identificação de pessoas. Um dos passos iniciais destes sistemas de reconhecimento é a localização dos olhos que possibilitam a normalização de tamanho e alinhamento de faces ou íris [BKRV03].

Como uma das mais salientes características da face humana, os olhos representam uma importante função na análise automática de expressões faciais, pois apresentam alguns pontos fiduciais bastante estáveis em comparação com outras características faciais. Por isto,

muitos sistemas de reconhecimento de expressões detectam primeiramente os olhos para posterior detecção de outras características faciais (como sobrancelhas e boca), visto que a localização destes últimos pode ser estimada utilizando as coordenadas dos olhos como referencial [PCRK05].

Em sistemas de vídeo conferência, quando a câmera que captura a face do usuário não está alinhada com a mesma, ocorre um fenômeno chamado “falta de contato com os olhos”. Se a câmera estiver montada acima do monitor, a face aparenta estar olhando para baixo. Por outro lado, a face vai parecer estar olhando para cima se a câmera estiver montada embaixo do monitor [Yip05]. Através da detecção de olhos é possível enquadrar corretamente a face dentro do quadro e ainda selecionar regiões que devem receber uma fraca compressão (pouca perda de informação visual) no procedimento de compactação utilizado em sistema de vídeo conferência.

Vale salientar que a necessidade por mecanismos de localização dos olhos não se resume apenas as aplicações apresentadas, porém uma enumeração completa de todas as aplicações possíveis fugiria ao escopo deste trabalho.

1.2 Objetivos e relevância

Diante do contexto apresentado na seção anterior, o objetivo geral desta dissertação é propor uma nova solução passiva e não intrusiva (definições destes termos se encontram na Seção 2.2) para o problema de detecção de olhos em imagens com faces humanas, que apresente um desempenho superior, em termos da taxa de detecção, aos obtidos com o uso de técnicas existentes.

Entre os objetivos específicos, cabe mencionar os seguintes: (a) analisar a aplicação de técnicas de pré-processamento de imagens, verificando seu impacto na melhoria do desempenho na detecção de olhos; (b) propor um sistema híbrido de detecção composto por três estratégias distintas e (c) investigar abordagens para combinação de resultados de vários detectores de olhos independentes.

A relevância desta dissertação é justificada pelo aspecto inovador do uso de técnicas para melhoria da detecção de olhos em imagens com a presença de iluminação lateral, da utilização de operações pontuais para melhoria de desempenho em situações com imagens

capturadas com sub-exposição (imagens escuras) ou super-exposição (imagens muito claras) e da combinação de diversas estratégias de classificação (esta configuração permite ocultar situações em que a utilização de evidências de uma única estratégia não apresentaria um resultado satisfatório). A aplicação destes procedimentos visou sempre superar as dificuldades apresentadas pelas técnicas já existentes.

1.3 Estrutura da dissertação

Este trabalho se encontra estruturado conforme discutido a seguir. No Capítulo 2, são apresentadas as categorias e estratégias mais usuais na classificação de trabalhos relacionados à detecção de olhos. Além disto, o Capítulo 2 ainda contém uma revisão bibliográfica de artigos relevantes sobre este tema, produzidos nos últimos dez anos.

A abordagem proposta para detecção de olhos é descrita no Capítulo 3. Esta descrição é feita de forma modular e, sempre que possível, composta de: (a) princípio de funcionamento, técnica e/ou metodologia utilizada na construção de cada módulo; (b) explanação sobre o embasamento matemático (fórmulas e parâmetros necessários) presentes em cada módulo; (c) objetivos esperados com a aplicação das operações presentes em cada módulo; (d) exemplos de imagens utilizadas como entrada para cada módulo e as respectivas imagens geradas após o processamento.

O sistema implementado com base na abordagem proposta é descrito no Capítulo 4. Aspectos como a organização interna em termos de pacotes, relacionamentos e atributos dos principais componentes e fluxo de processamento são ilustrados através de diagramas UML (*Unified Modelling Language*).

A viabilidade da abordagem proposta é validada através de comparação com algumas das abordagens apresentadas no Capítulo 2. Os resultados dos experimentos comparativos são exibidos no Capítulo 5.

No Capítulo 6, são apresentadas as conclusões obtidas a partir deste estudo, destacadas as principais contribuições e sugeridos trabalhos futuros, que podem ser derivados a partir do exposto ao longo desta dissertação.

1.4 Notação adotada

Para facilitar a leitura e o entendimento do texto e garantir a organização e a consistência da descrição das abordagens apresentadas, a notação adotada e aplicada ao longo de todo este trabalho é a seguinte:

- valores escalares: são representados por letras minúsculas do alfabeto ocidental. Ex: x, y, w e h ;
- vetores e matrizes: são representados por letras minúsculas do alfabeto grego. Ex: $\alpha, \beta, \gamma, \delta$. Os elementos de uma matriz são acessados pela letra correspondente ao vetor ou à matriz seguida de dois ou mais pares de números limitados por parênteses. Ex: $\alpha(1), \beta(2), \gamma(1, 2)$ e $\delta(1, 2, 3)$. Imagens são tratadas como matrizes multidimensionais;
- funções: são representadas por um nome composto de letras minúsculas do alfabeto ocidental seguida dos argumentos recebidos entre parênteses.

Capítulo 2

Revisão Bibliográfica

Geralmente, os métodos para detecção de olhos em imagens são classificados em duas categorias: ativos ou passivos. Os métodos ativos se caracterizam pela detecção de olhos através da análise de vídeo ou sequência de imagens, enquanto a detecção de olhos com métodos passivos requer uma única imagem. Esta última classe pode ainda ser dividida em três tipos, de acordo com a estratégia empregada: (a) análise da aparência, (b) aprendizagem e (c) modelos [JWDF05]. Na Seção 2.1, apresentam-se detalhes e limitações da utilização de métodos ativos e a Seção 2.2 descreve e analisa as diferenças entre as diversas estratégias utilizadas pelos métodos passivos. Finalmente, na Seção 2.3, apresentam-se um sumário dos trabalhos revisados neste capítulo.

2.1 Métodos ativos

A detecção de piscadas de olhos e a utilização de iluminação infravermelha intermitente são técnicas frequentemente utilizadas na detecção de olhos a partir de vídeo. A detecção de olhos através de piscadas é feita com a busca por padrões temporais que surgem nas imagens resultantes da subtração de quadros consecutivos em uma sequência de vídeo [BKR03][MBZ02]. A detecção de olhos com iluminação infravermelha é feita a partir de um feixe de luz dirigido sobre os olhos. A córnea reflete a luz, produzindo o efeito conhecido como “olhos vermelhos”, muito comum em fotos produzidas com o uso de *flash*. Este fenômeno produz pupilas com alto brilho em imagens em níveis de cinza. A posição dos olhos é determinada através destas regiões de alto brilho [YC05][AZVK05].

A necessidade de uma sequência de vídeo ou de aparelhagem especial (diodos emissores de luz infravermelha e câmeras sensíveis a este tipo de iluminação), bem como restrições no cenário de captura da foto (não deve haver superfícies refletoras deste tipo de iluminação) tornam impraticáveis ou limitam o uso de métodos ativos em imagens capturadas previamente sem estes requisitos [TB05].

2.2 Métodos passivos

Conforme apresentado no início deste capítulo é possível dividir os métodos passivos em 3 categorias, conforme o tipo de estratégia utilizada: (a) análise da aparência; (b) aprendizagem; e (c) modelos [JWDF05]. Na abordagem baseada na análise da aparência, características tais como cor, textura e formas presentes na imagem são consideradas. Através de análises estatísticas da pele humana em imagens, agrupamentos de *pixels* com cor de pele são definidos em algum espaço cromático, nos trabalhos de Schmuege et al [SJST07] e Kakumanu et al [KMB07] são apresentadas extensas revisões bibliográficas a respeito de definições de agrupamento de cor de pele em diversos espaços de cores. Os limites destes agrupamentos são usados como limiares para binarização de imagens de faces humanas. Em seguida, uma análise de componentes conectados limitados por região de pele é feita utilizando o resultado da binarização. Regras de conhecimento são aplicadas para eliminação de falsos candidatos e determinação da real posição dos olhos. Na maioria das vezes, a detecção de olhos baseada na análise da aparência requer poucos recursos computacionais, devido à redução do espaço de busca e à simplicidade de cálculo de características invariantes à rotação e escala, como cores e formas. A principal desvantagem desta abordagem é a dificuldade de adotar um critério satisfatório para identificar regiões de pele, que seja robusto a variações étnicas e mudanças de iluminação [TB05].

Para detecção baseada em aprendizagem, um conjunto de imagens de exemplos e contra exemplos são utilizados para treinamento de um classificador. Os sistemas que utilizam este tipo de abordagem usualmente diferem nas características usadas para treinamento, nas técnicas empregadas para extração destas características a partir das imagens e no tipo de classificador adotado, como por exemplo: Redes Neurais, Modelos Escondidos de Markov (*Hidden Markov Models* - HMM), Máquinas de Vetores de Suporte (*Support Vector Machi-*

nes - SVM), etc. Os olhos correspondem às regiões da imagem que produzem uma resposta positiva do classificador. Esta abordagem pode apresentar invariância a rotações, mudanças de escala e iluminação, através da geração sintética de exemplos e contra exemplos para utilização no treinamento. A dificuldade deste procedimento se encontra na falta de critérios específicos para geração de contra exemplos em imagens genéricas [TB05].

Detecção baseada em modelos procura minimizar uma função de energia, obtida através do casamento de um dado modelo com a imagem. O padrão normalmente é composto por um círculo e duas curva parabólicas, desta forma modelando o aspecto de um olho. Os métodos que utilizam esta abordagem diferem basicamente pelo modo de encontrar a localização inicial do padrão. Na busca pelos olhos, o modelo é deslocado de acordo com um processo de minimização de uma função de energia. O deslocamento do modelo permite a detecção até mesmo de olhos rotacionados e apresenta tolerância a variações de abertura dos olhos. Algumas desvantagens do uso desta abordagem são as dificuldades para encontrar os pontos iniciais para encaixe do modelo e o alto custo computacional [TB05].

Vale destacar uma categorização pouco comum, encontrada no trabalho de Yoo [YC05], em que os métodos de detecção de olhos são classificados em intrusivos e não intrusivos. Os métodos intrusivos são caracterizados pela utilização de aparato auxiliar (como lentes de contato especiais, adesivos coloridos e emissores eletromagnéticos). O desconforto causado por estes aparatos fazem com que os métodos intrusivos, em geral, não sejam muito utilizados.

As subseções seguintes apresentam uma revisão bibliográfica de trabalhos recentes sobre detecção de olhos utilizando a abordagem passiva, fortemente relacionados com o método proposto nesta dissertação. Esta revisão apresenta abordagens baseadas na análise da aparência, em aprendizagem e em modelos, respectivamente. A escolha pela abordagem passiva é justificada pela maior abrangência da mesma, que permite o processamento de imagens e quadros de vídeo capturados previamente sem a utilização de nenhum aparato auxiliar (como feixe de iluminação infravermelha) que facilite a localização dos olhos.

2.2.1 Abordagens baseadas na análise da aparência

Huang et al [HSLW98] propuseram uma combinação de abordagens para solucionar o problema da detecção robusta de faces e olhos. Em imagens de faces de baixa resolução, os

olhos e as sobrancelhas se apresentam como barras escuras que podem ser detectadas utilizando a segunda derivada de um filtro Gaussiano. Por isso, Huang et al [HSLW98] aplicaram filtros Gaussianos e direcionais para detectar candidatos a olhos, sobrancelhas, boca e nariz. Após este passo, um modelo estrutural é criado para agrupar pontos característicos de candidatos à face usando relações geométricas. Na face detectada, a região dos olhos/sobrancelhas é determinada utilizando a estrutura da face. Para olhos abertos, o *pixel* com menor intensidade é considerado o centro do olho (pupila). Para olhos fechados, uma linha ou uma curva escura é definida e o ponto médio deste elemento determina o centro do olho. Em experimentos realizados com uma base proprietária formada por 680 imagens de 30 pessoas, esta abordagem apresentou falha na detecção de olhos em 14 imagens (97,4% de acerto). Estas falhas foram causadas pela inclinação de algumas faces e problemas de iluminação (iluminação lateral e pouca iluminação).

Um algoritmo baseado em conhecimento para detecção de olhos foi apresentado por Zhang e Lenders [ZL00]. O algoritmo pode ser descrito da seguinte forma: dada a localização inicial da face, limiares de binarização são definidos baseados na distribuição bimodal presente no histograma da face (a binarização é simplificada assumindo que a imagem contém apenas uma única face humana). O processo de binarização separa as regiões escuras (cabelos, sobrancelhas e olhos) das regiões de pele. As regiões com os olhos são localizadas usando regras baseadas em informações da posição dos cabelos. Os olhos estão contidos em regiões próximas da linha imaginária que liga o fim das duas costeletas. Após a seleção das regiões dos olhos, detecção de bordas e regras de conhecimento são aplicadas para localizar a íris e as pálpebras, determinando os quatro cantos de um retângulo que limita o olho. As regras de conhecimento utilizam o fato das bordas dos olhos apresentarem um formato arredondado e repetição em relação ao eixo horizontal, devido à simetria dos olhos direito e esquerdo. Zhang e Lenders [ZL00] declararam que a aplicação do método proposto na base de faces Yale (a base de faces Yale contém 11 imagens de 15 sujeitos em uma variedade de condições incluindo presença de óculos, variações de iluminação e alterações de expressões faciais [Gro05]) mostrou-se bastante rápida e eficiente em relação ao tempo de execução e as taxas de detecção, porém desvantagens e resultados numéricos não foram apresentados pelos mesmos.

Um sistema de detecção de olhos sob condições variadas de iluminação foi proposto por

Samad et al [SHT01]. Uma grade com 64 (8 x 8) células é construída com a intensidade média dos *pixels* contidos em cada célula. Esta grade desliza pela face, definindo candidatos a olhos através de regras que limitam a diferença de intensidade entre células de uma mesma grade. Em uma segunda etapa, um mapa de contorno é gerado a partir dos candidatos encontrados. Regiões com o maior número de contornos indicam a presença de olhos. Pode-se fazer analogia do mapa de contorno com um mapa topográfico onde os contornos definem regiões com a mesma altura (no mapa de contorno, altura é substituída por intensidade do nível de cinza). A etapa final é a detecção dos olhos através da eliminação sucessiva dos candidatos a olhos utilizando relações de distância vertical e horizontal entre os lados da face (direito e esquerdo). Experimentos, realizados com 200 imagens da base Olivetti (a base de imagens Olivetti contém 400 imagens de 10 pessoas diferentes com variações na iluminação, nas expressões faciais e nos detalhes faciais, como óculos e barba [Gro05]), apresentaram índice de acerto na detecção de olhos de 94%. O algoritmo apresentou falhas na detecção devido ao posicionamento dos sujeitos em relação à câmera e às condições de iluminação nas imagens de face.

Kumar et al [KRR02] propuseram um algoritmo de três etapas: (a) definição de possíveis candidatos a olhos, (b) eliminação de candidatos e (c) determinação do centro dos olhos através de funções de projeção. A localização de possíveis candidatos a olhos é feita através de binarizações definidas a partir de modelos de cor de pele nos espaços HSV e RGB, combinados através da operação E (AND) lógico entre as duas binarizações. Os modelos de cor de pele foram definidos experimentalmente utilizando imagens de faces com uma ampla variedade étnica. A eliminação de candidatos a olhos é realizada através de quantização espacial com análise de componentes conectados e regras de conhecimento baseadas em relações antropométricas. Funções de projeção de média e de variância são aplicadas aos candidatos remanescentes para detecção do centro dos olhos. Os resultados obtidos se mostraram bastantes sensíveis a variações na iluminação que provocam alterações drásticas na aparência da pele, mas bastante promissores em situações de condições de iluminação controlada e com faces frontais.

Han et al [HKN02] não utilizaram técnicas de detecção de pele. Sua estratégia é baseada em morfologia matemática em níveis de cinza. Para detecção de olhos, primeiramente, é aplicada uma operação morfológica de fechamento na imagem de intensidade (níveis de

cinza) fornecida como entrada. Em seguida, um mapa de vales é obtido através da subtração da imagem resultante do fechamento morfológico pela imagem de intensidade de entrada. A geração deste mapa de vales permite a detecção de bochechas, boca e olhos através da utilização de um algoritmo iterativo de crescimento de regiões que é aplicado no mapa. Este algoritmo expande regiões elípticas posicionadas em lugares previamente estabelecidos comparando *pixels* claros e escuros entre regiões adjacentes. A combinação de áreas circulares com alto brilho e elípticas com baixo brilho satisfazendo um dado conjunto de condições indicam a localização de bochechas, boca e olhos. Os resultados experimentais, realizados com 450 imagens da base de faces Caltech [Web], apresentaram um taxa média de acerto de 98,8% na detecção de olhos. As imagens utilizadas contém faces de 26 sujeitos (homens e mulheres) apresentando variações de expressões faciais e iluminação. As falhas na detecção ocorreram, principalmente, devido à falta de iluminação adequada e fechamento dos olhos.

Para detecção do centro e dos limites dos olhos, Hua e Geng [ZHG04] definiram uma Função de Projeção Generalizada (*Generalized Projection Function* - GPF) que é calculada através da Função de Projeção Integral (*Integral Projection Function* - IPF) e da Função de Projeção de Variância (*Variance Projection Function* - VPF). A função GPF se caracteriza por herdar a robustez da IPF e a sensibilidade da VPF. Estas funções de projeção são comumente utilizadas em algoritmos de segmentação de imagens e têm o cálculo de somatórios de linhas (projeção horizontal) e colunas (projeção vertical) como ponto em comum e as regras de definição de transições entre dois segmentos como diferencial entre elas. Experimentos em base de faces mostram que todos os casos especiais de GPF são efetivos na detecção de olhos. Além disto, Zhou e Geng [ZHG04] perceberam que IPF é mais eficiente em bases de faces de ocidentais do que em bases de faces de orientais e o contrário é válido para VPF. Uma análise das detecções revelou que este comportamento se deveu a diferenças nas sombras causadas pelo nariz e pela depressão dos olhos de indivíduos de diferentes etnias. A avaliação da função proposta para detecção do centro dos olhos utilizou a base de faces NJUFace (a base de imagens NJUFace consiste de 359 imagens, com uma resolução de 380×285 *pixels* em escala de cinza, capturadas com variações de iluminação, expressão, pose e tamanho das faces de modelos chineses [ZHG04]) e obteve um taxa de correta detecção de 98,4%.

Fathi e Manzuri [FM04] detectaram pele utilizando agrupamentos definidos em um es-

paço de cor RG normalizado. Uma das principais vantagens da utilização do espaço de cor RG é o baixo custo computacional, devido à redução de um modelo de três dimensões (RGB) para um modelo de duas dimensões (RG normalizado), além da redução da sensibilidade a alterações nas condições de iluminação. A procura por candidatos a olhos é feita em regiões delimitadas pela pele detectada e com *pixels* de baixa intensidade, devido à presença de elementos mais escuros (como a íris e os cílios) em relação à pele. Candidatos a olhos são selecionados baseados em sua localização e através da aplicação de regras de conhecimento. Candidatos localizados na metade inferior da região de pele detectada e com área menor ou maior que determinados limiares são eliminados. Os candidatos restantes são validados por um filtro de variância de olhos. A avaliação de desempenho realizada com base em sequências de vídeo, geradas pelos autores, aponta um desempenho de 98% de acerto na detecção de olhos em condições ótimas. Os erros relatados se devem, em grande parte, a problemas de iluminação.

Para superar limitações impostas por iluminação e posicionamento da face no processo de detecção de olhos, Xingming e Huangyuan [XH06] propuseram um algoritmo baseado em normalização de iluminação que utiliza uma combinação de transformação *Affine* (*Affine Transformation* - AF) e compensação de iluminação baseada em modelos de regressão múltipla (*Illumination Compensation based on the Multiple Regression Model* - ICR) e em detecção de face utilizando um classificador *Adaboost*. Na face detectada, a probabilidade de ocorrência de cor de pele é utilizada para eliminar sombras, após o que, a área da pele é binarizada. Para fazer a remoção de falsos candidatos a olhos, regras básicas (incluindo limitações para tamanho do arco, largura, posição, centro de gravidade, ângulo e distância) são aplicadas. Finalmente, o par de olhos restante é verificado utilizando um classificador do tipo SVM. Em testes utilizando a base SCUT (Xingming e Huangyuan [XH06] não disponibilizaram referências para esta base), uma implementação para este algoritmo apresentou uma taxa de 96,2% de acerto na detecção de olhos. As falhas na detecção se deram, em grande parte, ao uso de óculos. A base SCUT é formada por 450 imagens de 50 indivíduos capturadas sob três condições diferentes de iluminação.

A Tabela 2.1 apresenta um sumário das características usadas nos artigos de detecção de olhos baseados na análise da aparência revisados. A primeira coluna apresenta os autores dos artigos; a segunda, a estratégia adotada para extrair as características utilizadas para encon-

trar os olhos e a terceira, o critério de decisão utilizado na eliminação de falsos candidatos.

Tabela 2.1: Sumário de abordagens para detecção de olhos baseadas na análise da aparência.

Autores	Estratégia	Decisão
Huang et al [HSLW98]	Filtros direcionais	Minímo local
Zhang e Lenders [ZL00]	Imagem binarizada	Localização
Samad et al [SHT01]	Mapa de contornos	Distância entre candidatos
Kumar et al [KRR02]	Filtros de pele	Altura e largura
Han et al [HKN02]	Morfologia Matemática	Intensidade de brilho
Hua e Geng [ZHG04]	Funções de projeção	Variação de projeção
Fathi e Manzuri [FM04]	Filtros de pele	Altura e largura
Xingming [XH06]	Morfologia Matemática	SVM

2.2.2 Abordagens baseadas em aprendizagem

Peng et al [PZB98] propuseram um método neural híbrido para localização de olhos humanos. O pré-processamento de imagens inclui três fases: redução de resolução, normalização para níveis de cinza e equalização de histogramas. Estas fases têm o objetivo de reduzir a quantidade de informação a ser processada e diminuir a sensibilidade às variações de iluminação. De acordo com os autores, uma Rede Neural de Base Radial (*Radial Basis Function - RBF*) foi utilizada devido ao rápido tempo de treinamento, boa generalização e simplicidade quando comparada a redes do tipo Perceptron de Multi Camadas (*Multi-Layer Perceptron - MLP*). A rede RBF serve como filtro entre a imagem de face de entrada e a imagem com o mapa de resposta apresentado pela rede. Os picos de intensidade presentes neste mapa são referenciados como possíveis regiões de olhos. Regras de conhecimento são utilizadas para validar as regiões candidatas a olhos. Estas regras de conhecimento consideram aspectos como a área, o tamanho e a inclinação ocupada pelo candidato no mapa, bem como a distância média entre os dois candidatos. No experimento reportado pelos autores, foi construída uma base de 300 imagens de faces capturadas de 18 estudantes chineses e uma implementa-

ção do método proposto foi aplicado nesta base, produzindo uma taxa de acerto de 97,41% de correta detecção.

No sistema robusto proposto por D’Orazio et al. [DLSG04] foi utilizado um algoritmo capaz de trabalhar com imagens complexas sem restrições de fundo (*background*) ou cor de pele e sem limitações para as regiões dos olhos. Um detector de círculos baseado em transformadas direcionais de *Hough* é aplicado na imagem inteira, para identificar regiões limitadas por bordas com configuração circular, similar à que se apresenta na íris. Como não se sabe o raio da íris a ser detectada, o mesmo algoritmo é aplicado diversas vezes, com variações do raio da configuração circular. As subimagens contendo o resultado do processo de detecção de círculos são aplicadas em um filtro *wavelet* cuja função é extrair características de baixas e altas frequências e organizá-las de forma hierárquica. A saída deste filtro é aplicada à entrada de uma rede neural treinada para detectar olhos em imagens. Para testar a eficácia, uma base de faces de 1474 imagens de seis pessoas diferentes (com olhos de cores diferentes) obtidas em diversos tamanhos e posições foi utilizada. Os resultados apresentaram 96% de taxa de correta detecção para imagens com olhos abertos. O desempenho do sistema não foi satisfatório para olhos semi-abertos (45%) ou fechados (0%) devido a falhas na detecção de regiões com configurações circulares.

Wu e Trivedi [WT05] propuseram um sistema para solucionar o problema da detecção de olhos em faces humanas através da utilização de estruturas estatísticas com representações em forma de árvore binária. Para extrair os padrões necessários que permitem a construção das árvores binárias, os autores utilizaram a técnica de Análise de Componentes Independentes (*Independent Component Analysis* - ICA). Esta técnica permite separar características condicionalmente independentes em diferentes sub-árvores, a partir de amostras de exemplos e contra exemplos. O algoritmo de k-médias é usado para agrupar padrões similares. A detecção dos olhos é feita através da extração de características de uma imagem e do encaminhamento na árvore binária de cima para baixo (*top-down*). As imagens utilizadas como exemplos de olhos foram extraídas da base FERET (a base FERET é formada por 14.051 imagens com resolução de 256×384 *pixels* em escala de cinza de 1199 sujeitos, capturadas com variações na expressão facial, iluminação e pose dos modelos [Gro05]) e os exemplos de não-olhos foram gerados a partir da base Caltech (a base de imagens Caltech contém 450 imagens frontais de faces de 27 sujeitos com variações de condições de iluminação e expres-

sões faciais [Web]). Este conjunto de exemplos e contra exemplos também foi utilizado para determinação dos limiares de proximidade mais adequados para indicar o agrupamento de características. A avaliação do desempenho utilizando a base FERET apresentou um índice de correta detecção de 92,43%. Um fato curioso é que o sistema é capaz de discernir a sutil diferença entre o olho esquerdo e direito, visto que foi treinado apenas com imagens de olho esquerdo e os resultados da avaliação apresentam apenas 5,99% de casos de detecção de olho direito (falso positivo).

Baseado em Discriminantes Não Paramétricos Recursivos (*Recursive Nonparametric Discriminant - RND*), Wang et al [WGJW05] propuseram o treinamento de classificadores probabilísticos e a combinação de múltiplos classificadores usando *Adaboost* para formar um detector de olhos, robusto e preciso segundo os autores. O método de localização dos olhos segue um procedimento hierárquico: a face é detectada (o método de detecção de faces também é baseado em *Adaboost*); restrições geométricas são aplicadas para localizar os olhos (restrições só permitem a busca na parte superior da face); um classificador é utilizado para indicar presença ou não de olhos. Para garantir robustez, diversas fontes foram utilizadas para gerar a base de imagens de olhos utilizada, principalmente a base FERET e imagens de olhos capturadas na *World Wide Web*. Além disto, o conjunto de olhos foi aumentado através da geração sintética de exemplos de olhos a partir da aplicação de pequenas rotações nas imagens de olhos inicialmente obtidas. Na prática, apenas o detector de olho esquerdo foi treinado devido à simetria das imagens. O olho direito é detectado através de rebatimento da imagem. Resultados mostram que taxas de acerto da ordem de 94,5% foram obtidas utilizando a base FRGC 1.0 [FRG]. Os erros na detecção ocorreram devido à iluminação inadequada.

Wang e Yin [WY05] derivaram um mapa topográfico a partir da imagem original em níveis de cinza usando análise de Esboço Topográfico Primário (*Topographic Primal Sketch*). Nesta técnica, os *pixels* em níveis de cinza de uma imagem são tratados como uma superfície em três dimensões (3D), sendo a altura de um determinado *pixel* definida por sua intensidade. A análise de esboço topográfico classifica as regiões como picos, valas, sulcos, colinas e planos. *Pixels* rotulados como valas no mapa de terreno são marcados como candidatos para seleção de pares. Janelas centradas nestes candidatos são extraídas e suavizadas com um filtro Chebyshev [CSH⁺00], para remover eventuais ruídos. Um Modelo de Mistura de

Gaussianas (*Gaussian Mixture Model* - GMM) foi treinado com mapas topográficos de 293 imagens e utilizado como classificador para selecionar o correto par de olhos entre todos os candidatos. Nos experimentos realizados, o detector foi testado com a base JAFFE (a base JAFFE contém 213 imagens com resolução de 256×256 pixels em modo monocromático de 10 modelos japonesas, capturadas com variações de expressões [LAKG98]). Os olhos foram corretamente localizados em 95.8% dos casos. Inclinações na face influenciaram negativamente no desempenho apresentado pelo detector.

Outra abordagem foi apresentada por Jin et al [JYS⁺06], que propuseram um detector híbrido utilizando características não apenas de um único olho, mas também do par formado. A classificação considera um conjunto de características de baixa dimensão que são produzidas projetando uma imagem normalizada de olhos em um auto-espaço ponderado (selecionado através de filtragem). O conjunto destas características é classificado utilizando a técnica de Máxima Verossimilhança Simplificada (*Simplified Maximum Likelihood* - SML) e SVM. Para treinar estes classificadores foram utilizadas imagens de olhos extraídas das bases FERET e BioID [Sca]. Para aumentar o conjunto de treinamento, foram geradas variações nas imagens de olhos através de rotações de -15° , -10° , -5° , 0° , $+5^\circ$, $+10^\circ$ e $+15^\circ$. Os classificadores foram treinados apenas com imagens de olhos esquerdos, sofrendo os olhos direito rebatimento (espelhamento na direção vertical) nos processos de treinamento e de detecção. Para reduzir a taxa de falsos positivos, uma configuração de par de olhos médio é calculada entre todos os candidatos que obtiveram alta resposta na classificação. O cálculo da média permite a inclusão de um componente espacial no resultado. O desempenho deste classificador foi avaliado na bases de faces FERET e obteve um índice de acerto médio de 90,75%.

Os artigos que utilizam técnicas baseadas em aprendizado descritos nesta dissertação são sumarizados na Tabela 2.2. A primeira coluna contém os autores do artigo; a segunda coluna apresenta as características usadas como entrada do processo de classificação e a terceira, o tipo de classificador utilizado.

2.2.3 Abordagens baseadas em modelos

No algoritmo proposto por Lin e Yang [LY04] a detecção de faces em uma imagem colorida foi feita através de um modelo de cor de pele humana experimentalmente definido no espaço

Tabela 2.2: Sumário de abordagens para detecção de olhos baseadas em aprendizagem.

Autores	Característica	Classificador
Peng et al [PZB98]	Imagem pura	RBF
D’Orazio et al. [DLSG04]	Wavelet	Rede Neural
Wu e Trivedi [WT05]	ICA	Bayesiano
Wang et al [WGJW05]	RND	AdaBoost
Wang e Yin [WY05]	Mapa topográfico	Modelo Gaussiano
Jin et al [JYS+06]	Projeção em auto-espaço	SVM

HSI (*Hue, Saturation e Intensity*). Entretanto, imprecisões na definição de regiões de pele produzem regiões que provocam erros na detecção dos olhos. Para contornar este problema, o método FCF (*Face Circle Fitting*) foi proposto. A função do método FCF é incluir como candidatos a olhos apenas aqueles contidos dentro do círculo delimitado pela face. A localização precisa e verificação dos olhos são feitas com o algoritmo DPF (*Dark Pixel Filter*). Este algoritmo define um modelo para os olhos em função de valores de intensidade para a pele, para os olhos e para regiões de transição entre a pele e os olhos. Regiões com *pixels* escuros (que não contêm pele) são comparadas com o modelo. O candidato que obtiver melhor resposta é tomado como olho. O algoritmo para detecção de olhos foi testado utilizando a base de faces HHI [HHI]. Esta base contém 206 imagens de sujeitos de diferentes grupos étnicos capturadas sob iluminação superior e lateral. O desempenho atingido foi de 74% de acerto. Apesar do desempenho apresentado ser baixo, o algoritmo apresentou velocidade de resposta bem maior do que outras abordagens existentes, o que ressalta sua indicação para sistemas de detecção de olhos em tempo real.

Shinjiro e Kawato [KT04] propuseram um algoritmo inovador para detecção dos olhos. Inicialmente, posições candidatas a olhos são identificadas através de limiares de intensidade. O caráter inovador deste algoritmo reside no fato da validação dos candidatos a olhos não ser feita através da comparação do respectivo candidato com um modelo de olho. O algoritmo utiliza na verdade um padrão que descreve a região contida entre os dois olhos (“*Between-the-Eyes*”, como descrito no artigo). Segundo Shinjiro e Kawato [KT04], este padrão apresentou características mais fáceis de serem detectadas do que o olho propriamente

dito. Experimentos com uma implementação do algoritmo proposto foram realizados. Os resultados indicam que o algoritmo apresenta falhas quando não é possível definir claramente a posição da pupila. Resultados numéricos e a base de faces utilizada nos experimentos não foram mencionados.

Um sumário de modelos utilizados nos trabalhos revisados que utilizam esta abordagem são apresentados na Tabela 2.3. Os autores dos artigos são apresentados na primeira coluna e o modelo utilizado na segunda coluna.

Tabela 2.3: Sumário de abordagens para detecção de olhos baseadas em modelo.

Autores	Modelo
Lin e Yang [LY04]	Círculo-Face
Shinjiro e Kawato [KT04]	Padrão entre olhos

2.3 Sumário

Frequentemente, os métodos de detecção de olhos são classificados em duas grandes categorias: métodos ativos e passivos. Os métodos ativos se caracterizam pela utilização de sequências de imagens ou vídeos, pois trabalham com a detecção de alterações em sequências de imagens, do estado dos olhos (aberto/fechado) ou da reflexão de luz infravermelha no fundo da pupila. Os métodos passivos requerem apenas uma única imagem e se dividem em três estratégias de detecção de olhos: baseada na análise da aparência, em aprendizagem e em modelos:

- a abordagem baseada na análise da aparência leva em consideração características como cor, texturas e formas;
- a abordagem baseada em aprendizagem utiliza classificadores treinados com características extraídas de exemplos e contra exemplos de olhos;
- a abordagem baseada em modelos produz uma resposta derivada da avaliação de uma função de casamento entre um modelo de olho e a imagem;

As bases de imagens de faces e os desempenhos encontrados nos artigos revisados são sumarizados na Tabela 2.4. A primeira coluna enumera os autores; enquanto a base de imagens utilizada e o desempenho obtido são apresentados na segunda e terceira coluna, respectivamente.

Tabela 2.4: Base de imagens utilizadas e desempenho dos artigos revisados.

Autores	Base de faces	Desempenho
Huang et al [HSLW98]	Proprietária	97,9%
Peng et al [PZB98]	Proprietária	97,41%
Zhang e Lenders [ZL00]	Yale	Não informado
Samad et al [SHT01]	Olivetti	94,0%
Han et al [HKN02]	Caltech	98,8%
Kumar et al [KRR02]	Não informada	Não informado
Fathi e Manzuri [FM04]	Proprietária	98,0%
Hua e Geng [ZHG04]	NJUFace	98,4%
D’Orazio et al. [DLSG04]	Proprietária	89,0%
Lin e Yang [LY04]	HHI	74,0%
Shinjiro e Kawato [KT04]	Proprietária	Não informado
Wu e Trivedi [WT05]	FERET	92,43%
Wang et al [WGJW05]	FRGC 1.0	94,5%
Wang e Yin [WY05]	JAFFE	95,8%
Xingming e Huangyuan [XH06]	SCUT	96,2%
Jin et al [JYS ⁺ 06]	FERET	90,75%

Após esta revisão bibliográfica é possível concluir que: os métodos ativos se caracterizam pela precisão das respostas sendo indicadas em situações onde é possível utilizar todos os aparatos necessários para sua aplicação; abordagens baseadas na análise de aparência são recomendadas quando existem restrições de recursos computacionais (ex: sistemas embarcados) e as características consideradas apresentam um comportamento homogêneo na população onde o sistema será utilizado; abordagens baseadas em aprendizagem são sugeridas quando é necessária uma maior tolerância na variação das amostras de olhos a serem

detectados e a abordagem baseada em modelos é aconselhada em situações que requerem alta precisão mas onde o tempo de processamento não é um fator crucial (ex: processamento em lotes).

Observando o desempenho sumarizado na Tabela 2.4, é possível afirmar que as abordagens baseadas na análise de aparência apresentam, de uma forma geral, resultados melhores que as outras abordagens. No entanto, esta afirmação não é muito precisa, uma vez que uma comparação correta deve considerar o desempenho dos sistemas sobre uma base de imagens comum.

No Capítulo 3, apresentado a seguir, são definidos os módulos que compõem a abordagem para detecção de olhos em faces humanas proposta nesta dissertação.

Capítulo 3

Abordagem proposta

Neste capítulo, apresenta-se o sistema de detecção de olhos em imagens com faces humanas proposto nesta dissertação. A arquitetura geral do sistema é ilustrada na Figura 3.1.

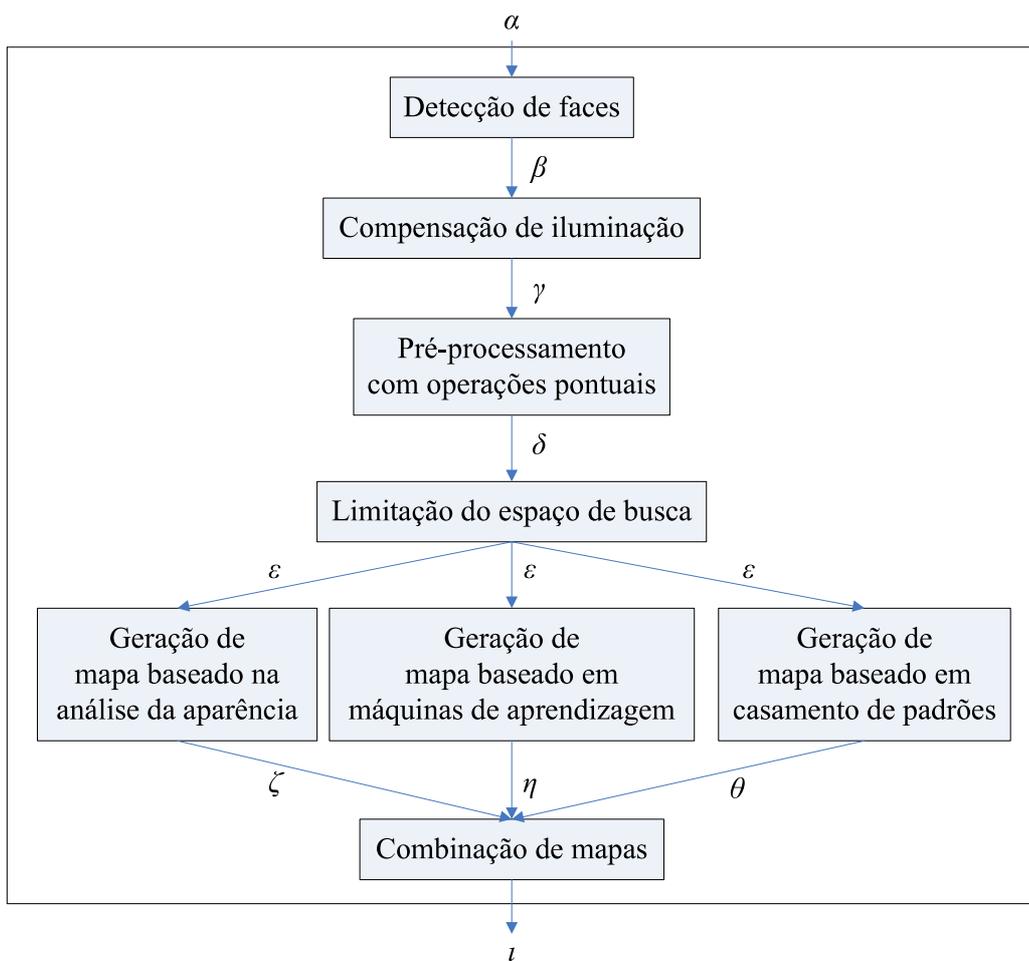


Figura 3.1: Diagrama com a arquitetura geral do sistema proposto.

A operação do sistema pode ser descrita, de maneira sucinta, pelas seguintes etapas:

1. Faces são detectadas em uma imagem de entrada α (os procedimentos seguintes são aplicados em cada imagem de face encontrada em α);
2. A imagem de face β passa por um processo de compensação de iluminação para atenuação de sombras geradas em casos de captura com iluminação lateral, produzindo uma imagem corrigida γ ;
3. A imagem corrigida γ tem seu brilho e contraste melhorados através da aplicação de operações pontuais, gerando uma imagem δ ;
4. Uma redução do espaço de busca é obtida através da determinação de uma subimagem ε que corresponde a uma região de interesse dentro da imagem δ ;
5.
 - (a) Um mapa de evidência ζ é preenchido com valores calculados utilizando a análise da aparência (cor) dos *pixels* da subimagem ε .
 - (b) Um mapa de evidência η é composto a partir da saída de uma máquina de aprendizagem treinada com imagens de olhos e não-olhos quando a subimagem ε é aplicada em sua entrada;
 - (c) Um mapa de evidência θ é produzido através da avaliação de uma função de casamento de modelos ao longo da subimagem ε ;
6. Os mapas de evidência ζ , η e θ são combinados utilizando um conjunto de regras pré-definidas. O resultado desta combinação produz um conjunto de coordenadas ι que corresponde à localização dos olhos.

Um aprofundamento, em termos de detalhes de organização interna, funcionamento e objetivos de cada módulo, é apresentado ao longo das Seções 3.1 até 3.8.

3.1 Detecção de faces

A detecção de faces é uma etapa fundamental na resolução de vários problemas práticos, como recuperação de imagens, automação de composição fotográfica, reconhecimento de faces em sistemas de segurança, dentre outros. O objetivo da detecção de faces é encontrar

rostos em imagens estáticas ou quadros de vídeo. Em casos de objetos complexos como faces humanas, é difícil encontrar características e heurísticas capazes de representar a imensa variedade de instâncias de uma face. Para detecção de objetos com este tipo de dificuldade, geralmente, é utilizado um classificador (modelo estatístico) [BKP05].

O módulo de detecção de faces tem como objetivo, localizar, recortar e normalizar o tamanho de faces humanas, como preparação para a etapa de detecção de olhos. O diagrama deste módulo é apresentado na Figura 3.2 e as descrições de cada componente do mesmo são apresentadas ao longo desta seção.

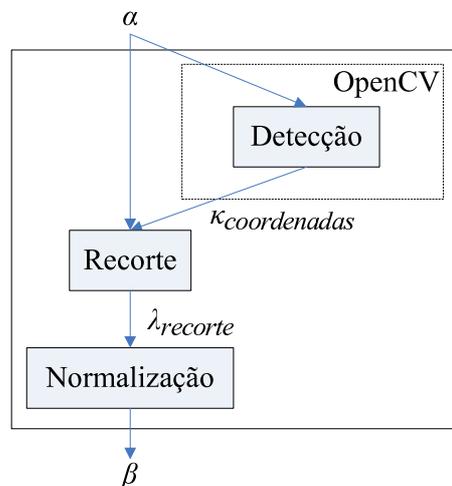


Figura 3.2: Diagrama com a arquitetura do módulo de detecção de faces.

A etapa de detecção é realizada com o uso da biblioteca de software OpenCV (*Intel Open Source Computer Vision Library*) [Ope]. Esta biblioteca é composta de funções e classes, escritas nas linguagens C e C++, que implementam diversos algoritmos de Processamento de Imagens e Visão Computacional. De acordo com Bradski [BKP05], o detector de faces da OpenCV é baseado em uma abordagem originalmente desenvolvida por Viola e Jones [VJ01] e estendida por Lienhart [LM02]. Este método utiliza características simples de *Haar* (assim chamadas porque estas características são calculadas de forma similar aos coeficientes das Transformadas *Wavelets* de *Haar*) e um conjunto de classificadores interligados em cascata como modelo estatístico. O detector de faces foi treinado com imagens de tamanho fixo (24×24) e a detecção é feita deslocando uma janela de busca deste mesmo tamanho sobre a imagem α e verificando se a região correspondente da imagem a cada deslocamento contém uma face. A detecção de faces de diferentes tamanhos é feita através de redimensionamento

da imagem original. Como resultado da etapa de detecção é produzido um conjunto de vetores. Cada vetor $\kappa_{coordenadas}$, contido neste conjunto, contém as coordenadas do canto superior esquerdo, largura e altura das faces que foram localizadas pelo detector na imagem de entrada α .

O procedimento de recorte consiste na cópia de um subconjunto de *pixels* da imagem original α para uma nova imagem $\lambda_{recorte}$. A região de onde os pixels são copiados é definida pelo vetor $\kappa_{coordenadas}$. Esta operação é descrita pela equação,

$$\lambda_{recorte}(i, j) = \alpha(y + i, x + j); \quad i = 1, \dots, h \text{ e } j = 1, \dots, w, \quad (3.1)$$

em que (x, y) indica a posição da coordenada do canto superior esquerdo, h indica a altura e w indica a largura da face detectada (estes parâmetros estão contidos no vetor $\kappa_{coordenadas}$).

O processo de normalização consiste na alteração do tamanho da imagem de face $\lambda_{recorte}$ para uma dimensão pré-definida de 150×150 *pixels*, utilizando interpolação de ordem zero [GW92]. Esta resolução foi definida utilizando o critério da facilidade de determinação do centro da pupila através da simples observação humana. Este processo é fundamental para aplicação das técnicas de detecção de olhos descritas nas Seções 3.5 e 3.6. No esquema de interpolação utilizado, os *pixels* da imagem resultante β são gerados através da replicação do vizinho mais próximo contido na imagem recortada $\lambda_{recorte}$, conforme

$$\beta(i, j) = \lambda_{recorte}(int(i \times r_h))(int(j \times r_w)); \quad i = 1, \dots, h \text{ e } j = 1, \dots, w, \quad (3.2)$$

em que int é um função que retorna a parte inteira de um valor real, r_h e r_w são razões das alturas e larguras entre as imagens β e $\lambda_{recorte}$, respectivamente.

3.2 Compensação de iluminação

Para um observador, uma imagem é formada a partir da luz refletida pelos objetos sendo observados. Num modelo simplificado de formação de imagens [Par97], a imagem percebida é resultado da reflectância de um objeto multiplicada pela iluminação do ambiente, conforme a equação

$$\beta = \mu_{reflec} \times \mu_{ilum}, \quad (3.3)$$

em que β é a imagem formada, μ_{reflec} é a reflectância apresentada pelo contexto da imagem e μ_{ilum} é a iluminação utilizada na captura da imagem. O conceito de formação da imagem é

apresentado de forma prática nas Figuras de 3.3(a) a 3.3(c). A reflectância de um objeto (uma série de faixas claras e escuras neste caso) é apresentada na Figura 3.3(a). Um exemplo de um padrão de iluminação é ilustrado na Figura 3.3(b), enquanto na Figura 3.3(c) visualiza-se uma imagem resultante da multiplicação da reflectância pela iluminação [Smi97].

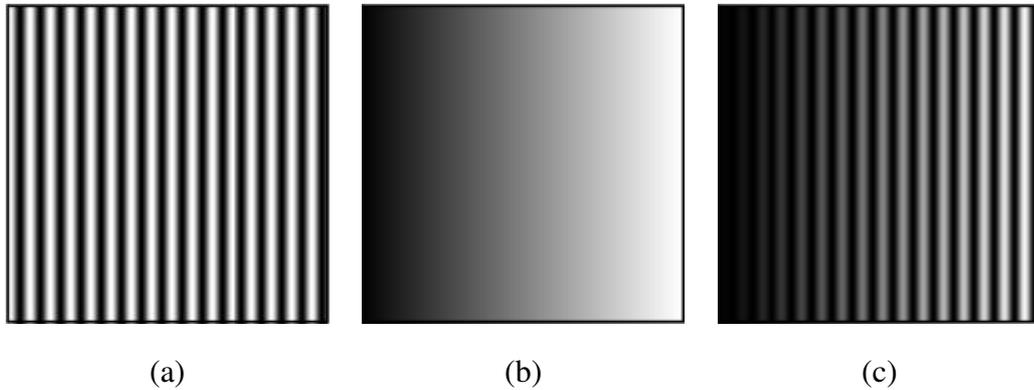


Figura 3.3: Formação da imagem a partir da reflectância e iluminação. Em (a) a reflectância de uma imagem, em (b) o padrão de iluminação e em (c) a imagem visualizada.

Para entender como a formação da imagem, a partir da reflectância e iluminação, está relacionada com a visão no cotidiano, basta imaginar duas pessoas vestidas com roupas escuras e idênticas, uma exposta à luz solar e outra sob a sombra de uma árvore. O percentual de luz que é refletido nas duas pessoas é o mesmo. Porém, como existe uma grande diferença entre a iluminação que atinge o sujeito exposto e o encoberto, a imagem da pessoa sob a sombra se apresenta mais escura e o contraste entre a face (região clara) e as roupas (regiões escuras) também é menor.

O modelo de formação de imagens descrito anteriormente é conhecido por modelo de iluminação-reflexão e tem sido usado para melhoria da qualidade de imagens adquiridas sob condições de iluminação não adequadas para análise automática [DGK06].

O objetivo do módulo de compensação de iluminação é remover o componente de iluminação de uma imagem de face, de forma que a imagem final apresente a reflectância da face independente das condições de iluminação. A filtragem homomórfica foi o procedimento utilizado para separar estes componentes, conforme apresentado no diagrama de blocos presente na Figura 3.4. Este diagrama é detalhado nos parágrafos seguintes.

O processo de filtragem homomórfica, bastante utilizada em Processamento Digital de Sinais e Processamento de Imagens, permite manipular sinais combinados através de ope-

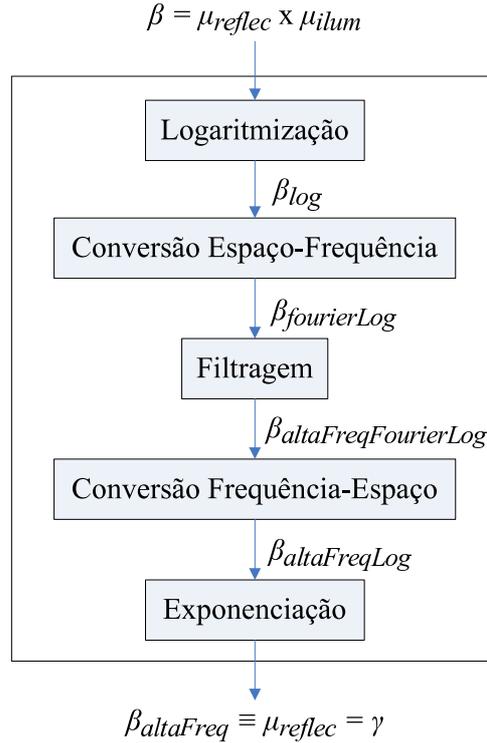


Figura 3.4: Diagrama com a arquitetura do módulo de compensação de iluminação.

rações não lineares. A estratégia consiste em transformar o problema da combinação não linear em um problema de combinação linear, através do uso do princípio da convolução por separabilidade e de operações matemáticas apropriadas [Smi97].

O processo de filtragem homomórfica utilizado foi adaptado da implementação apresentada por Parker [Par97] e começa com a logaritmização dos elementos na Equação 3.3, obtendo

$$\begin{aligned}
 \beta_{log} &= \log(\beta) \\
 &= \log(\mu_{reflec} \times \mu_{ilum}) \\
 &= \log(\mu_{reflec}) + \log(\mu_{ilum}),
 \end{aligned} \tag{3.4}$$

em que β é a imagem resultante da combinação não linear da reflectância (μ_{reflec}) e da iluminância (μ_{ilum}). Desta forma, o problema da separação de sinais combinados através da multiplicação é convertido em um problema de separação de sinais adicionados. Esta conversão torna-se possível através do uso do operador logarítmico.

Em geral, é possível relacionar a reflectância com sinais de alta frequência, enquanto a iluminância apresenta características de baixa frequência. Este relacionamento é ilustrado

nas Figuras 3.3(a) e 3.3(b) [Par97]. Assim, uma equivalência para a reflectância pode ser obtida com uma filtragem do tipo passa-alta (*passa_alta*),

$$\begin{aligned}\mu_{reflec} &\equiv passa_alta(\alpha) \\ &\equiv \beta_{altaFreq}.\end{aligned}\tag{3.5}$$

Esta filtragem foi realizada, no módulo de compensação de iluminação, utilizando uma representação da imagem obtida através da Transformada Rápida de Fourier Bidimensional (*Fast Fourier Transform Bidimensional* - FFT 2D). A Transformada de Fourier é uma operação matemática que converte a representação espacial de uma imagem em uma representação no domínio da frequência [Smi97]. Esta operação de conversão espaço-frequência é representada pela equação

$$\beta_{fourierLog} = fft(\beta_{log}),\tag{3.6}$$

em que $\beta_{fourierLog}$ é a representação da imagem no domínio da frequência e fft é o operador da Transformada Rápida de Fourier. Para permitir a utilização do algoritmo FFT em todas as imagens, estas são preenchidas com bordas brancas para torná-las quadradas e com largura/altura igual a uma potência de dois, quando necessário. A implementação da FFT 2D utilizada foi inspirada na implementação apresentada por Parker [Par97]. Um exemplo de uma mesma imagem representada no domínio do tempo e da frequência, respectivamente, é apresentado nas Figuras 3.5(a) e 3.5(b). A função da filtragem homomórfica é atenuar o componente de iluminação na face e ressaltar a reflectância através da aplicação de um filtro passa-alta, cuja função de transferência (ganho H em função da frequência angular w) é ilustrada na Figura 3.6. O resultado obtido após aplicação da filtragem é expresso por

$$\beta_{altaFreqFourierLog} = passa_alta(\beta_{fourierLog}),\tag{3.7}$$

em que $\beta_{altaFreqFourierLog}$ são as componentes de alta frequência obtidas a partir da aplicação de um filtro passa-alta (*passa_alta*) em $\beta_{fourierLog}$ (representação da imagem no domínio da frequência). Finalmente, é necessária a aplicação da Transformada Inversa Rápida de Fourier (IFFT) para que ocorra a conversão frequência-espço, descrita pela equação

$$\beta_{altaFreqLog} = ifft(\beta_{altaFreqFourierLog}),\tag{3.8}$$

em que $\beta_{altaFreqLog}$ é a representação da reflectância no domínio espacial, obtido pelo módulo de conversão frequência-espço, que implementa a transformada inversa de Fourier (*ifft*).

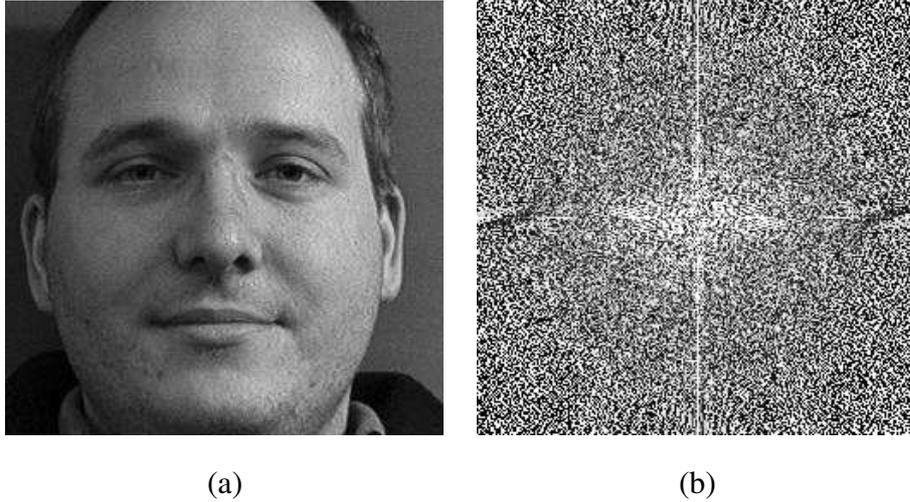


Figura 3.5: Representações no domínio do espaço e da frequência. Em (a) uma imagem representada no domínio do espaço e em (b) a mesma imagem representada no domínio da frequência.

Para inverter a operação de logaritmo aplicada na etapa inicial da filtragem homomórfica e que permitiu a utilização do princípio da convolução por separabilidade, é utilizado o módulo de exponenciação, cuja operação é descrita por:

$$\beta_{altaFreq} = \exp(\beta_{altaFreqLog}). \quad (3.9)$$

Comparando as Equações 3.9 e 3.5, comprovamos que uma equivalência para a refletância foi obtida, como desejado,

$$\beta_{altaFreq} \equiv \mu_{reflec} = \gamma. \quad (3.10)$$

Um exemplo da utilização da compensação de iluminação através da filtragem homomórfica para a imagem da Figura 3.7(a) é apresentado na Figura 3.7(b). É possível observar a remoção de sombras (consequência da captura com iluminação lateral), o que facilita o processo de detecção de olhos.

3.3 Pré-processamento com operações pontuais

De acordo com Bovik [Bov05], uma operação pontual em uma imagem γ é uma função f de uma única variável aplicada identicamente em todos os *pixels* na imagem, gerando uma

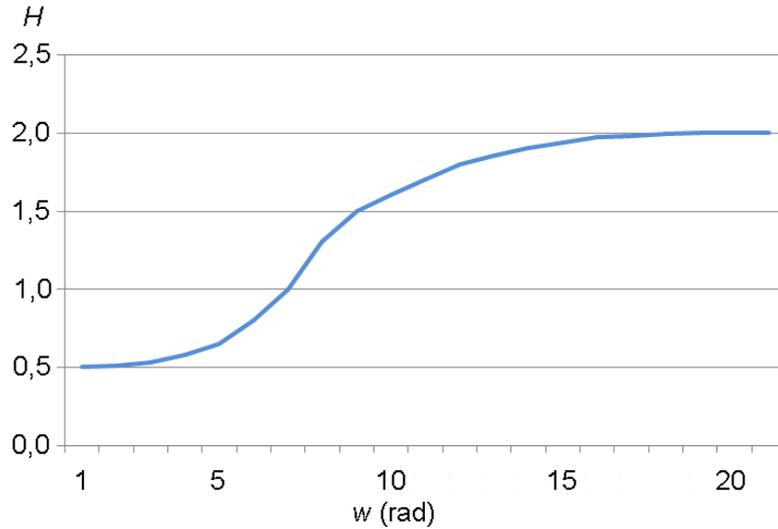


Figura 3.6: Função de transferência do filtro passa-alta utilizado na filtragem homomórfica.

imagem modificada δ . Portanto para cada coordenada (x, y) é obtida

$$\delta(x, y) = f(\gamma(x, y)). \quad (3.11)$$

Operações pontuais não afetam o posicionamento espacial de objetos em uma imagem, nem suas formas, mas apenas incrementam ou decrementam (ou não alteram) o valor de cada *pixel* de acordo com a função f utilizada. Portanto, uma operação pontual f altera apenas o histograma e a aparência geral de uma imagem. O módulo de pré-processamento com operações pontuais, cujo diagrama de blocos é apresentado na Figura 3.8, utiliza as operações pontuais de expansão linear de histograma e equalização de histograma apresentadas por Bovik [Bov05]. Estas operações são detalhadas nas Seções 3.3.1 e 3.3.2.

3.3.1 Expansão de histograma

Uma imagem com uma ampla distribuição em seu histograma tende a apresentar uma melhor distinção visual. Uma expansão linear de histograma, também conhecida como expansão de contraste, é uma operação pontual linear que expande o histograma de uma imagem de forma a utilizar toda a faixa de valores disponíveis [Bov05].

O diagrama de blocos do módulo de expansão de histograma contido na etapa de pré-processamento do detector de olhos é apresentado na Figura 3.9. A explicação do funcionamento de cada etapa se encontra nos parágrafos seguintes.

O primeiro passo para expansão do histograma de uma imagem γ é determinar os extremos do histograma. Os valores destes extremos são definidos nas etapas de cálculo do mínimo e cálculo do máximo, expressa por:

$$a = \text{minimo}(\gamma) \quad (3.12)$$

$$b = \text{maximo}(\gamma), \quad (3.13)$$

em que a e b correspondem ao menor e maior valor de intensidade encontrado na imagem γ .

O objetivo da expansão linear de histograma é mapear os valores entre a e b na imagem para valores situados entre 0 e $k - 1$, em que $k - 1$ é o valor máximo de intensidade que um *pixel* pode assumir. Esta operação pode ser expressa em duas equações lineares

$$pa + l = 0 \quad (3.14)$$

$$pb + l = k - 1 \quad (3.15)$$

em que as duas incógnitas (p e l) apresentam as seguintes soluções

$$p = \left(\frac{k - 1}{b - a} \right) \quad (3.16)$$

$$l = -a \left(\frac{k - 1}{b - a} \right), \quad (3.17)$$

portanto, a etapa de expansão é dada por:

$$\delta(i, j) = \text{expansao}(\gamma(i, j)) = \left(\frac{k - 1}{b - a} \right) (\gamma(i, j) - a). \quad (3.18)$$

Através da expansão linear de histograma é possível obter um aprimoramento do contraste e da visibilidade de detalhes em uma imagem. A expansão linear de histograma pode ser aplicada para melhoria de imagens capturadas com super-exposição (imagens muito claras) ou com falta de iluminação adequada. Um exemplo de melhoria da qualidade de uma imagem capturada com excesso de iluminação, através da expansão linear de histograma, é apresentado nas Figuras 3.10(a) e 3.10(b). Observando as imagens, pode se perceber que a região escura da pupila é ressaltada, o que facilita o trabalho de detecção dos olhos.

É importante salientar que a determinação dos valores de a e b utilizados no módulo de pré-processamento com operações pontuais é feita levando em consideração apenas a região de espaço de busca definido na Seção 3.4, pois o interesse é ressaltar o contraste entre a região da pele e dos olhos.

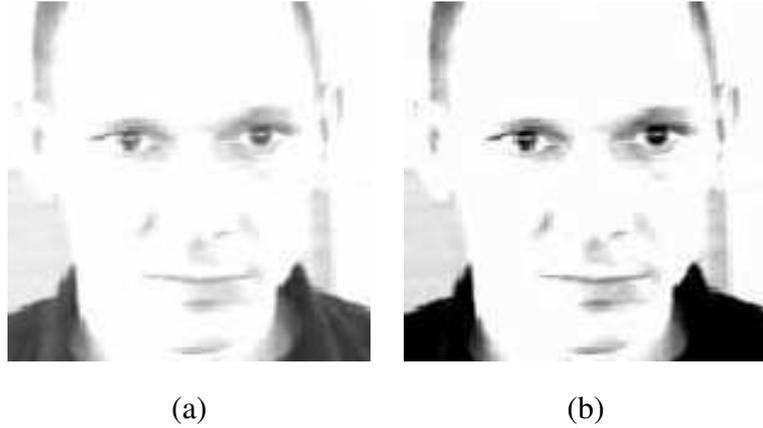


Figura 3.10: Exemplo de expansão linear de histograma em imagem com excesso de iluminação. Em (a) uma imagem capturada com excesso de iluminação e (b) a mesma imagem após o processo de expansão linear de histograma.

Outro exemplo de aprimoramento com expansão linear de histograma, agora de uma imagem capturada com pouca iluminação, é ilustrado nas Figuras 3.11(a) e 3.11(b). A Figura 3.11(a) é a imagem original da face e a Figura 3.11(b) é o resultado da aplicação da expansão linear de histograma. O efeito de “clareamento” da imagem é resultante do espalhamento da

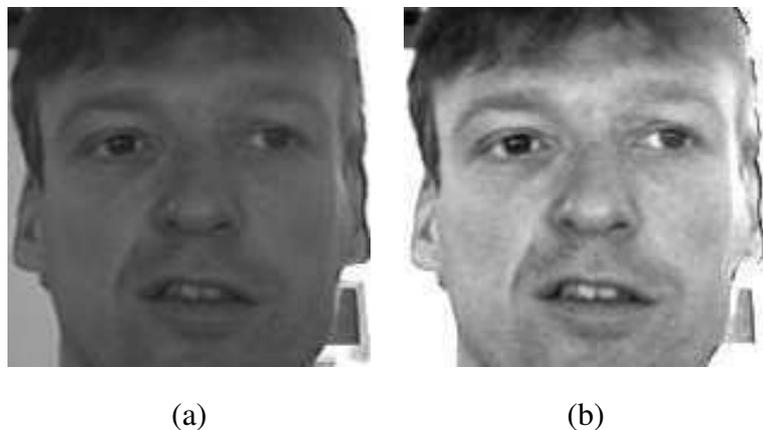


Figura 3.11: Exemplo de expansão linear de histograma em imagem com pouca iluminação. Em (a) uma imagem capturada com pouca iluminação e (b) a mesma imagem após o processo de expansão linear de histograma.

distribuição de valores das regiões mais escuras (mais à esquerda) para regiões mais claras, conforme ilustrado nas Figuras 3.12(a) e 3.12(b), que correspondem aos histogramas das Figuras 3.11(a) e 3.11(b), respectivamente.

É importante observar que em imagens que já possuem grande utilização da faixa de

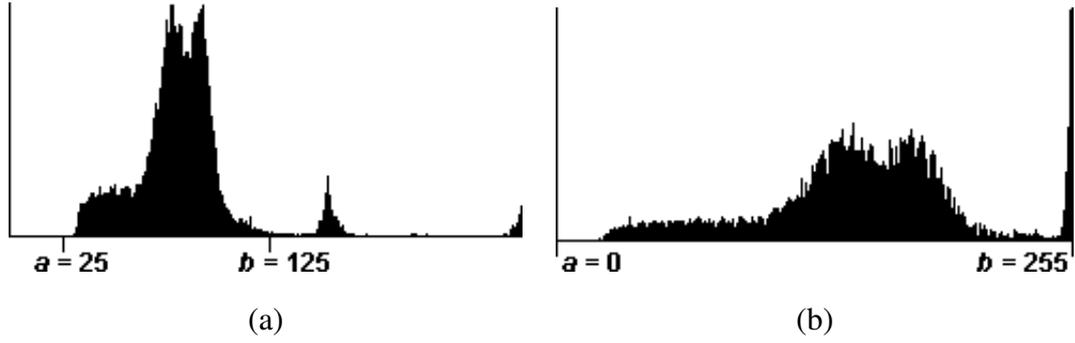


Figura 3.12: Expansão linear de histograma em imagem com pouca iluminação. Em (a) o histograma de uma imagem capturada com excesso de iluminação e (b) o histograma após o processo de expansão.

valores de intensidade disponível, a expansão linear de histograma provoca pouca ou mesmo nenhuma alteração.

3.3.2 Equalização de histograma

A equalização de histograma é uma operação não linear que estende a idéia da expansão linear de histograma descrita na Seção 3.3.1. Além da distribuição do histograma preencher toda a faixa disponível, esta distribuição deve ser uniforme por toda a faixa de valores. Considerando o conceito de máxima entropia, uma imagem com histograma perfeitamente uniforme contém a maior quantidade de informação ou complexidade [Bov05]. O diagrama de blocos do módulo de equalização de histograma utilizado na etapa de pré-processamento do detector de olhos é apresentado na Figura 3.13.

Para uma imagem γ contendo n pixels com um histograma ν , o cálculo do histograma normalizado é dado por

$$\xi_{norm}(i) = \frac{1}{n}\nu(i); i = 0, \dots, k - 1, \quad (3.19)$$

lembrando que esta função tem a seguinte propriedade

$$\sum_{i=0}^{k-1} \xi_{norm}(i) = 1. \quad (3.20)$$

O histograma normalizado ξ corresponde à densidade de probabilidade de valores de intensidade de uma imagem. Em outras palavras, se a coordenada de um pixel é escolhida aleatoriamente, então $\xi(q) = Pr\{\gamma(x, y) = q\}$.

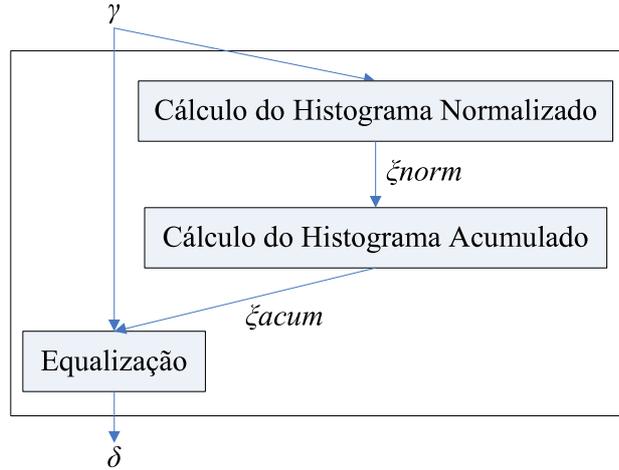


Figura 3.13: Diagrama do módulo de equalização de histograma.

O cálculo do histograma acumulado é definido em termos do histograma normalizado como:

$$\xi_{acum}(r) = \sum_{i=0}^r \xi_{norm}(i); r = 0, \dots, k - 1, \quad (3.21)$$

em que a função ξ_{acum} é uma função de distribuição de probabilidade acumulada, portanto trata-se de uma função não decrescente com $\xi_{acum}(k - 1) = 1$. Esta função permite a interpretação probabilística de que, para uma coordenada (x, y) aleatoriamente selecionada, $\xi_{acum}(r) = Pr\{\gamma(x, y) \leq r\}$. A partir da Equação 3.21, o histograma normalizado é expressado como:

$$\xi_{norm}(q) = \xi_{acum}(q) - \xi_{acum}(q - 1); q = 0, \dots, k - 1. \quad (3.22)$$

Ambas ξ_{norm} e ξ_{acum} , são descrições completas da distribuição dos valores de intensidade dos *pixels* de uma imagem γ .

Tratando ξ_{norm} e ξ_{acum} como funções de uma variável contínua x , estes vetores podem ser vistos como uma função de densidade de probabilidade (fdp) e uma função de distribuição acumulada (FDA) (valendo o relacionamento $\xi_{norm} = d\xi_{acum}(x)/dx$), portanto, a imagem produzida pela operação de equalização pode ser descrita por

$$\delta(i, j) = equalizacao(\gamma(i, j)) = \xi_{acum}(\gamma(i, j)), \quad (3.23)$$

em que imagem δ tem um histograma com distribuição plana.

Um exemplo de utilização da equalização de histogramas em uma imagem de face é apresentado nas Figuras 3.14(a) e 3.14(b). A imagem de face original se encontra na Figura 3.14(a) e a versão após a equalização de histogramas é apresentada na Figura 3.14(b).

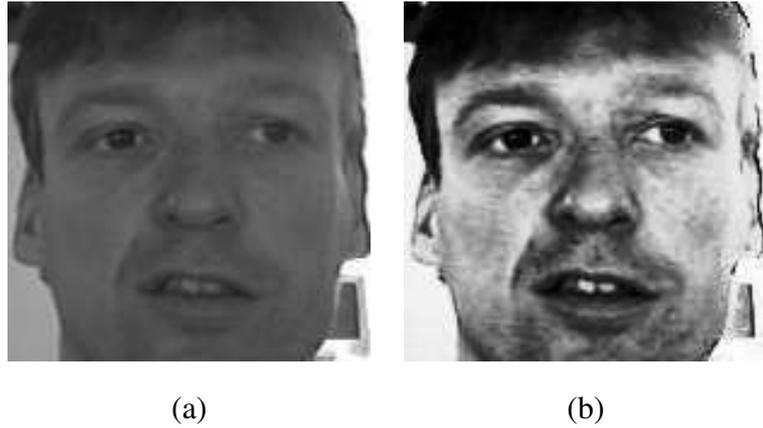


Figura 3.14: Exemplo de equalização de histograma. Em (a) uma imagem sem equalização. Em (b) a mesma imagem após a equalização de histograma.

Um histograma não pode ser completamente equalizado com a utilização deste procedimento quando os *pixels* possuem valores discretos, isto ocorre porque os valores contidos no histograma acumulado ξ_{acum} precisam ser quantizados para valores disponíveis na representação considerada [Bov05]. Esta limitação pode ser observada através dos histogramas, após a aplicação do processo de equalização, que se encontram nas imagens das Figuras 3.15(a) e 3.15(b), respectivamente. Após o processo de equalização, o histograma apresenta um aspecto de distribuição que se assemelha, porém não é igual, a uma distribuição uniforme.

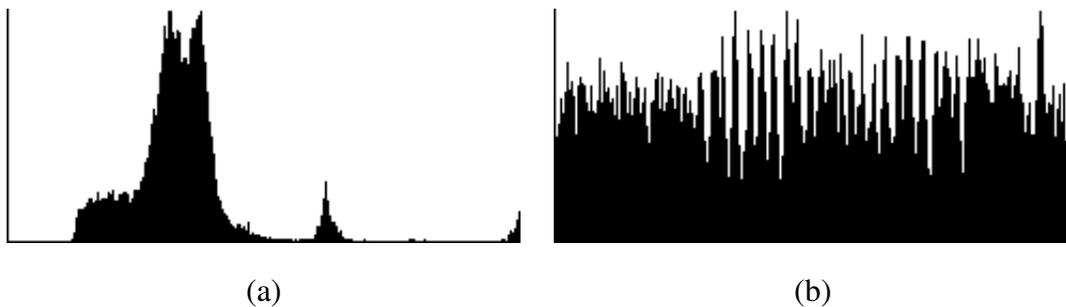


Figura 3.15: Equalização de histograma. Em (a) o histograma de uma imagem sem equalização e em (b) o histograma após o processo de equalização.

3.4 Limitação do espaço de busca

Para diminuir o tempo de execução e aumentar a taxa de acerto do algoritmo de detecção de olhos proposto nesta dissertação, é necessário definir uma área de espaço de busca em

uma imagem de face. A função do módulo de limitação do espaço de busca é restringir a busca por olhos apenas nesta sub-região da imagem. A operação de limitação presente neste módulo expressa por:

$$\begin{aligned}\varepsilon(i, j) &= \text{limitacao}(\delta) = \\ &= \delta(x + j, y + i); i \in \varphi_{ver} \text{ e } j \in \varphi_{hor},\end{aligned}\tag{3.24}$$

em que ε é o espaço de busca reduzido para a imagem δ , utilizando os limites de busca φ_{ver} e φ_{hor} , cuja determinação é apresentada nos parágrafos seguintes. O diagrama do bloco deste módulo se encontra na Figura 3.16.

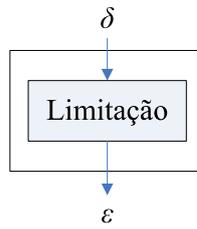


Figura 3.16: Diagrama do módulo de limitação do espaço de busca.

A determinação do espaço de busca para olhos foi feito através do cálculo da localização média dos olhos em imagens de faces detectadas pelo detector do OpenCV [Ope]. O processo para determinação destas posições médias consistiu na aplicação do detector de faces do OpenCV na base de faces BioID [Sca] e na leitura das coordenadas dos olhos em arquivos que acompanham esta base. Todos os valores destas coordenadas foram normalizados através da translação da origem do sistema de coordenadas do canto superior esquerdo da imagem para o canto superior esquerdo da face detectada e posterior divisão pela largura ou altura da face detectada, de acordo com o eixo da coordenada. A transformação aplicada nestas coordenadas é explicitada por

$$x_{norm} = \frac{(x_{olho} - x_{face})}{w}\tag{3.25}$$

$$y_{norm} = \frac{(y_{olho} - y_{face})}{h},\tag{3.26}$$

em que (x_{norm}, y_{norm}) é o par de coordenadas normalizadas, (x_{olho}, y_{olho}) é o par de coordenadas do olho obtida de arquivo da base de dados BioID, (x_{face}, y_{face}) é a coordenada do canto superior esquerdo da face e w e h correspondem à largura e altura da face detectada, respectivamente. É importante observar que os valores da coordenada normalizada se encontram no intervalo $[0; 1]$.

Para certificação da validade das coordenadas dos olhos disponíveis nas 1521 imagens contidas na base BioID [Sca] e do processo de normalização das coordenadas descrito previamente, recortes das imagens com a face detectada e os olhos marcados (utilizando as coordenadas normalizadas) foram gerados. Através da inspeção visual destes recortes, foram observadas ocorrências de:

- falsa detecção de face em 100 imagens. Exemplos de falsa detecção são apresentados nas Figuras de 3.17(a) a 3.17(d);
- dupla detecção de face em 4 imagens. A Figura 3.18(a) com a Figura 3.18(b) e a Figura 3.18(c) com a Figura 3.18(d) ilustram exemplos de dupla detecção de uma mesma face em uma imagem;
- ausência de detecção de faces em 47 imagens.

As coordenadas normalizadas geradas a partir da segunda detecção para os casos de dupla detecção (Figuras 3.18(b) e 3.18(d)) foram descartadas. A validade das coordenadas fornecidas com a base BioID foi comprovada ao final da inspeção visual.

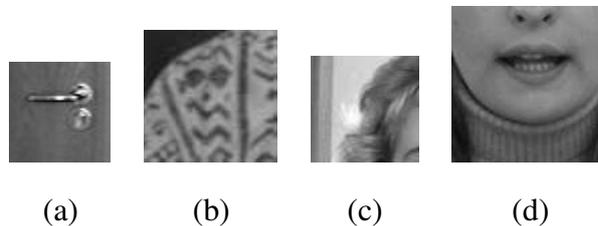


Figura 3.17: Exemplos de falsa detecção de faces.

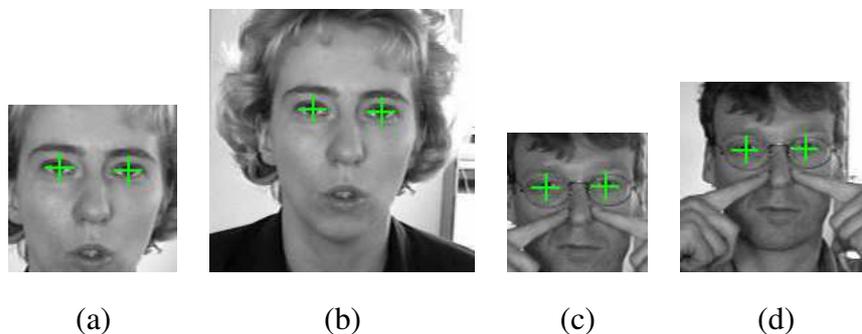


Figura 3.18: Exemplos de dupla detecção de mesma face. A faces detectadas em (a) e (c) estão contidas nas detecções em (b) e (d), respectivamente.

As estatísticas obtidas das coordenadas normalizadas geradas no procedimento anterior são apresentadas na Tabela 3.1. A primeira coluna indica o olho (esquerdo ou direito); a segunda coluna, o eixo da coordenada; a terceira coluna, a média da coordenada; a quarta coluna, o desvio padrão da coordenada média e a quinta coluna, a largura média para cada olho. As posições médias dos olhos foram definidas utilizando imagens com os olhos ali-

Olho	Eixo	Média (μ)	Desvio Padrão (σ)	Largura
Direito	X	0,30	0,018	0,13
	Y	0,38	0,020	
Esquerdo	X	0,68	0,018	0,13
	Y	0,38	0,021	

Tabela 3.1: Estatísticas das coordenadas normalizadas

nhados com o eixo horizontal. Para permitir a detecção de olhos em faces com inclinação de até 45° no plano da face é necessário incorporar esta variação de inclinação conforme as transformações descritas a seguir. As coordenadas médias obtidas para os olhos direito e esquerdo foram

$$\overline{olho_{dir}} = (0, 30; 0, 38) \quad (3.27)$$

$$\overline{olho_{esq}} = (0, 68; 0, 38). \quad (3.28)$$

O resultado da aplicação da translação da origem do sistema de coordenadas para a posição central da imagem (0,50; 0,50) resulta em:

$$olho_{dirTrans} = (-0, 20; -0, 12) \quad (3.29)$$

$$olho_{esqTrans} = (0, 18; -0, 12). \quad (3.30)$$

A aplicação de operações de rotação de $+45^\circ$ e -45° (no plano da face) nas posições transladadas dos olhos resulta nas coordenadas que são definidas pelas equações

$$olho_{dirRot+45} = (-0, 22; 0, 06) \quad (3.31)$$

$$olho_{esqRot+45} = (0, 04; -0, 20) \quad (3.32)$$

$$olho_{dirRot-45} = (-0, 06; -0, 22) \quad (3.33)$$

$$olho_{esqRot-45} = (0, 20; 0, 04). \quad (3.34)$$

Finalmente, como resultado da translação do sistema das coordenadas rotacionadas para a posição do sistema de coordenadas inicial são obtidos os seguintes valores:

$$olho_{dirTransRot+45} = (0, 28; 0, 56) \quad (3.35)$$

$$olho_{esqTransRot+45} = (0, 54; 0, 30) \quad (3.36)$$

$$olho_{dirTransRot-45} = (0, 44; 0, 28) \quad (3.37)$$

$$olho_{esqTransRot-45} = (0, 70; 0, 54). \quad (3.38)$$

Tomando como base os valores máximos e mínimos definidos pelas coordenadas descritas nas Equações 3.35, 3.36, 3.37 e 3.38 para o eixo horizontal e considerando um deslocamento horizontal igual à largura média dos olhos apresentada na Tabela 3.1, define-se o intervalo de busca horizontal φ_{hor} conforme

$$\begin{aligned} \varphi_{hor} &= (0, 28 - 0, 13; 0, 70 + 0, 13 + 2) \\ &= (0, 15; 0, 85), \end{aligned} \quad (3.39)$$

em que o valor 2, na parcela referente ao eixo vertical, foi adicionado visando uma simetria do espaço de busca em relação ao centro da imagem de face. O deslocamento horizontal igual à largura dos olhos é suficiente, visto que é aproximadamente sete vezes maior que o desvio padrão (σ) da coordenada dos olhos, considerando o eixo X, apresentado na Tabela 3.1.

Novamente, utilizando os valores máximos e mínimos definidos pelas coordenadas descritas nas Equações 3.35, 3.36, 3.37 e 3.38 para o eixo horizontal, define-se o intervalo de busca vertical apresentado na Equação 3.40,

$$\varphi_{ver} = (0, 28; 0, 56). \quad (3.40)$$

O diagrama na Figura 3.19 ilustra o procedimento descrito nos parágrafos anteriores. O retângulo vermelho representa a área delimitada pelas operações de rotações das coordenadas do centro dos olhos subtraída/adicionada com a distância média dos olhos na direção horizontal. Exemplos de espaço de busca por olhos em faces, gerados com a utilização dos limites definidos pelas Equações 3.39 e 3.40, são apresentados nas Figuras de 3.20(a) a 3.20(d).

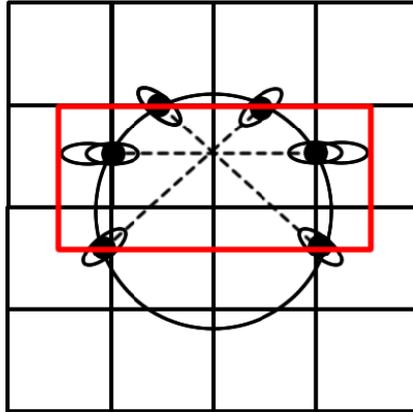


Figura 3.19: Diagrama ilustrativo da definição do espaço de busca.

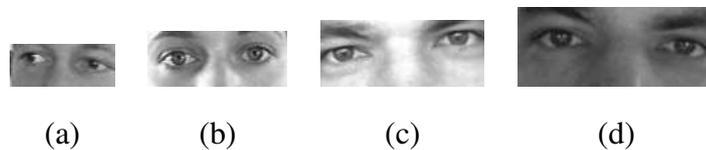


Figura 3.20: Exemplos de espaço de busca de olhos em imagens de faces.

Deste ponto em diante, por simplicidade, o termo detecção de olhos em faces estará relacionado com a detecção de olhos não na face detectada como um todo e sim no espaço de busca, definido nesta seção, contido na face.

3.5 Geração de mapa de evidência baseado na análise da aparência

Máquinas de aprendizagem são modelos computacionais que aplicam o princípio da aproximação de função através de exemplos, significando que elas podem “aprender” uma função através da observação de alguns exemplos da mesma. Para que uma máquina de aprendizagem seja capaz de solucionar um problema, é necessário que este problema seja passível de representação em forma de uma função para a qual um conjunto de entradas e saídas é conhecido. Este conjunto de entradas e saídas deve representar plenamente o comportamento da função [Nis05].

A análise de características dos *pixels* é feita no módulo de geração de mapa de evidência baseado na análise da aparência, utilizando uma máquina de aprendizagem do tipo rede neural, treinada com exemplos de pele, não-pele e área de transição entre estas duas classes.

O diagrama de blocos deste módulo é ilustrado na Figura 3.21 e sua descrição detalhada é apresentada nos parágrafos seguintes.

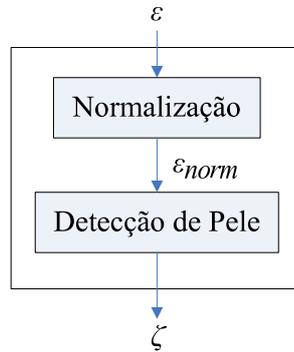


Figura 3.21: Diagrama do módulo de geração de mapa de evidência baseado na análise da aparência.

Redes neurais são treinadas utilizando exemplos. Um exemplo é um par de vetores que contém uma entrada definida e uma saída desejada para a função a ser aproximada. No processo de treinamento, os exemplos de entrada são apresentados e os pesos associados às conexões vão sendo ajustados de forma a produzir na saída da rede o resultado desejado para cada exemplo [Hay98]. No treinamento da rede utilizada na detecção de pele, foram utilizadas 122 imagens contidas na base rotulada e disponibilizada por Jones e Rehg [JR02]. Cada imagem do conjunto de treinamento foi segmentada manualmente. A cada um destes segmentos foi atribuída a uma de três classes distintas: pele, região de transição e não-pele. O efeito do rotulamento sobre as imagens das Figuras 3.22(a) e 3.22(b) é ilustrado nas Figuras 3.23(a) e 3.23(b), respectivamente.

Tanto na formação dos vetores utilizados no processo de treinamento, quanto na propagação dos valores no processo de utilização da rede neural, é necessário que as imagens utilizadas sejam normalizadas, de modo que a intensidade de cada canal RGB varie no intervalo $[0, 1]$. Esta normalização é expressa como:

$$\varepsilon_{norm}(i, j, R) = \frac{\varepsilon(i, j, R)}{max} \quad (3.41)$$

$$\varepsilon_{norm}(i, j, G) = \frac{\varepsilon(i, j, G)}{max} \quad (3.42)$$

$$\varepsilon_{norm}(i, j, B) = \frac{\varepsilon(i, j, B)}{max}. \quad (3.43)$$

onde max é o valor da intensidade máxima de um *pixel* (255 na representação de imagem



(a)

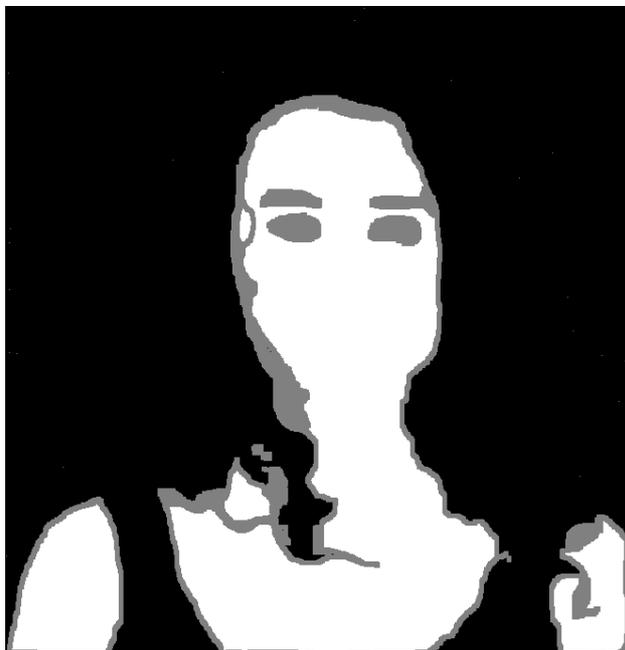


(b)

Figura 3.22: Imagens originais sem rotulamento disponibilizadas por Jones e Rehg [JR02].



(a)



(b)

Figura 3.23: Imagens com regiões rotuladas por Cardoso e Gomes [CG07]. Em (a) rotulamento da imagem 3.22(a) e em (b) rotulamento da imagem 3.22(b).

adotada neste trabalho).

O detector de pele pode ser visualizado como uma função capaz de mapear a cor de um *pixel* em um valor que indica uma evidência do *pixel* pertencer a uma região de pele. O módulo de detecção de pele foi adaptado do trabalho de Cardoso e Gomes [CG07] e utiliza uma rede neural do tipo Perceptron de Multicamadas (*Multilayer Layer Perceptron - MLP*) implementada com a biblioteca de *software Fast Artificial Neural Network Library - FANN* [Nis05] e treinada com o algoritmo de *backpropagation*. A saída o de cada neurônio da rede é obtida através da aplicação de uma função de ativação (f) sobre a soma das entradas (i_0, \dots, i_k) ponderadas pelos pesos (w_0, \dots, w_k) associados a cada conexão. Este procedimento é ilustrado na Figura 3.24.

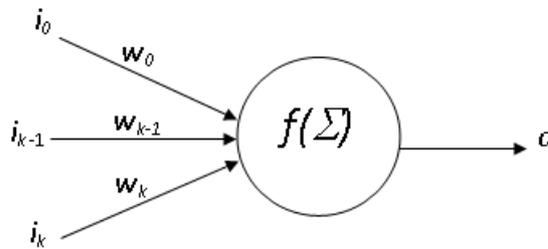


Figura 3.24: Arquitetura básica de um neurônio artificial.

A utilização de uma máquina de aprendizagem do tipo rede neural MLP se deve à sua adequação na solução de problemas linearmente inseparáveis [Bis95], visto que, as experiências realizadas por Cunha [CG05] e Chiang [CTY⁺03] apresentaram fortes evidências de que a definição de um modelo de cor de pele em um espaço de cores é um problema desta classe.

A rede neural utilizada na detecção de pele apresenta a arquitetura de três camadas com três neurônios na camada de entrada, nove neurônios na camada escondida e um neurônio na camada de saída ($3 \times 9 \times 1$), conforme o diagrama ilustrado na Figura 3.25. A três entradas correspondem aos valores dos canais vermelho, verde e azul (R, G, B) de cada *pixel* da imagem e a saída o indica a evidência do *pixel* pertencer a região de pele.

A rede foi treinada até que o erro médio quadrático (*Mean Square Error - MSE*) atingisse o valor 10^{-5} , resultando em uma taxa de acerto de 85,02% no processo de validação do tipo *hold-out* com 3/4 das amostras selecionadas para o treinamento e 1/4 para a validação.

O mapa de evidência ζ gerado pelo módulo de geração de mapa de evidência baseado na

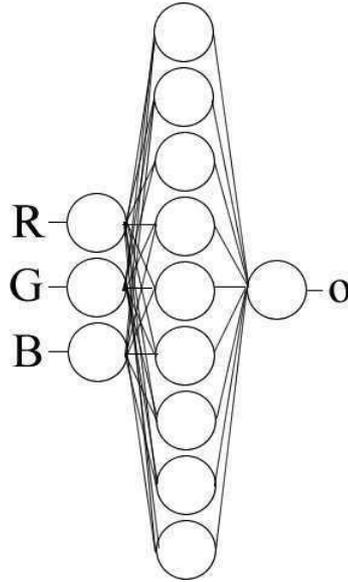


Figura 3.25: Organização da rede neural utilizada para detecção de pele.

análise da aparência é uma imagem que corresponde à resposta apresentada pela rede neural quando aplicada *pixel-a-pixel* na imagem ε_{norm} . Esta operação é representada pela seguinte equação:

$$\zeta(i, j) = 1 - \text{deteccao}(\varepsilon_{norm}(i, j, R), \varepsilon_{norm}(i, j, G), \varepsilon_{norm}(i, j, B)), \quad (3.44)$$

em que *deteccao* representa a função de saída da rede neural. É importante ressaltar que o mapa de evidência ζ representa o complemento da detecção de pele.

Exemplos de faces e seus respectivos mapas gerados através do módulo de geração de mapa de evidência baseado na análise da aparência são apresentados a seguir. Na Figura 3.26(a), ilustra-se um mapa com o comportamento desejado, ressaltando apenas as regiões dos olhos. Na Figura 3.26(b), além dos olhos, é possível perceber uma forte resposta na região das sobrancelhas e na Figura 3.26(c) esta resposta se apresenta também no cenário de fundo. O comportamento apresentado nestes dois últimos exemplos é esperado, visto que o objetivo do mapa é ressaltar regiões limitadas por pele.

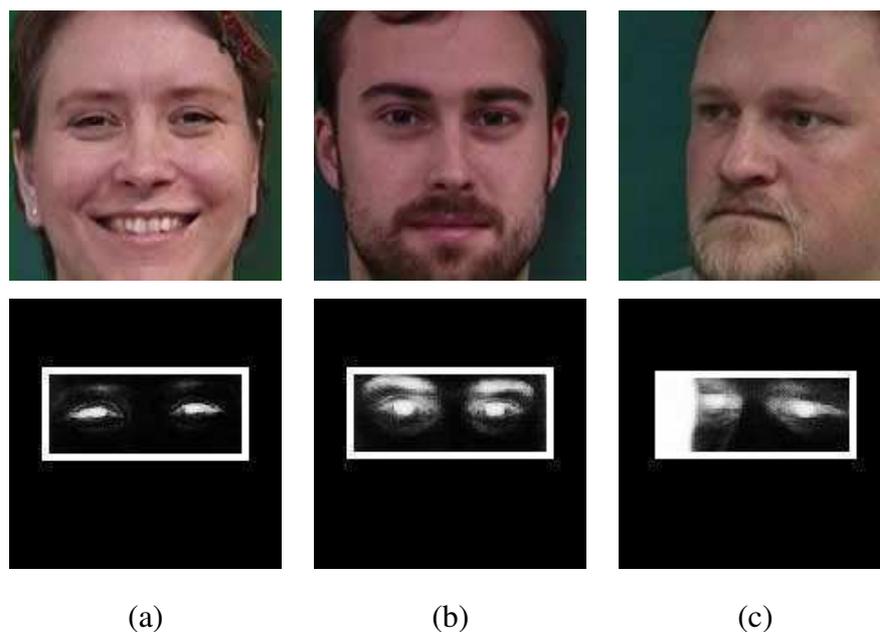


Figura 3.26: Mapas de evidência gerados com análise da aparência. Áreas de busca delimitadas por retângulos brancos.

3.6 Geração de mapa de evidência baseado em aprendizagem

Na resolução de problemas relacionados ao reconhecimento de padrões em imagens com a utilização de técnicas de aprendizagem, é comum representar entrada e saída com vetores numéricos. Os vetores de entrada são formados pelos próprios *pixels* que compõem a imagem e os vetores de saída são formados pelos valores desejados. Porém, para maximizar o desempenho de qualquer classificador utilizado, é importante que o conjunto de valores escolhidos como entrada da máquina seja representativo para o problema a ser solucionado. Seria impossível, para qualquer abordagem baseada em treinamento, diferenciar duas simples classes se esta abordagem fosse treinada com exemplos de apenas uma classe ou com características que não fossem discriminativas entre estas duas classes. Características desejáveis para utilização no processo de treinamento são aquelas que apresentam uma baixa variação entre padrões de uma mesma classe e alta variação entre padrões de classes diferentes.

O módulo de geração de mapa de evidência baseado em aprendizagem utiliza uma rede neural MLP implementada com a biblioteca de *software Stuttgart Neural Network Simulator - SNNS [ZMV⁺]* e treinada com o algoritmo de *backpropagation*. A arquitetura e o

algoritmo de treinamento são similares aos utilizados no módulo de geração de mapa de evidência baseado na análise da aparência descrito na Seção 3.5. A diferença entre estas duas redes, além do conjunto utilizado no treinamento, fica por conta da quantidade de neurônios em cada camada e a quantidade de características utilizadas como entrada. A quantidade de características utilizadas no módulo de geração de mapa de evidência baseado em aprendizagem é centenas de vezes maior do que a requerida pela rede neural do módulo de geração de mapa de evidência baseado na análise da aparência. Desta forma, para reduzir o tamanho do vetor de entrada e selecionar características discriminativas para o problema de classificação em questão, utilizamos Análise de Componentes Principais (*Principal Component Analysis* - PCA) para seleção de características. O diagrama de blocos do módulo de geração de mapa de evidência baseado em aprendizagem é ilustrado na Figura 3.27 e sua descrição detalhada é apresentada nos parágrafos seguintes.

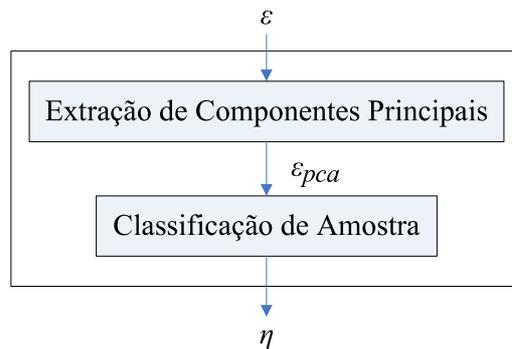


Figura 3.27: Diagrama do módulo de geração de mapa de evidência baseado em aprendizagem.

Para determinação dos valores da matriz dos componentes principais e treinamento da rede neural, foi necessário montar um base de imagens contendo exemplos de olhos e não-olhos. O procedimento para construção desta base seguiu três etapas.

Primeiramente, utilizando a base de imagens BioID [Sca] e uma base formada por imagens capturadas na *World Wide Web*, faces foram detectadas, extraídas e normalizadas para o tamanho 150×150 pixels [LPG⁺07]. Esta etapa foi realizada com o uso do módulo detector de faces descrito na Seção 3.1.

Em seguida, os candidatos a olhos e não-olhos foram extraídos do espaço de busca definido na Seção 3.4 de cada imagem de face normalizada. A redução do número de candidatos que passaram para a próxima etapa foi feita com o auxílio de um procedimento de filtra-

gem de candidatos inspirado no trabalho de Viola [VJ04]. Este procedimento utiliza uma representação intermediária da imagem, chamada imagem integral, definida como

$$\varepsilon_f(i, j) = \varepsilon(i, j) \quad (3.45)$$

$$+ \varepsilon_f(i, j - 1) \quad (3.46)$$

$$+ \varepsilon_f(i - 1, j) \quad (3.47)$$

$$- \varepsilon_f(i - 1, j - 1), \quad (3.48)$$

em que $\varepsilon_f(i, j - 1)$, $\varepsilon_f(i - 1, j)$ e $\varepsilon_f(i - 1, j - 1)$ são chamadas recursivas a própria função imagem integral. A vantagem da utilização da função da imagem integral é que ela permite o rápido cálculo do valor médio dos pixels de uma região começando em x, y e com largura w e altura h , como apresentado na equação

$$\begin{aligned} media(x, y, w, h) = & \varepsilon_f((y + h, x + w)) \\ & - \varepsilon_f(y, x + w) \\ & - \varepsilon_f(y + h, x) \\ & + \varepsilon_f(y, x). \end{aligned} \quad (3.49)$$

Para a filtragem de candidatos, uma grade do mesmo tamanho das imagens de olhos utilizadas no treinamento (36×21 pixels) e organizada como uma matriz de nove células com tamanho 12×7 cada, como ilustrada na Figura 3.28 [LPG⁺07].

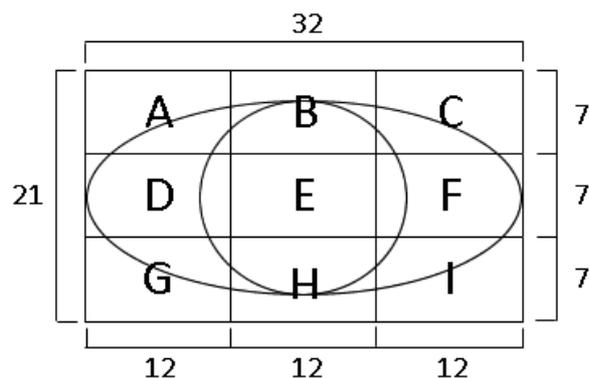


Figura 3.28: Organização da grade de filtragem de olhos.

A grade para filtragem de candidatos foi deslocada por toda a imagem integral pré-calculada e o valor médio de cada célula foi calculado a cada movimentação. Foram se-

lecionados os candidatos cujas células satisfizeram a seguinte expressão lógica:

$$\begin{aligned}
& (media(ABC) > media(DEF)) \\
& \wedge (media(GHI) > media(DEF)) \\
& \wedge ((media(E) < media(D) \vee media(E) < media(F)) \\
& \wedge ((media(E) < media(G) \vee media(E) < media(I)) \\
& \wedge (media(E) < media(H)), \tag{3.50}
\end{aligned}$$

em que A, \dots, I representam cada uma das células da grade e $media(x)$ é o valor médio dos pixels contidos na célula (ou grupo de células) X , calculado utilizando a imagem integral. Analisando as condições presentes na expressão lógica da Equação 3.50, pode-se concluir que a aplicação destas permite reduzir os exemplos de olhos a imagens com uma região predominante mais escura na sua região central (aparência do olho).

Finalmente, os candidatos selecionados são rotulados manualmente como exemplo de olho ou exemplo de não-olho. A distribuição final da base se encontra na Tabela 3.2. Nesta tabela a primeira coluna informa as classes em que as imagens foram classificadas; a segunda, a base de origem das imagens; a terceira, a quantidade absoluta de imagens; a quarta, o subtotal de imagens agrupado por classes; a quinta, apresenta a quantidade relativa de imagens e a sexta, apresenta a quantidade relativa de imagens agrupada por classes. Exemplos de amostras de olho são ilustrados nas Figuras de 3.29(a) a 3.29(f), enquanto amostras do tipo não-olho são exemplificadas nas Figuras de 3.29(g) a 3.29(l). Estas amostras foram utilizadas no processo de treinamento da rede neural. Na determinação do rótulo, só foram considerados exemplos de olhos cuja pupila estivesse na região central da imagem. Por exemplo, os olhos contidos nas Figura 3.29(f) e 3.29(l) foram extraídos da mesma imagem de face. No entanto, o primeiro foi classificado como exemplo de olho e o segundo como exemplo de não-olho.

A etapa de extração de características é baseada na Análise de Componentes Principais (PCA). Dada a dificuldade na determinação de padrões em um conjunto de dados com alta dimensionalidade, a técnica PCA foi utilizada como ferramenta para seleção de características discriminativas e, conseqüentemente, para redução do tamanho do vetor de entrada da rede neural do módulo de aprendizagem. O procedimento para determinação da matriz de extração de componentes principais descrito nos parágrafos seguintes foi aplicado individu-

Tabela 3.2: Distribuição da base de imagens de olhos e não-olhos.

Rótulo	Base	Quantidade	Subtotal	Percentual	Subtotal
Olhos	BioID	1122	1320	42,5%	50%
	<i>World Wide Web</i>	198		7,5%	
Não-olhos	BioID	1122	1320	42,5%	50%
	<i>World Wide Web</i>	198		7,5%	
Total		2640	2640	100,0%	100,0%

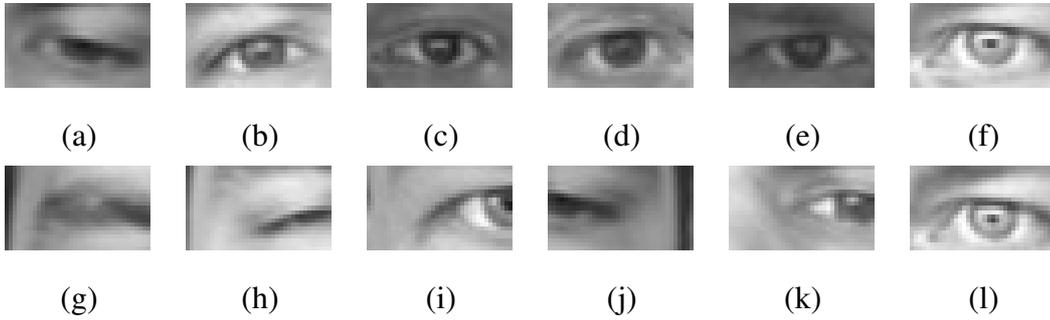


Figura 3.29: Exemplos de olhos e não-olhos utilizados no treinamento da rede neural. Exemplos de olhos de (a) a (f) e exemplos de não-olhos de (g) a (l).

almente na base de olhos e não-olhos.

Para possibilitar a aplicação da técnica PCA, cada imagem da base de olhos e não-olhos foi convertida em um vetor utilizando a equação

$$\rho((x \times w) + y) = \varepsilon(x, y); \quad x = 1, \dots, w, \quad e \quad y = 1, \dots, h; \quad (3.51)$$

em que w e h são os valores da largura e da altura das imagens, respectivamente. No caso particular da abordagem proposta, $w = 32$ e $h = 21$ são utilizados.

Para a determinação da matriz de redução de dimensionalidade da técnica PCA é necessária a determinação de um vetor médio $\bar{\rho}$ para cada conjunto de k vetores (agrupados em exemplos de olhos e não-olhos) obtidos no passo anterior ($k = 1320$),

$$\bar{\rho} = \frac{\sum_{i=1}^k \rho(i)}{k}. \quad (3.52)$$

Em seguida, o conjunto de vetores é ajustado através da subtração de cada vetor pelo

vetor médio ρ , gerando um novo conjunto de vetores ρ_{ajust} ,

$$\rho_{ajust}(i) = \rho(i) - \bar{\rho}; i = 1, \dots, k. \quad (3.53)$$

Covariância é uma medida de dispersão em um conjunto de dados [Smi02]. Uma matriz de covariância de tamanho $wh \times wh$ é gerada conforme a equação,

$$\sigma(j, i) = \frac{\sum_{p=1}^k \rho_{ajust}(i) \times \rho_{ajust}(j)}{k - 1}; i = 1, \dots, (wh) \text{ e } j = 1, \dots, (wh); \quad (3.54)$$

em que $\sigma(j, i)$ é a covariância entre os elementos nas posições i e j dos vetores contidos em cada conjunto.

O conjunto de autovalores τ_{valor} e autovetores normalizados τ_{vetor} da matriz de covariância σ (matriz quadrada) são calculados. Em seguida, os pares de autovalores e autovetores são ordenados de acordo com a magnitude de cada autovalor, em uma lista σ_{ord} ,

$$\sigma_{ord} = \{(\tau_{valor}(1); \tau_{vetor}(1)), \dots, (\tau_{valor}(wh); \tau_{vetor}(wh)) | \tau_{valor}(n) \leq \tau_{valor}(n + 1)\}. \quad (3.55)$$

Os autovetores associados aos autovalores de maior magnitude possuem uma maior relevância na descrição dos dados contidos no conjunto. Como o objetivo é selecionar características discriminantes para o contexto em questão, o autovetor $\tau_{vetor}(wh)$ (o qual é associado ao maior autovalor) é descartado da lista σ_{ord} . Visto que este autovetor aproxima-se do vetor médio do conjunto $\bar{\rho}$ que, a princípio, não possui características discriminativas.

As matrizes dos conjuntos de olhos e não-olhos resultantes do processo descrito anteriormente têm suas dimensionalidades reduzidas considerando a escolha de um pivô: todos os autovetores cujos índices são menores que o pivô são descartados. O calculo do pivô requer um vetor adicional que armazena a distribuição acumulada normalizada dos autovalores em σ_{ord} ,

$$\tau_{acum}(i) = \frac{\tau_{acum}(i - 1) + \tau_{valor}(i)}{\tau_{valor}(wh - 1)}; i = 1, \dots, wh - 1. \quad (3.56)$$

A fórmula que define o pivô p para uma retenção de 95% da variabilidade dos dados é dada por

$$p = \{i | \tau_{acum}(i - 1) < 0,95 \leq \tau_{acum}(i)\}; i = 1, \dots, wh - 1. \quad (3.57)$$

Como a magnitude do autovalor corresponde à sua representatividade no conjunto de exemplos, o pivô definido na Equação 3.57 tem a função de particionar o conjunto de au-

tovetores em duas partes. Os autovetores cujos autovalores acumulados representam, relativamente, até 5% da variação são descartados e os autovetores na partição com 95% de representatividade formam as colunas da matriz de redução de dimensionalidade da técnica PCA (σ_{olhos} e $\sigma_{nao-olhos}$) de cada conjunto. O tamanho desta partição foi definido empiricamente.

Em cada uma das lista ordenadas, correspondente aos exemplos de olhos e não-olhos, o valor encontrado para o pivô foi de 723. Portanto, a matriz de redução de dimensionalidade da técnica PCA (σ_{olhos} e $\sigma_{nao-olhos}$) de cada conjunto possui o tamanho 756×33 . O valor 756 corresponde ao tamanho dos exemplos (21×36) e o valor 33 corresponde a quantidade de autovetores selecionados. A matriz de redução de dimensionalidade da técnica PCA do módulo de geração de mapa de evidência baseado em aprendizagem é determinada através da concatenação das matrizes correspondentes a técnica PCA dos exemplos de olhos σ_{olhos} e não-olhos $\sigma_{nao-olhos}$, produzindo uma matriz σ_{pca} de dimensão 756×66 .

A extração de característica utilizando a técnica PCA ocorre com a multiplicação do vetor de entrada ρ , gerado a partir da imagem ε (em níveis de cinza), pela matriz σ_{pca} conforme,

$$\begin{aligned} \varepsilon_{pca\ 66 \times 1} &= extracao(\varepsilon) \\ &= \rho_{1 \times 756} \times \sigma_{pca\ 756 \times 66}. \end{aligned} \tag{3.58}$$

A mesma base utilizada na geração da matriz σ_{pca} foi utilizada no treinamento da rede utilizada para classificação de amostra. A arquitetura da rede neural utilizada possui um topologia triangular com 66 neurônios na camada de entrada, 33 neurônios na camada escondida e 1 neurônio em sua camada de saída. O tamanho da camada de entrada corresponde a saída da etapa de extração de componentes principais. O tamanho das camadas restantes é definido pela topologia triangular, que foi adotada após realização de experimentos preliminares com outras arquiteturas. Dez redes foram treinadas com inicialização aleatória de pesos, utilizando o algoritmo de *backpropagation* por 50.000 ciclos. O conjunto de amostras utilizado no treinamento foi particionado de acordo com a Tabela 3.3. Para cada rede treinada, foi selecionado o conjunto de pesos que apresentou o menor soma dos erros quadráticos (*Sum Squared Error* - SSE) entre todos os ciclos do treinamento. Ao final, foi selecionada a rede que apresentou a maior taxa de classificação na etapa de teste (95,20% de acerto).

Na definição da saída desejada pela rede foi atribuído o valor zero (0,0) para exemplos

Tabela 3.3: Distribuição da base de imagens de olhos e não-olhos.

Etapa	Imagens	Percentual
Treino	1584	60%
Validação	396	15%
Teste	660	25%
Total	2640	100%

de não-olhos e um (1,0) para exemplos de olhos. Portanto, o mapa de evidência η gerado na saída do módulo de geração de mapa de evidência baseado em aprendizagem corresponde à resposta apresentada pela rede neural quando aplicado às características extraídas de uma janela de busca deslizante que percorre todo o espaço de busca definido na face. Esta operação é representada pela equação

$$\eta(i, j) = \text{classificacao}(\varepsilon_{pca}(i, j)), \quad (3.59)$$

em que *classificacao* representa o valor de saída da rede neural.

A vantagem na utilização de Redes Neurais para detecção de olhos é a capacidade de generalização que as mesmas apresentam. A rede é capaz de indicar se uma imagem contém ou não um olho ainda que a imagem em questão não tenha sido utilizada no processo de treinamento.

Exemplos de mapas gerados através do módulo de geração de mapa de evidência baseado em aprendizagem são apresentados nas Figuras 3.30(a) e Figura 3.30(b). O mapa na Figura 3.30(a) apresenta um ótimo resultado para uma face com iluminação lateral. Um exemplo de desempenho satisfatório em faces com traços orientais é ilustrado na Figura 3.30(b). Finalmente, a Figura 3.30(c) apresenta um mapa com forte dispersão, mas com picos concentrados nas regiões centrais dos olhos.

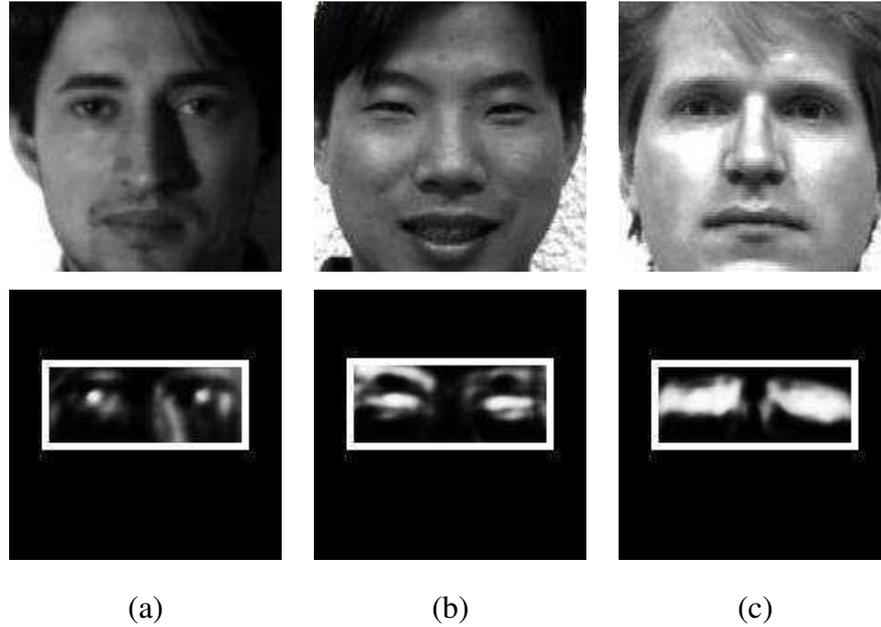


Figura 3.30: Mapas de evidência baseados em aprendizagem. Áreas de busca delimitadas por retângulos brancos.

3.7 Geração de mapa de evidência baseado em modelos

Métodos de decisão teórica para o reconhecimento de padrões utilizam funções de decisão ou discriminantes [GW92]. Seja um vetor de padrões n -dimensional

$$v = (v(1), v(2), \dots, v(n))^T. \quad (3.60)$$

Para m classes de padrões $\phi_1, \phi_2, \dots, \phi_m$, o problema básico do reconhecimento de padrões por decisão teórica é encontrar m funções de decisão (ou discriminantes) $f_1(v), f_2(v), \dots, f_m(v)$ com a propriedade que, se o padrão v pertencer à classe ϕ_i então

$$f_i(v) > f_j(v); \quad j = 1, \dots, M; \quad j \neq i. \quad (3.61)$$

O módulo de geração de mapa de evidência baseado em modelos utiliza uma simplificação da estratégia de reconhecimento de padrões por decisão teórica apresentada no parágrafo anterior. Para detecção de olhos, apenas duas classes de padrões ϕ_{olhos} e $\phi_{nao-olhos}$ foram consideradas. Além disto, uma única função de decisão f_{olhos} foi definida. O mapa obtido na saída deste módulo contém os resultados da avaliação da função f_{olhos} para cada ponto dentro do espaço de busca definido na Seção 3.4. O diagrama de blocos do módulo de geração

de mapa de evidência baseado em modelos se encontra na Figura 3.31 e seus pormenores são apresentados nos próximos parágrafos.

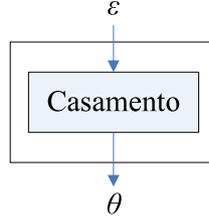


Figura 3.31: Diagrama do módulo de geração de mapa de evidência baseado em modelo.

A função de decisão utilizada no reconhecimento de padrões de olhos considera como base a correlação entre uma subimagem modelo v de tamanho $w_v \times h_v$ e subimagens de mesmas dimensões contidas em uma imagem ε (em níveis de cinza) de tamanho $w_\varepsilon \times h_\varepsilon$. A correlação entre ε e v é dada por

$$c(s, t) = \sum_x \sum_y \varepsilon(x, y) v(x - s, y - t); s = 1, \dots, w_\varepsilon \text{ e } t = 1, \dots, h_\varepsilon. \quad (3.62)$$

A soma é realizada sobre a região em que ε e v se sobrepõem. Este procedimento é ilustrado na Figura 3.32, a origem de ε é definida no canto superior esquerdo e a origem de v se localiza no centro. Para qualquer valor de (s, t) em ε , a avaliação da Equação 3.62 gera um valor para c . À medida que s e t variam, v é deslocada sobre ε , resultando na função $c(s, t)$. O valor máximo de $c(s, t)$ indica a posição em que ε apresenta um melhor casamento com v .

A função de correlação dada na Equação 3.62 possui a desvantagem de ser sensível a mudanças na amplitude da intensidade dos *pixels* de ε e de v . Como c é diretamente proporcional à imagem ε e ao padrão v , os valores de c não são normalizados. Para contornar esta desvantagem, a solução proposta por Gonzalez [GW92] foi adotada na implementação do módulo em questão. Esta solução consiste em calcular o coeficiente de correlação, ao invés da correlação. O mapa gerado na etapa de casamento utiliza esta solução, e sua operação é matematicamente descrita por

$$\begin{aligned} \theta(t, s) &= \text{casamento}(\varepsilon) \\ &= \frac{1}{2} + \frac{\sum_x \sum_y (\varepsilon(x, y) - \bar{\varepsilon}(x, y))(v(x - s, y - t) - \bar{v})}{2\sqrt{\sum_x \sum_y (\varepsilon(x, y) - \bar{\varepsilon}(x, y))^2 \sum_x \sum_y (v(x - s, y - t) - \bar{v})^2}} \end{aligned} \quad (3.63)$$

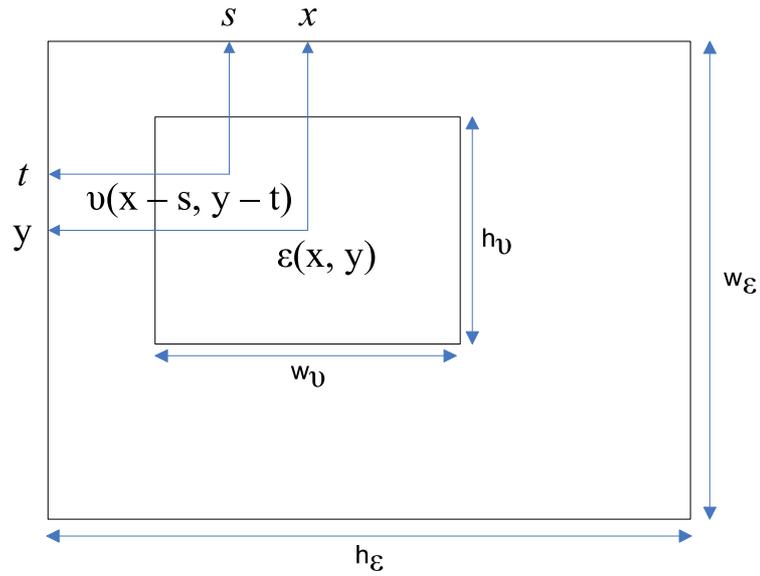


Figura 3.32: Arranjo para obtenção da correlação.

em que $s = 1, \dots, w_\varepsilon$ e $t = 1, \dots, h_\varepsilon$, \bar{v} é o valor médio dos pixels em $v(x, y)$, $\bar{\varepsilon}(x, y)$ é o valor médio de $\varepsilon(x, y)$ na região coincidente com a posição corrente de v . O coeficiente de correlação $\theta(x, y)$ varia no intervalo $[0, 1]$.

O padrão de olho v utilizado foi inspirado no trabalho de Akashi [AWT06] e se encontra ilustrado na Figura 3.33. O tamanho da matriz é 19×9 , sendo a área cinzenta correspondendo ao valor 154 (tom de cinza). A elipse tem eixos de tamanho 17 e 7, e é preenchida com o valor de intensidade máxima (255), enquanto o círculo tem raio 7 e valor de intensidade mínima (0). Tanto a elipse quanto o círculo são posicionados no centro do padrão.



Figura 3.33: Padrão de olho utilizado na etapa de casamento.

Exemplos de mapas gerados através do módulo de geração de mapa de evidência baseado em modelos são apresentados nas figuras seguintes. É possível perceber, nas Figuras 3.34(a) e 3.34(b), que a intensidade da resposta aumenta quando o padrão se “encaixa” na região dos olhos. Um caso em que a eficiência do casamento de modelos é satisfatória mesmo quando

o olho se apresenta semi-aberto é apresentado na Figura 3.34(c). Uma possível limitação do modelo adotado está relacionado com a aplicação do mesmo em olhos com íris claras.

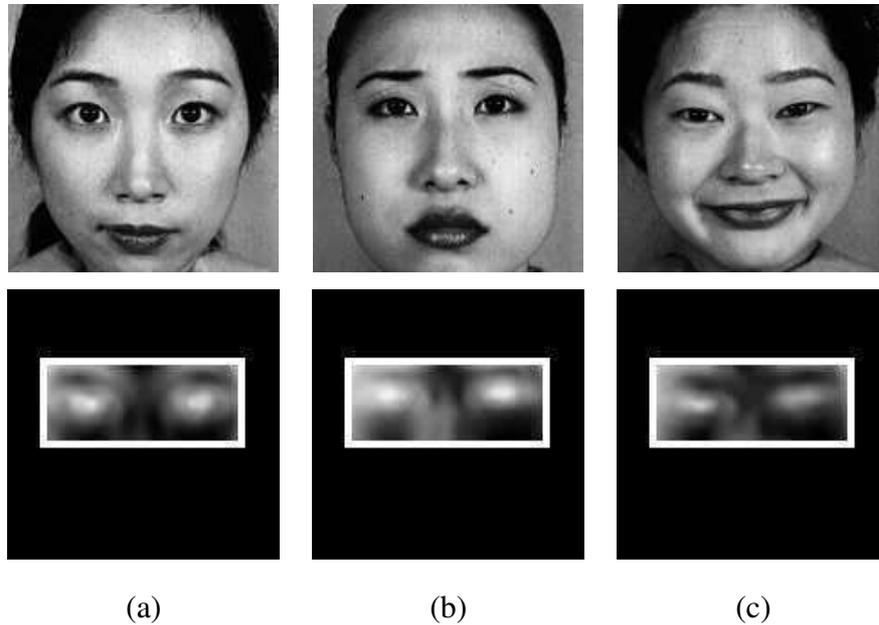


Figura 3.34: Mapas de evidência baseados em modelos. Retângulos brancos delimitam áreas de busca.

3.8 Combinação de mapas de evidência

Em tarefas relacionadas com o reconhecimento de padrões, tem se tornado crescente o uso de múltiplos classificadores, de forma a aumentar a taxa de acerto de sistemas de classificação. Muitos comparativos utilizando regras alternativas de combinação têm sido realizados [ISB03], dentre as quais, as mais utilizadas são:

- Regra do Produto: obtém uma classificação baseada na multiplicação das saídas apresentadas pelos classificadores;
- Regra da Média: calcula um valor médio da classificação a partir dos resultados individuais de cada classificador;
- Regra do *Ranking*: determina a classificação utilizando uma ordenação ascendente ou descendente dos resultados gerados pelos classificadores;

Bilmes e Kirchhoff [BK00] apresentam estudos que indicam que a Regra do Produto apresenta um melhor resultado se os classificadores presentes na combinação estão correlacionados. A Regra da Média é recomendada quando os classificadores não estão correlacionados entre si. A Regra do *Ranking* é mais robusta quando existem elementos com alta discrepância em relação ao padrão.

Para todos este conjunto de regras, podem haver variações que incluam uma estratégia de ponderação para cada classificador utilizado. Estas regras e/ou variações são utilizadas no momento da combinação do mapas de evidência gerados pelo módulo de geração de mapas.

O objetivo do módulo de combinação de mapas de evidência é produzir um mapeamento topográfico de que indique a evidência da ocorrência de olhos em uma face. Este mapeamento é composto pela combinação de diversos mapas, em que cada mapa representa a distribuição espacial da resposta apresentada por uma das estratégias de detecção de olhos descritas nas Seções 3.5, 3.6 e 3.7. O diagrama do módulo de combinação de mapas é ilustrado na Figura 3.35.

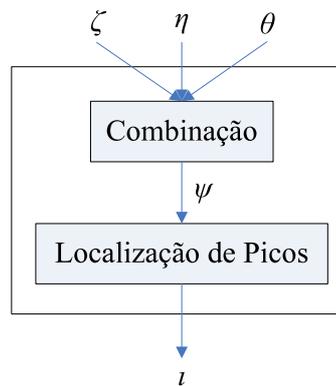


Figura 3.35: Diagrama do módulo de combinação de mapas de evidência.

A operação da etapa de combinação pode ser representado pela equação

$$\psi = \text{combinacao}(\zeta, \eta, \theta), \quad (3.64)$$

em que ψ é o mapa de evidência resultante e *combinacao* representa a função utilizada na combinação. Na implementação do módulo de combinação foram implementadas três funções que correspondem a cada uma das regras de combinação descritas no início desta seção. Na Tabela 3.4, são apresentados as regras e as respectivas definições de funções.

Tabela 3.4: Funções de combinação utilizadas no módulo de combinação de mapas de evidência.

Regra	Função
Produto	$produto(i, j) = \zeta(i, j) \times \eta(i, j) \times \theta(i, j)$
Média	$media(i, j) = \frac{\zeta(i, j) + \eta(i, j) + \theta(i, j)}{3}$
Ranking	$ranking(i, j) = maximo(\zeta(i, j), \eta(i, j), \theta(i, j))$

Finalmente, a etapa de localização de picos é responsável por encontrar as coordenadas do mapa de evidência ψ onde a função de combinação (*combinacao*) apresenta a maior resposta. Os dois picos encontrados, um localizado na parte esquerda do mapa e outro na parte direita, correspondem as coordenadas dos olhos direito e esquerdo, respectivamente. Esta etapa é descrita por

$$l_{dir} = (x, y) | x \in [0, \frac{w_\psi}{2}], y \in [0, h_\psi] e \psi(x, y) \geq \psi(i, j) \text{ com } x \neq i \text{ e } y \neq j; \quad (3.65)$$

$$l_{esq} = (x, y) | x \in [\frac{w_\psi}{2} + 1, w_\psi], y \in [0, h_\psi] e \psi(x, y) \geq \psi(i, j) \text{ com } x \neq i \text{ e } y \neq j. \quad (3.66)$$

Em que w_ψ e h_ψ correspondem a largura e altura do mapa de evidência ψ , $\frac{w_\psi}{2}$ é coordenada no eixo X que delimita a fronteira entre os dois lados do mapa e l_{dir} e l_{esq} , são os vetores com as coordenadas dos olhos direito e esquerdo, respectivamente.

A expectativa na utilização de um resultado composto a partir das respostas apresentadas por diversas estratégias é que a combinação de evidências ressalte a localização dos olhos e que, na ocorrência de situações de desvantagem para utilização de uma estratégia, esta situação seja compensada pelas respostas das outras estratégias utilizadas. Em resumo, a composição de estratégias deve apresentar um resultado, na média, superior ao produzido por cada estratégia de forma individual.

3.9 Sumário

A abordagem de detecção de olhos proposta é dividida em módulos. Na etapa inicial ocorre uma redução do espaço de busca através da detecção de faces. As faces detectadas passam um processamento para atenuação de sombras geradas em caso de captura com iluminação

lateral. Em seguida, as faces sofrem um ajuste de brilho e contraste com operações pontuais (expansão ou equalização de histograma). Uma etapa de redução de espaço é executada em seguida através da limitação da busca para uma região pré-definida a partir de estatísticas. A imagem resultante da restrição de busca é processada utilizando três abordagens: uma abordagem centrada na análise da aparência, outra abordagem baseada em aprendizagem e uma última que utiliza um esquema de casamento de modelos. Os resultados gerados pelas abordagens são combinados utilizando regras de combinação de classificadores e, finalmente, os picos definidos em cada lado do resultado da combinação indicam a posição dos olhos detectados.

No Capítulo 4, são apresentados os detalhes da organização e do funcionamento do *software* implementado com base na abordagem proposta neste capítulo.

Capítulo 4

Implementação da abordagem proposta

No Capítulo 3, foi detalhada a abordagem para detecção de olhos em faces humanas proposta nesta dissertação. Para realização dos experimentos comparativos de validação descritos no Capítulo 5, foi necessário implementar um sistema para detecção de olhos de acordo com as etapas e o fluxo de dados descritos na abordagem proposta. Neste capítulo são apresentados, de forma analítica, os detalhes organizacionais e funcionais do sistema criado.

A implementação foi realizada utilizando a linguagem C++ com o ambiente de desenvolvimento integrado (*Integrated Development Environment - IDE*) *Microsoft Visual Studio 2005*. A escolha desta linguagem é justificada por vários fatores:

- suporte a orientação a objetos;
- desempenho superior comprovado em relações a outras linguagens (como Java, por exemplo);
- uso difundido na área de Visão Computacional;
- existência de bibliotecas para leitura e escrita de imagens em diferentes formatos (como OpenCV [Ope], por exemplo);
- existência de bibliotecas para treinamento e execução de Redes Neurais (como FANN [Nis05] e SNNS [ZMV⁺], por exemplo);
- disponibilidade de utilização da IDE *Microsoft Visual Studio 2005* através do programa *MSDN Academic Alliance* [Mic].

Entretanto, há também algumas desvantagens com a escolha da linguagem e do ambiente de desenvolvimento. Dentre as desvantagens decorrentes do uso da linguagem C++ no ambiente *Visual Studio*, é possível citar: a dependência em relação ao sistema operacional *Microsoft Windows*; a dificuldade de portabilidade do código para outros compiladores, devido as peculiaridades da sintaxe e da interface para programação de aplicações (*Application Program Interface - API*) adotadas nesta IDE e a suscetibilidade a problemas de alocação/gerenciamento de memória em relação à outras linguagens com limpeza automática de memória (*Garbage Collector - GC*).

A organização dos sistema em termos de pacotes e classes é apresentada na Seção 4.1 e seu funcionamento é ilustrado na Seção 4.2.

4.1 Organização do sistema

Para garantir a organização, a possibilidade de expansão e o reaproveitamento de código, o sistema foi dividido em sete pacotes (*packages*). O diagrama de *deployment* apresentado na Figura 4.1 ilustra a organização definida para o sistema implementado: o pacote *ImageTypes* contém as classes criadas para os diferentes tipos de representações de imagens requeridos pelo sistema, o pacote *Detectors* armazena os detectores utilizados pelo sistema (detectores de face e olhos), o pacote *HomomorphicProcess* acomoda os algoritmos relacionados com o processamento homomórfico, o pacote *Enhancer* encapsula processos de melhoria de contraste e brilho, o pacote *Maps* disponibiliza abordagens para geração dos mapas de evidência, o pacote *Mergers* é composto pelas representações das regras de combinação disponíveis no sistema e o pacote *Shapes* define abstrações de elementos e formas geométricas básicas utilizadas pelos outros pacotes.

Na Figura 4.2, é apresentado o diagrama de classes do sistema implementado. O pacote *ImageTypes* contém uma classe base *MatrixImage* com rotinas básicas baseadas em *templates*. Derivando desta classe base, existem representações para imagens em ponto flutuante (*DoubleImage*) e em números complexos (*ComplexImage*), além de imagens multicanal (*Image*) e monocanal (*GrayImage*).

Como elemento centralizador das operações, aparece a classe *EyeDetectorApp* que gerencia operações de leitura e escrita de arquivos de imagens e geração de arquivos com as

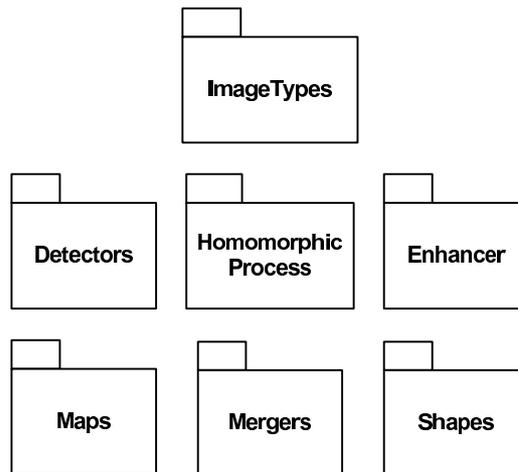


Figura 4.1: Diagrama de *deployment* do sistema implementado.

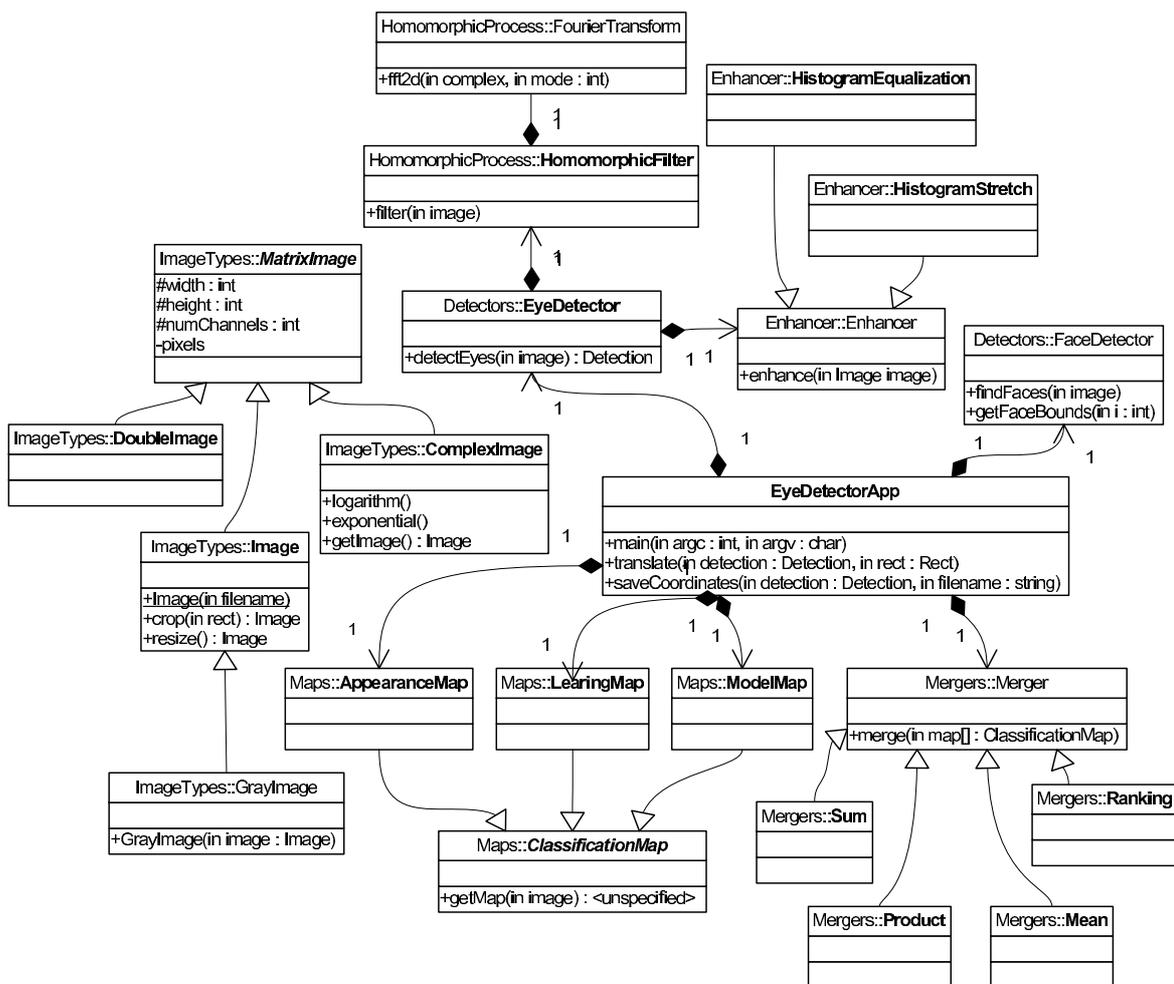


Figura 4.2: Diagrama de classes do sistema implementado.

coordenadas dos olhos. A classe *FaceDetector*, contida no pacote *Detectors*, comporta-se como um *wrapper*, possibilitando agregar o paradigma da orientação a objetos ao detector de faces contido na biblioteca OpenCV [Ope] (disponibilizado originalmente no paradigma procedural). Ainda faz parte deste pacote, a classe *EyeDetector* que controla o fluxo dos dados entre cada etapa do processo de detecção de olhos.

O pré-processamento para ajuste no contraste e no brilho nas imagens de faces é feito com base na *interface* da classe *Enhancer*. Duas implementações para esta *interface* estão disponíveis, *HistogramEqualization* e *HistogramaStretch*, que implementam os algoritmos de equalização e expansão linear de histograma, respectivamente. Estas três classes estão acomodadas no pacote *Enhancer*.

A funcionalidade de filtragem homomórfica, contida no pacote *HomomorphicProcess*, é executada pela classe *HomomorphicFilter* utilizando instância da classe *FourierTransform* para execução das conversões entre os domínios do espaço e da frequência.

A geração de mapas de evidência é feita no pacote *Maps*, por classes que estendem a classe base *ClassificationMap*. Na implementação presente neste pacote, foram disponibilizadas três classes derivadas da classe principal. A classe *AppearanceMap* gera um mapa de evidência utilizando técnicas de análise de aparência com Redes Neurais. Outra rede neural é utilizada para geração da mapa de evidência na classe *LearningMap*. No último mapa de evidência disponibilizado, é utilizada uma estratégia de casamento de modelos presente na classe *ModelMap*.

A combinação dos mapas de evidência gerados é feita pelas classes contidas no pacote *Mergers*. Estas classes implementam algumas regras de combinação, como a regra do produto (classe *Product*), a regra da média (classe *Mean*) e a regra do *ranking* (classe *Ranking*). Todas estas classes derivam da classe *Merger*, também contida neste pacote.

4.2 Funcionamento do sistema

A sequência de execução do sistema segue a mesma idéia da abordagem proposta no Capítulo 3. Uma visão geral desta sequência é apresentada no diagrama da Figura 4.3. A classe *EyeDetectorApp* carrega uma imagem de um arquivo cujo nome é fornecido como argumento no construtor de um objeto da classe *Image*. Esta imagem é passada como parâmetro para

a etapa de detecção de faces descrita na Seção 3.1 e executada pelo método *findFaces* da classe *FaceDetection*. Em seguida, ocorre uma iteração sobre cada face detectada, sendo as coordenadas do retângulo que delimitam a face acessadas através do método *getFaceBounds* e o recorte da face sendo obtido através de chamada ao método *crop* da classe *Image*. Cada face recortada é passada como argumento para detecção de olhos via método *detectEyes* da classe *EyeDetector*. Após a detecção dos olhos, as coordenadas sofrem uma translação (método *translate*) para que apresentem localizações em relação ao canto superior esquerdo da imagem. As coordenadas são salvas em arquivo através do método *saveCoordinates*.

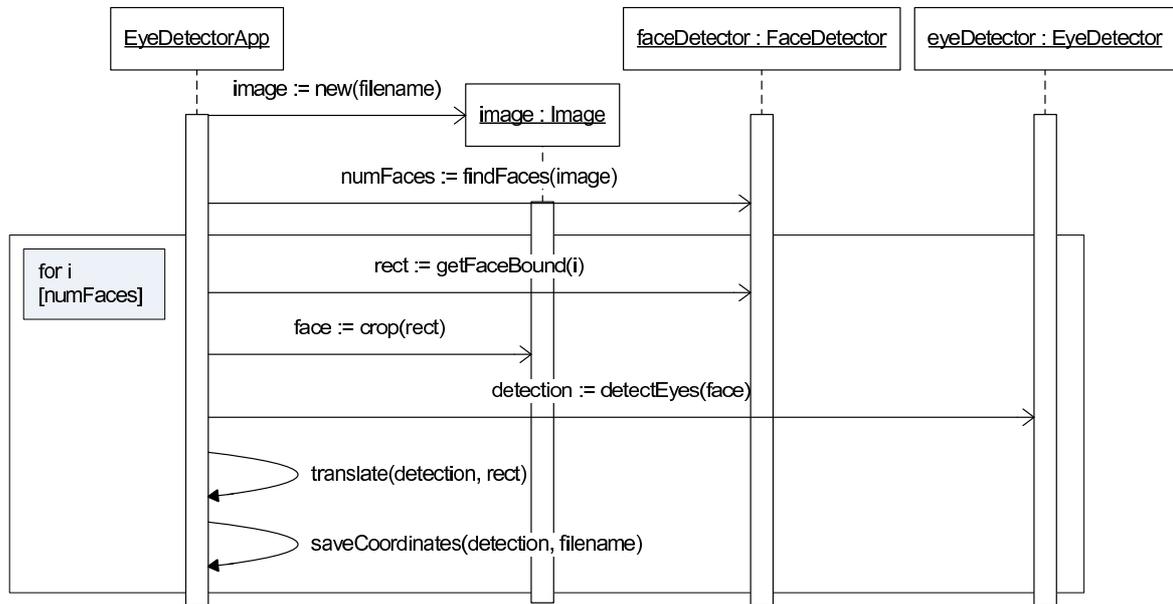


Figura 4.3: Diagrama da sequência de execução do programa principal.

O método de detecção de olhos *detectEyes* pode ser dividido em duas etapas: pré-processamento e geração/combinção de mapas de evidência. Na etapa de pré-processamento, apresentada na Figura 4.4, a imagem com a face é redimensionada para o tamanho 150×150 pixels através do método *resize* da classe *Image*. Em seguida, a imagem redimensionada sofre uma filtragem para atenuação de sombras com compensação de iluminação, descrita na Seção 3.2 e implementada através do método *filter* da classe *HomomorphicFilter*. Uma versão da imagem filtrada em níveis de cinza é gerada a partir da passagem de uma imagem colorida como argumento para o construtor de um objeto da classe *GrayImage*. Esta versão em níveis de cinza sofre um pré-processamento para melhoria de contraste e brilho através de aplicação um método polimórfico da classe abstrata *ImageEnhancer*. Duas implementações

para esta classe abstrata foram definidas no sistema, a classe *HistogramStretch* e a classe *HistogramEqualization* que correspondem às técnicas descritas nas Seções 3.3.1 e 3.3.2. A escolha da implementação que é feita através de argumentos passados pelo usuário na linha de comando para execução do sistema.

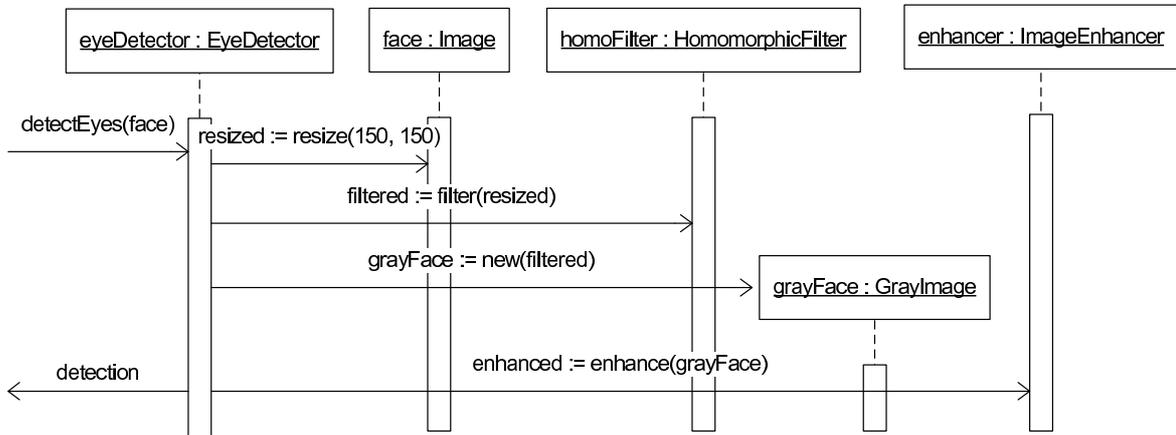


Figura 4.4: Diagrama de sequência do pré-processamento.

Para realização do procedimento de compensação de iluminação descrito na Seção 3.2, foi utilizado o método *filter* da classe *HomomorphicFilter*. Inicialmente, este método converte a imagem redimensionada da face para uma representação utilizando números complexos através da construção de imagem da classe *ComplexImage*. Em seguida, a operação de logaritmo é aplicada em todos os *pixels* com a chamada ao método *logarithm*. A representação no domínio da frequência é obtida com a execução do método *fft2d* da classe *FourierTransform* que recebe a imagem complexa e um *flag* de indicação de direção da conversão (*FORWARD*). A filtragem homomórfica é feita pelo método *homoFilter*. O método *fft2d* com flag de direção reversa (*BACKWARD*) é novamente executado para obter a forma de representação espacial. A operação de exponenciação é aplicada em todos os *pixels* através do método *exponential* da classe *ComplexImage*. O resultado obtido no final deste processo é convertido da representação em forma de números complexos para números inteiros através da execução do método *getImage* também da classe *ComplexImage*.

A segunda etapa do método de detecção de olhos *detectEyes* é apresentada na Figura 4.6. O procedimento de limitação do espaço de busca descrito na Seção 3.4 é embutido na etapa da geração de mapas de evidência. Os mapas de evidência são gerados levando em consideração apenas os *pixels* contidos dentro do espaço de busca definido. A obtenção

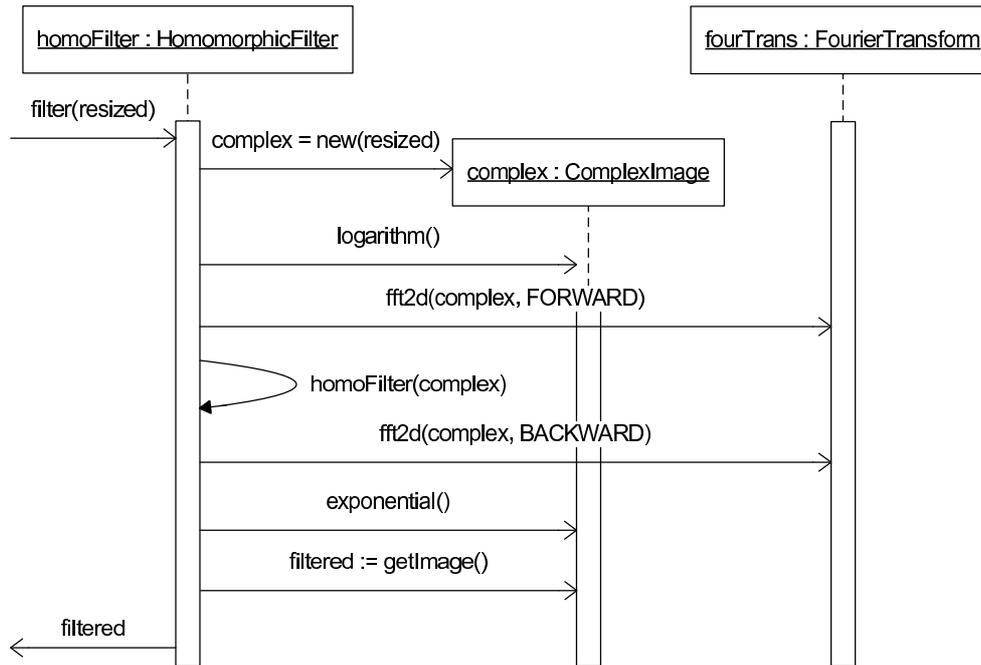


Figura 4.5: Diagrama de sequência da filtragem homomórfica.

dos mapas é feita através do método *getMap* definido na classe abstrata *ClassificationMap* e implementado pelas três classes que a estendem. A classe *AppearanceMap* gera um mapa baseado nas técnicas de análise de aparência descritas na Seção 3.5. A classe *LearningMap* produz um mapa utilizando a proposta de aprendizagem definida na Seção 3.6 e o mapa construído a partir do procedimento de casamento de modelos determinado na Seção 3.7 é gerado pela classe *ModelMap*. A combinação dos mapas é feita utilizando o método *merge* contido na *interface* da classe abstrata *Merger*. No sistema, três implementações para a classe *Merger* foram disponibilizadas: *Product*, *Mean* e *Ranking*. Cada uma delas utilizando uma abordagem de combinação de classificação diferente. Para detecção das coordenadas dos olhos obtidas no final desta etapa, o mapa gerado a partir da combinação das abordagens é passado para o método *findPeaks* que encontra as localizações onde o mapa apresenta as máximas respostas para cada um dos lados da face, as quais correspondem as posições dos olhos direito e esquerdo.

O detalhamento do funcionamento da geração de mapa de evidência através da análise da aparência, descrito na Seção 3.5, é apresentado na Figura 4.7. A imagem recebida como argumento pelo método *getMap* da classe *AppearanceMap* sofre uma varredura na área de busca definida na Seção 3.4. As intensidades dos canais RGB de cada *pixel* são armazenadas

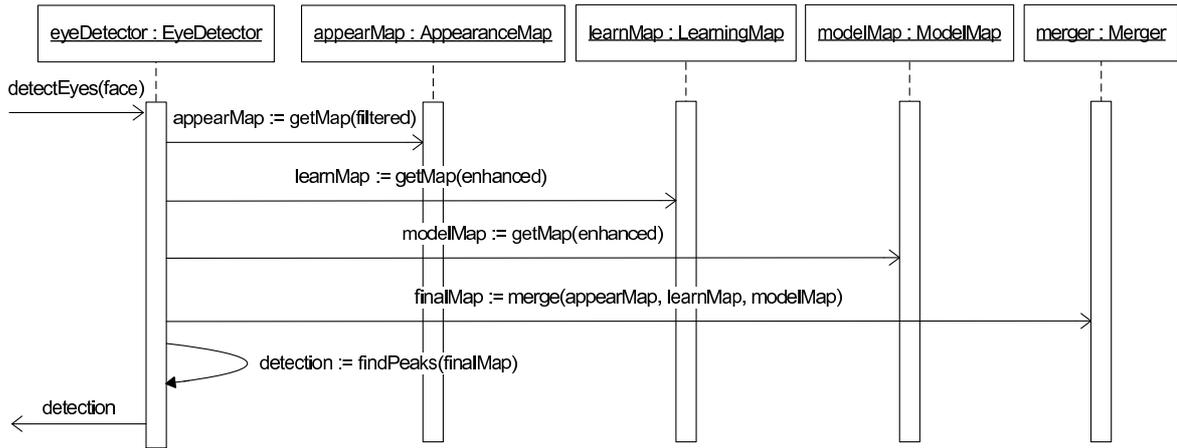


Figura 4.6: Diagrama de sequência de geração e combinação de mapas de evidência.

em um vetor através da execução do método *getPixel*. Este vetor é propagado pela rede neural (proposta na Seção 3.5) através do método *propagation* que também recebe os pesos da rede.

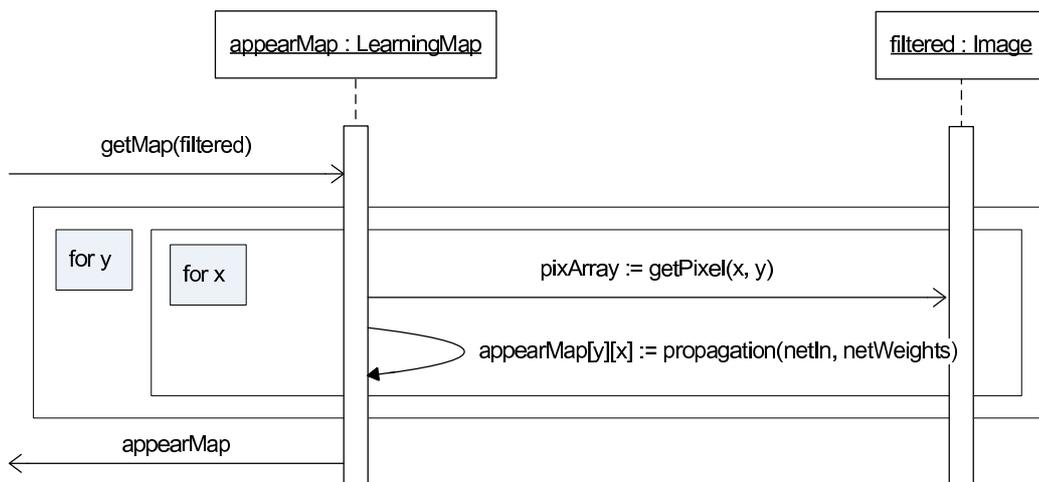


Figura 4.7: Diagrama de sequência da geração de mapa de evidência baseada na análise de aparência.

A sequência para geração de mapa de evidência baseado em aprendizagem de acordo com a Seção 3.6 está presente na Figura 4.8. A classe *LearningMap* gera um vetor a partir dos *pixels* contidos em uma matriz definida pelos argumentos da chamada ao método *asArray* da classe *Image*. A redução de dimensionalidade é obtida através da aplicação da técnica PCA no método *reduction*. A saída produzida pela técnica PCA é então propagada pela rede neural no método *propagation*. O resultado gerado pela rede é armazenado no mapa

retornado pelo método *getMap*.

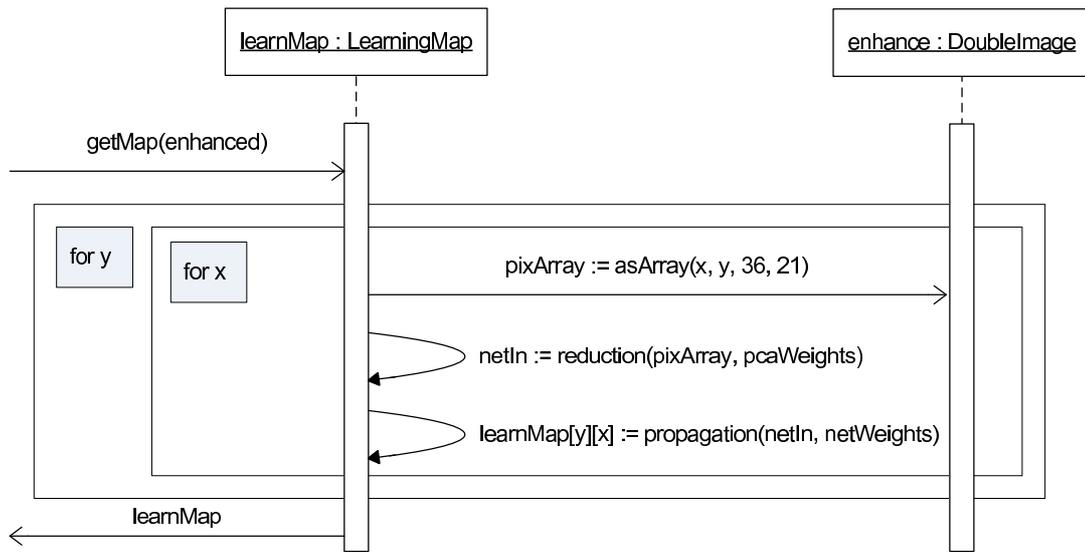


Figura 4.8: Diagrama de sequência da geração de mapa de evidência baseado em aprendizagem.

A sequência de operações para geração do mapa de evidência baseado em modelos, apresentado na Seção 3.7, é ilustrada na Figura 4.9. Primeiramente, a face recebida como entrada é convertida para uma representação em ponto flutuante através da construção de um objeto da classe *DoubleImage*. Uma segunda imagem em ponto flutuante é criada para representar o modelo. Este modelo é preenchido através de duas chamadas ao método *fillEllipse*. Uma chamada corresponde ao preenchimento da elipse branca da esclera e a outra ao círculo escuro da íris. O mapa retornado é o resultado da convolução do modelo com o espaço de busca da imagem recebida como parâmetro pelo método *getMap* da classe *ModelMap*. Esta convolução é executada pelo método *convolve* da classe *DoubleImage*.

4.3 Sumário

Para realização dos experimentos comparativos entre a abordagem proposta nesta dissertação e as propostas existentes, foi implementado um sistema de acordo com as definições apresentadas no Capítulo 3. A implementação do sistema utilizou a linguagem de programação C++ e foi dividida em sete pacotes para garantir a organização, a possibilidade de expansão e o reaproveitamento de código. A Tabela 4.1 apresenta a relação entre os procedimentos

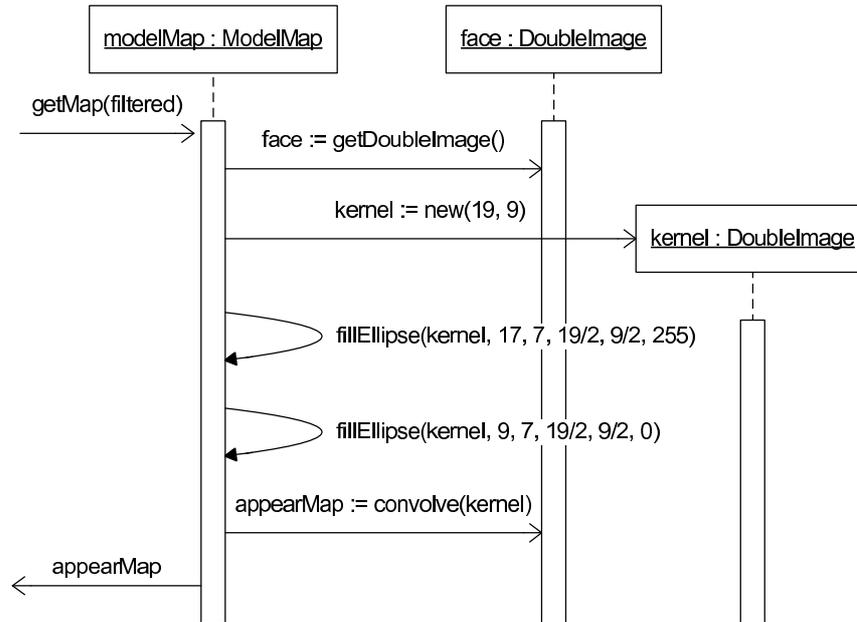


Figura 4.9: Diagrama de seqüência da geração de mapa de evidência baseado em modelos.

definidos na abordagem proposta (primeira coluna) e o pacote responsável pela respectiva implementação (segunda coluna).

No Capítulo 5, são apresentados os procedimentos aplicados e resultados produzidos durante a execução de experimentos comparativos entre a abordagem proposta e um subconjunto das abordagens discutidas na revisão bibliográfica presente no Capítulo 2.

Tabela 4.1: Relação entre procedimento proposto e respectivo pacote de implementação.

Procedimento	Pacote
Detecção de faces	<i>Detectors</i>
Compensação de iluminação	<i>HomomorphicProcess</i>
Pré-processamento com operações pontuais	<i>Enhancer</i>
Limitações do espaço de busca	<i>Maps</i>
Geração de mapa de evidência baseado na análise da aparência	<i>Maps</i>
Geração de mapa de evidência baseado em aprendizagem	<i>Maps</i>
Geração de mapa de evidência baseado em modelos	<i>Maps e ImageTypes</i>
Combinação de mapas de evidência	<i>Mergers</i>

Capítulo 5

Experimentos

O objetivo deste capítulo é descrever os experimentos realizados para comprovação do aperfeiçoamento obtido na abordagem proposta com a aplicação de filtragem homomórfica para compensação de iluminação lateral e melhoramento de brilho e contraste com a utilização de operações pontuais. Além disto, é apresentado um estudo comparativo entre a abordagem proposta (descrita no Capítulo 4) e uma seleção de trabalhos representativos, que foram apresentados na revisão bibliográfica (apresentada no Capítulo 2).

A Seção 5.1 apresenta uma visão geral das bases de imagens de faces utilizadas nos experimentos realizados. Na Seção 5.2 é definido o critério de tolerância a erro adotado nos experimentos. O impacto da utilização de compensação de iluminação lateral é apresentado na Seção 5.3. A Seção 5.4 traz um comparativo envolvendo brilho, contraste e as operações pontuais de expansão e equalização de histograma. A Seção 5.5 resume o experimento comparativo de desempenho entre a abordagem proposta com sistemas existentes.

5.1 Bases de imagens de faces

Quando considerado o desempenho apresentado por detectores de olhos, é importante observar a base de imagens de faces utilizada na avaliação do desempenho, visto que este fator influencia fortemente no resultado obtido [LPG⁺07]. Algumas bases utilizadas em trabalhos presentes na revisão bibliográfica não estão disponíveis publicamente, o que torna a utilização impossível. Em outros casos, o procedimento para obter acesso a base é complicado e moroso, desencorajando seu próprio uso.

Os artigos publicados por Huang et al [HSLW98], Fathi e Manzuri [FM04], Kumar et al [KRR02], D’Orazio et al. [DLSG04], Peng et al [PZB98], Shinjiro e Kawato [KT04] não indicam qual a base utilizada em testes para determinação do desempenho das soluções propostas para a detecção de olhos.

Os trabalhos de Xingming e Huangyuan [XH06] e Hua e Geng [ZHG04] utilizam as bases de imagens de faces SCUT e NJFace que aparentemente foram produzidas exclusivamente para realização destes trabalhos. Os mesmos não apresentam referências bibliográficas para estas bases. As referências apresentadas por Lin e Yang [LY04] para a base HHI e por Samad et al [SHT01] para a base Olivetti apresentam apontadores para arquivos não mais disponíveis na *World Wide Web*.

Em todos estes casos, permissão de acesso as bases de dados utilizadas foi solicitada através de correio eletrônico, mas não foi possível estabelecer contato com os autores. As tentativas de acesso as estas bases utilizando ferramentas de busca na *World Wide Web* (*Google*, *Yahoo* e *Live*) também não lograram êxito.

Obteve-se sucesso no contato através de correio eletrônico com os responsáveis pelas bases FERET [FER] e FRGC [FRG], mas verificou-se que as mesmas não são disponibilizadas para o público comum. A acesso as base FERET, utilizada por Wu e Trivedi [WT05] e Jin et al [JYS⁺06], e FRGC, utilizada por Wang et al [WGJW05], só é permitido através da existência de acordo entre a instituição mantenedora da base e a instituição que deseja fazer o uso, o que não foi possível ser concretizado no tempo do desenvolvimento deste trabalho.

Apesar dos problemas com as bases de dados acima, foi possível identificar e ter acesso a outras bases de dados de domínio público, com possibilidade de utilização no problema de detecção de olhos, a saber: IMM, Caltech e JAFFE.

A base de faces IMM engloba 240 imagens de faces de 40 indivíduos, todos sem óculos. A distribuição em termos de gênero é de 7 mulheres e 33 homens. As imagens foram capturadas em Janeiro de 2001 com resolução de 640×480 pixels usando o formato JPEG com uma camera Sony DV DCR-TRV900E. Variações de pose (face frontal ou rotacionada em aproximadamente 30 graus), expressões faciais (alegre, neutra e arbitrária) e iluminação (difusa e lateral) foram consideradas na captura das imagens, conforme apresentado nas Figuras de 5.1(a) a 5.1(d). Estruturas faciais como sobrancelhas, olhos, nariz, boca e queixo foram marcadas manualmente produzido um total de 58 pontos para cada face [SEL03].



Figura 5.1: Exemplos de imagens contidas na base IMM.

A base de imagens Caltech [Web] contém 450 imagens frontais de faces de 27 sujeitos capturadas por Markus Weber no Instituto de Tecnologia da Califórnia (*California Institute of Technology*). Esta base inclui imagens com variações de condições de iluminação em ambientes internos e externos; iluminação natural, lateral e com *flash* e alterações na expressão facial. Exemplos de imagens de faces contidas nesta base são ilustradas nas Figuras de 5.2(a) a 5.2(d). As imagens coloridas possuem resolução 896×592 *pixels*. O desempenho de 98,8% de acerto apresentado por Han et al [HKN02] foi obtido utilizando esta base, conforme discutido na Seção 2.2.1.

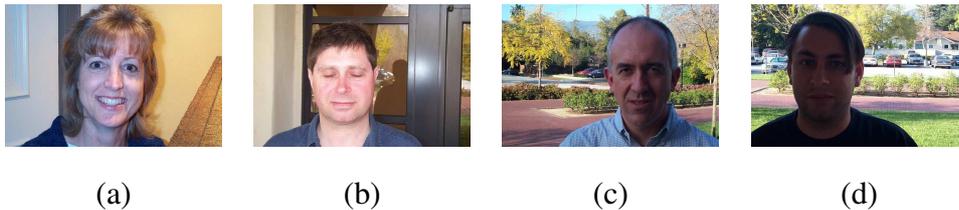


Figura 5.2: Exemplos de imagens contidas na base Caltech.

A base JAFFE contém 213 imagens de 10 modelos japonesas [LAKG98]. As imagens de cada modelo foram capturadas três ou quatro vezes enquanto apresentavam seis emoções básicas e uma face neutra, alguns exemplos destas imagens são mostrados nas Figuras de 5.3(a) a 5.3(d). As imagens foram disponibilizadas com tamanho 256×256 *pixels* em modo monocromático [Gro05]. O trabalho de Wang e Yin [WY05] utilizou esta base de imagens e apresentou uma taxa de acerto 95,8%, conforme descrito na Seção 2.2.2.

As principais características de cada uma das bases de imagens de face, utilizadas nos experimentos comparativos de melhoria de brilho e contraste com operações pontuais e de desempenho entre a abordagem proposta e sistemas existentes, são apresentadas na Tabela 5.1.

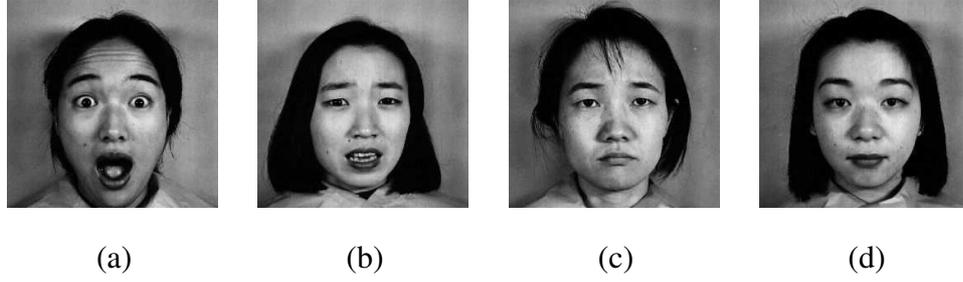


Figura 5.3: Exemplos de imagens contidas na base JAFFE.

Tabela 5.1: Características das base de imagens utilizadas nos experimentos.

Base	Imagens	Sujeitos	Cor	Variações			
				Expressão	Rotação	Iluminação	Cenário
IMM	240	40	Sim	Sim	Sim	Sim	Não
Caltech	450	27	Sim	Sim	Não	Sim	Sim
JAFFE	213	10	Não	Sim	Não	Não	Não

5.2 Tolerância a erro

A precisão na localização dos olhos é usualmente expressa como uma estatística de uma medida de erro, considerando como verdade (*ground truth*) os centros dos olhos marcados manualmente [MDWW01]. Para facilitar a comparação entre o desempenho de detectores de olhos quando testados em bases de imagens de faces diferentes, é recomendável a utilização de um procedimento para normalização de resultados baseado na escala da face, que pode ser estimada utilizando a distância entre os olhos [MGS07]. A forma de normalização adotada nesta dissertação segue a proposta de Jesorsky et al [JKF01] e é descrita pelas seguinte equações

$$d_{dir} = \frac{\sqrt{(X_{dir} - x_{dir})^2 + (Y_{dir} - y_{dir})^2}}{\sqrt{(X_{esq} - X_{dir})^2 + (Y_{esq} - Y_{dir})^2}} \quad (5.1)$$

$$d_{esq} = \frac{\sqrt{(X_{esq} - x_{esq})^2 + (Y_{esq} - y_{esq})^2}}{\sqrt{(X_{esq} - X_{dir})^2 + (Y_{esq} - Y_{dir})^2}} \quad (5.2)$$

em que d_{dir} e d_{esq} são as medidas de erro normalizadas para os olhos direito e esquerdo, respectivamente. (X_{dir}, Y_{dir}) e (X_{esq}, Y_{esq}) são as coordenadas dos centros dos olhos direito e esquerdo marcadas manualmente. As coordenadas (x_{dir}, y_{dir}) e (x_{esq}, y_{esq}) correspondem

às localizações centrais para os olhos direito e esquerdo produzidas pelo detector de olhos.

Maia [MGS07] define como aceitável um erro na detecção do centro dos olhos conforme expresso a seguir:

$$d \leq 0,25 \quad (5.3)$$

em que d representa a distância normalizada entre o olho marcado manualmente e o olho localizado pelo detector. Este critério é adotado como tolerância no confronto entre a abordagem proposta e os trabalhos utilizados nos experimentos comparativos. É importante observar que, apesar da importância da precisão da detecção, poucos trabalhos apresentam o erro médio obtido na realização de experimentos. Entre os trabalhos revisados no Capítulo 2, apenas Jin et al [JYS⁺06] ($2,92 \pm 1,10$ pixels em faces com distâncias intraoculares entre 25 e 120 pixels), Huang et al [HSLW98] (aproximadamente 6 pixels em faces com distâncias intraoculares entre 46 e 74 pixels) e Wang et al [WGJW05] (6,4 pixels, o tamanho médio das faces utilizadas não foi publicado) publicaram o erro médio obtido nos respectivos experimentos.

Todos os resultados produzidos nos experimentos que são apresentados a seguir foram obtidos de forma automática. Esta automação foi conseguida com o desenvolvimento de um sistema capaz de comparar as coordenadas produzidas pelo detector com as coordenadas geradas manualmente. A saída produzida contém estatísticas como erro médio e desvio padrão apresentado em um determinado conjunto de imagens e métricas de erro absoluto e erro normalizado para cada imagem processada. A Figura 5.4 ilustra o fluxograma adotado para geração destas estatísticas e métricas. As imagens são aplicadas no processo pré-definido de detecção de olhos, as coordenadas produzidas pelo detector de olhos são comparadas com as marcações manuais (*ground truth*) produzindo os valores desejados.

5.3 Avaliação do desempenho com uso da compensação de iluminação

O objetivo do experimento de avaliação do desempenho com uso da compensação de iluminação é comprovar a obtenção de melhores resultados na detecção de olhos com a aplicação de filtragem homomórfica para compensação de iluminação lateral. O experimento consiste

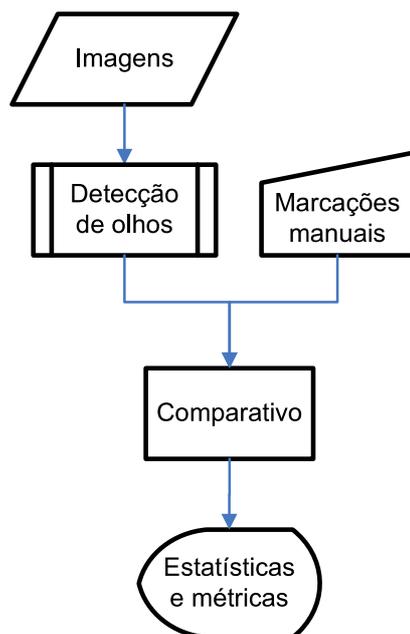


Figura 5.4: Fluxograma da geração de estatísticas e métricas.

na detecção de olhos com e sem a filtragem homomórfica, utilizando os módulos de geração de mapas de evidência baseados na análise de aparência (Seção 3.5), aprendizagem (Seção 3.6) e modelos (Seção 3.7).

Na realização deste experimento, foi utilizada a base IMM [SEL03], pois a mesma apresenta imagens coloridas com variações de iluminação lateral e é disponibilizada acompanhada das marcações manuais de diversos pontos da face (conforme visto na Seção 5.1). Os resultados apresentados neste experimento não incluem processos para melhoria de brilho e contraste. Portanto, o funcionamento do módulo de Pré-processamento com operações pontuais (abordado na Seção 3.3) foi inibido. Resultados relacionados à utilização deste módulo são apresentados na seção seguinte.

Nas Figuras de 5.5(a) a 5.5(d), apresentam-se os mapas de evidência produzidos utilizando a análise de aparência, aprendizagem e modelos além do mapa combinado (neste caso, utilizando a Regra do Produto), respectivamente, para uma determinada imagem da base IMM utilizando a técnica de compensação de iluminação. As Figuras 5.6(a) a 5.6(d) correspondem aos mesmos mapas para a mesma imagem, porém sem a utilização de compensação de iluminação. Comparando as figuras, é possível perceber o efeito benéfico da compensação de iluminação. A resposta dos classificadores apresenta uma maior simetria

na direção horizontal nas imagens processadas com a compensação de iluminação. Nas imagens que não passaram pelo processo de compensação de iluminação, a resposta apresenta uma atenuação no lado sombreado da face.

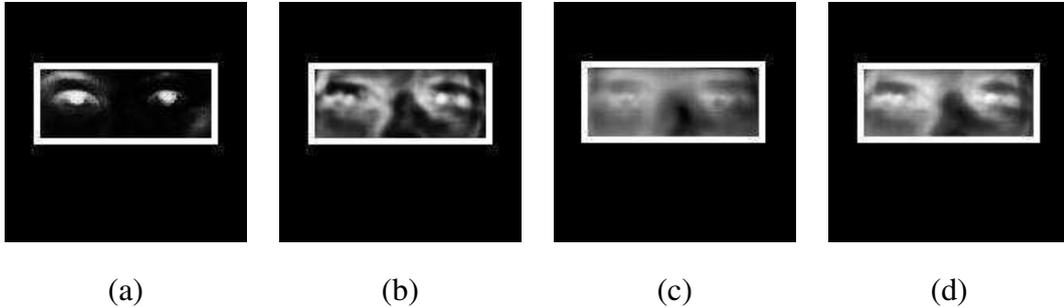


Figura 5.5: Mapas de evidências obtidos com a utilização da compensação de iluminação.

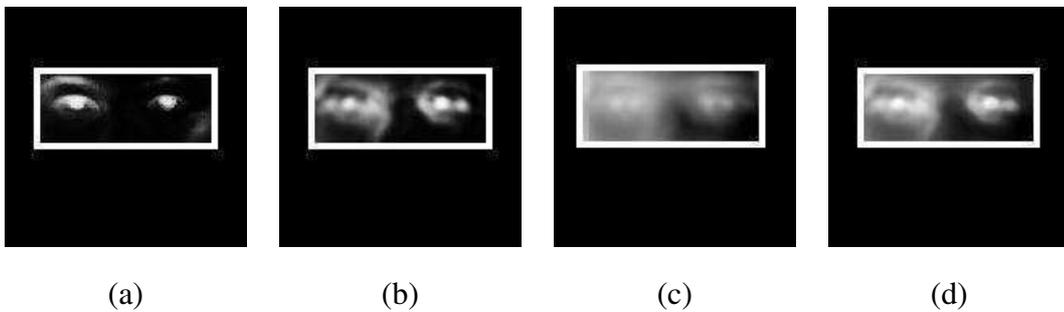


Figura 5.6: Mapas de evidências obtidos sem a utilização da compensação de iluminação.

Os resultados obtidos para toda a base são sumarizados na Tabela 5.2. A primeira coluna indica as alternativas de abordagem ou combinação de abordagens adotada (Seção 3.8). Na segunda coluna, é indicado o uso ou não da compensação de iluminação. Nas colunas restantes, são apresentados a taxa de acerto (considerando a tolerância declarada na Seção 5.2), o erro médio (\bar{d}) e o desvio padrão (σ) obtido em cada cenário do experimento.

O experimento de avaliação de desempenho com uso da compensação de iluminação, permite concluir que a utilização da filtragem homomórfica proporciona um ganho considerável na taxa de acerto e na redução do erro médio e desvio padrão em todas as alternativas de combinação de classificadores utilizadas. Outra constatação obtida com a realização deste experimento foi quanto ao desempenho superior apresentado pelas estratégias baseadas em combinação de classificadores, quando comparadas com a aplicação dos classificadores de forma isolada. É possível afirmar que a regra que apresentou o melhor desempenho dentre todas as estratégias de combinação aplicadas foi a Regra do Produto, indicando sua melhor

Tabela 5.2: Resultados obtidos com compensação de iluminação lateral na base IMM.

Regra	Compensação	Acerto	Erro médio (\bar{d})	Desvio padrão (σ)
Produto	Sim	97,041%	0,051	0,066
	Não	88,690%	0,124	0,102
Média	Sim	93,976%	0,047	0,065
	Não	89,222%	0,120	0,144
Ranking	Sim	92,857%	0,049	0,079
	Não	87,193%	0,106	0,124
Aprendizagem	Sim	90,617%	0,049	0,080
	Não	87,381%	0,106	0,124
Casamento	Sim	84,615%	0,121	0,103
	Não	83,529%	0,136	0,093
Aparência	Não	82,738%	0,081	0,170

adequação na resolução do problema proposto.

Alguns exemplos de imagens produzidas no experimento de avaliação de desempenho com uso da compensação de iluminação são apresentados nas Figuras de 5.7(a) a 5.7(f). As Figuras de 5.7(a) a 5.7(c) correspondem as detecções obtidas com o uso da compensação de iluminação, enquanto as Figuras de 5.7(d) a 5.7(f) correspondem as detecções obtidas sem a compensação da iluminação. Comparando as imagens, é possível comprovar visualmente a informação presente na Tabela 5.2: a aplicação da compensação de iluminação melhora a taxa de detecção em imagens capturadas em condições de iluminação lateral. Um conjunto mais abrangente de detecções geradas neste experimentos é apresentado no Apêndice A.

5.4 Comparativo entre operações pontuais para melhoria de brilho e contraste

O resultado esperado com a execução do experimento comparativo entre operações para melhoria de brilho e contraste, disponíveis no módulo de Pré-processamento como operações

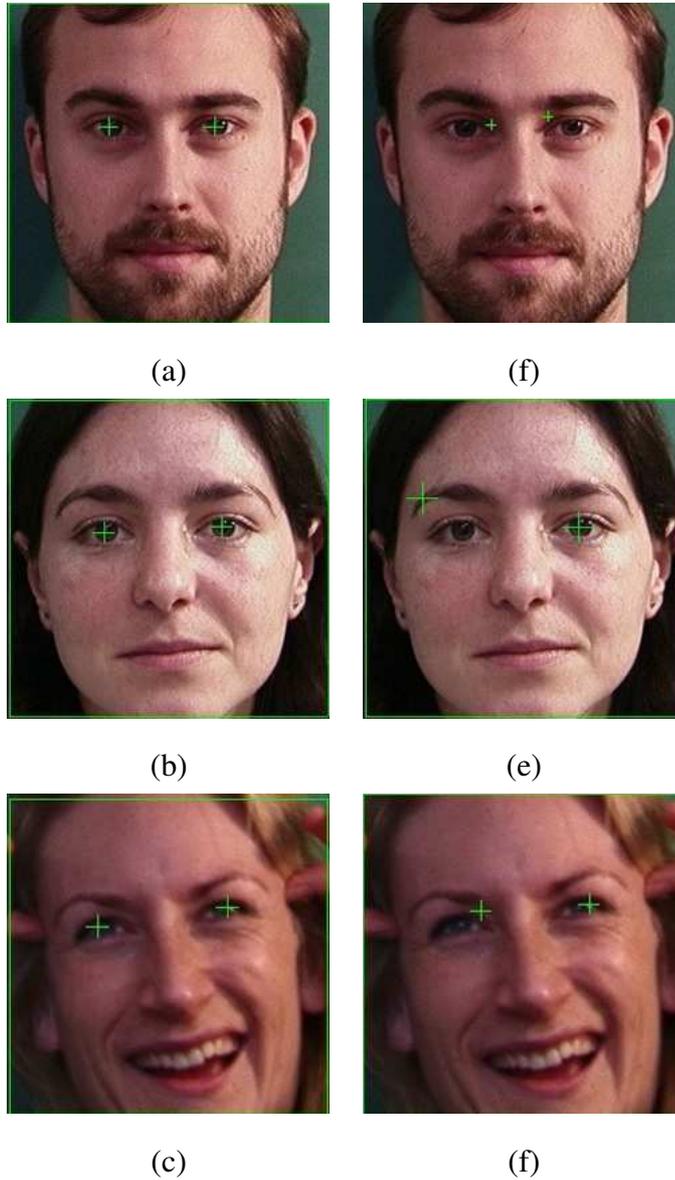


Figura 5.7: Comparação de desempenho relativa ao uso de compensação de iluminação.

pontuais, é a indicação da operação mais adequada (expansão ou equalização de histograma) para detecção de olhos, utilizando a estratégia de combinação de classificadores proposta nesta dissertação. Este experimento consistiu da aplicação da expansão ou da equalização de histograma na etapa de pré-processamento, antes da geração dos mapas de evidência e da utilização de compensação de iluminação e de classificadores combinados, devido à comprovação de melhor desempenho na detecção de olhos com o uso destas configurações.

A Tabela 5.3 apresenta os resultados produzidos neste experimento. A alternativa de combinação de classificadores adotada (Seção 3.8) é indicada na primeira coluna. A operação pontual de pré-processamento aplicada é apresentada na segunda coluna. Nas colunas restantes, são apresentadas a taxa de acerto (considerando a tolerância declarada na Seção 5.2), o erro médio (\bar{d}) e o desvio padrão (σ), obtidas no experimento. A combinação formada pela equalização seguida de expansão linear de histograma não foi considerada porque o efeito da expansão linear após a equalização é nulo, visto que a equalização já expande o histograma.

Tabela 5.3: Resultados obtidos com pré-processamento para melhoria de brilho e contraste na base IMM.

Regra	Operação Pontual	Acerto	Erro médio (\bar{d})	Desvio padrão (σ)
Produto	Expansão	97,605%	0,046	0,063
	Expansão e Equalização	96,471%	0,060	0,073
	Equalização	96,471%	0,060	0,075
Média	Expansão	97,024%	0,037	0,066
	Expansão e Equalização	97,006%	0,056	0,069
	Equalização	96,951%	0,056	0,070
Ranking	Expansão	96,471%	0,048	0,078
	Expansão e Equalização	91,617%	0,052	0,083
	Equalização	92,899%	0,060	0,075

Duas importantes conclusões podem ser extraídas do resultado do experimento comparativo entre operações pontuais para melhoria de brilho e contraste. A aplicação da expansão linear de histograma aumentou a taxa de acerto e reduziu o erro médio e o desvio padrão

em todos os cenários testados. A utilização da equalização de histograma, de forma isolada ou associada à expansão linear de histograma, não é capaz de superar o desempenho da aplicação da expansão linear de histograma de forma isolada, provocando uma perceptível degradação nas estatísticas de erro médio e desvio padrão para as técnicas de combinação adotadas nesta dissertação. Este efeito está relacionado com a destruição de informação gerada pela equalização e citada por Bovik [Bov05].

Alguns resultados produzidos no comparativo entre operações pontuais para melhoria de brilho e contraste são ilustrados nas imagens presentes nas Figuras de 5.8(a) a 5.8(l). Nas Figuras de 5.8(a) a 5.8(d) são apresentados os resultados obtidos com a utilização da equalização de histograma, nas Figuras de 5.8(e) a 5.8(h) são ilustrados os resultados produzidos com a combinação de operações pontuais (expansão e equalização de histograma) e nas Figuras de 5.8(i) a 5.8(l) são mostrados os resultados gerados com aplicação da expansão linear de histograma. A comparação entre as detecções estão de acordo com o comportamento descrito na Tabela 5.3: a taxa de acerto obtida com expansão linear de histograma é superior em relação a equalização e a combinação das duas. Mais exemplos de imagens produzidas neste comparativo são encontradas no Apêndice B.

5.5 Comparativo entre a abordagem proposta e sistemas existentes

O objetivo inicial pretendido com a realização de experimentos era comparar a implementação da abordagem proposta com todos os trabalhos de detecção de olhos baseados em abordagem passiva apresentados na revisão bibliográfica do Capítulo 2. Este objetivo não pôde ser alcançado devido a diversos fatores que impediram o acesso às bases de imagens de faces utilizadas nos demais trabalhos, conforme descrito na Seção 5.1, e as taxas de acerto não divulgadas em alguns trabalhos.

O experimento comparativo entre a abordagem proposta e os sistemas existentes foi realizado apenas com os trabalhos que utilizam bases de imagens de faces disponíveis publicamente e cujo desempenho foi quantitativamente descrito. Apesar de nenhum dos trabalhos revisados publicarem o critério de tolerância a erro adotado na definição da taxa de acerto, dois trabalhos apresentaram dados suficientes para a realização de comparação: Han et al

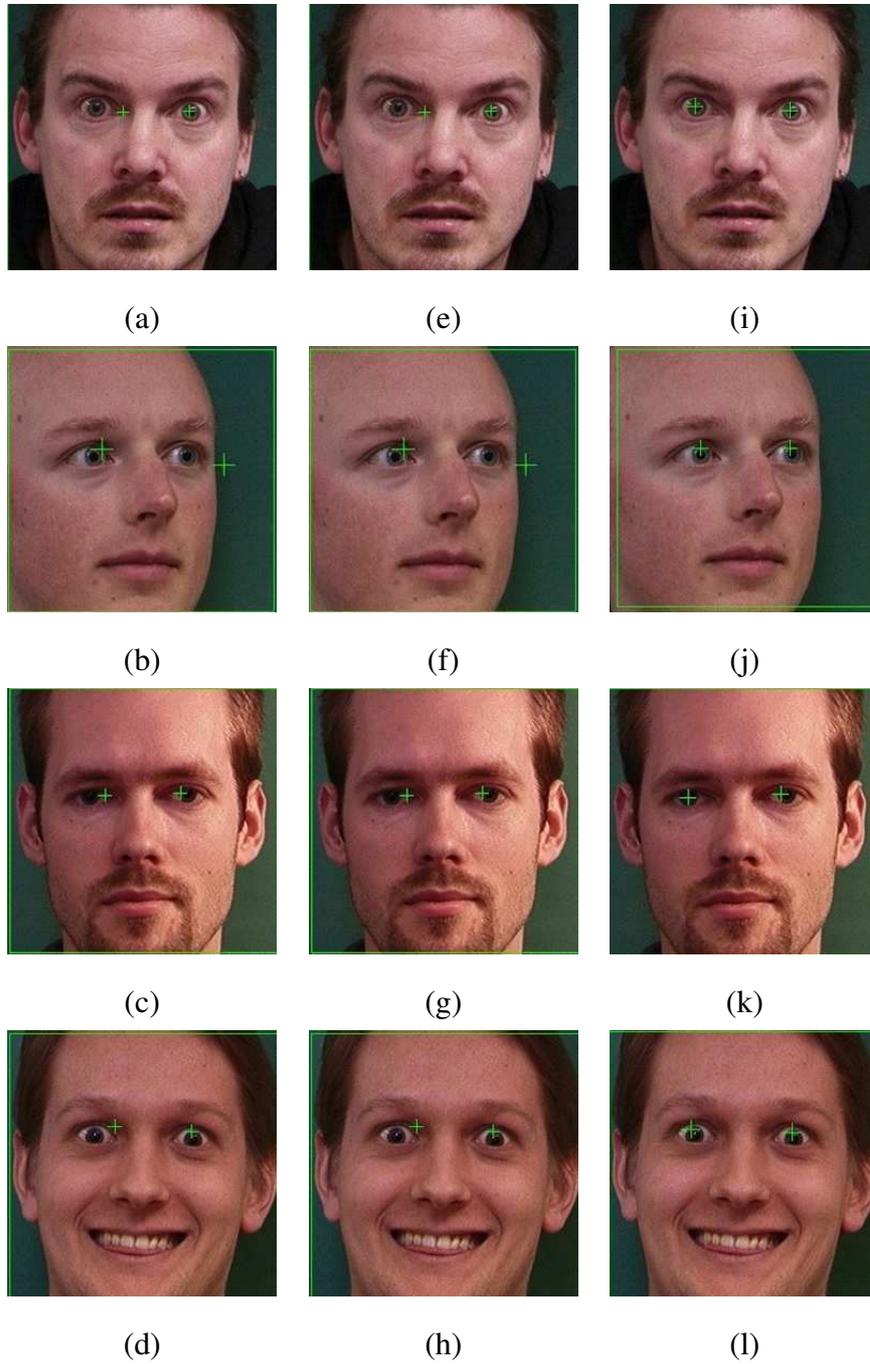


Figura 5.8: Comparação de desempenho relativa à melhoria de brilho e contraste.

[HKN02], que apresentou uma taxa de correta detecção de 98,8% utilizando a base de imagens Caltech [Web], e Wang e Yin [WY05], que indica um desempenho de 95,8% considerando a base de faces JAFFE [LAKG98]. O experimento ocorreu com a utilização das duas bases de imagens de faces citadas e considerando as três regras de combinação disponíveis, com aplicação de compensação de iluminação e expansão de contraste. As marcações manuais dos olhos em cada imagem das duas bases utilizadas foram gentilmente cedidas por Maia [MGS07]. Como a base JAFFE contém imagens monocromáticas e o módulo de geração de mapa de evidência baseado em aparência necessita de informação de cor para produzir resultado, o mesmo foi desabilitado nos testes que utilizam esta base.

O resultado deste experimento se encontra na Tabela 5.4. A organização das colunas segue a seguinte ordem: base de face utilizada, regra de combinação considerada, taxa de acerto, erro médio e desvio padrão normalizado.

Tabela 5.4: Tabela comparativa de desempenho entre a abordagem proposta e as existentes.

Base	Regra	Acerto	Erro médio (\bar{d})	Desvio padrão (σ)
Caltech	Produto	99,361%	0,041	0,076
	Média	97,833%	0,031	0,066
	<i>Ranking</i>	97,468%	0,054	0,058
	Han et al [HKN02]	98,800%	-	-
JAFFE	Produto	98,122%	0,064	0,043
	Média	98,020%	0,064	0,044
	<i>Ranking</i>	96,061%	0,084	0,050
	Wang e Yin [WY05]	95,800%	-	-

Considerando a base Caltech, os resultados apresentados pela implementação da abordagem proposta com as Regras da Média e do *Ranking* foram bastante promissores devido as altas taxas obtidas, porém estas taxas de acerto ainda se encontram abaixo do valor obtido por Han et al [HKN02]. No entanto, o resultado alcançado pela Regra do Produto (99,361%) superou o desempenho de Han et al [HKN02] (98,8%).

Observando os resultados produzidos com a base JAFFE, percebe-se que o sistema proposto obteve resultado entre o valor mínimo de 96,061% e o valor máximo de 98,122%

superando (independentemente da regra de combinação considerada) o resultado publicado por Wang e Yin [WY05] com 95,8% de acerto.

Alguns exemplos de detecção de olhos em imagens das bases Caltech e JAFFE, respectivamente, são ilustrados nas Figuras 5.9 e 5.10. Estas imagens foram produzidas no experimento comparativo entre a abordagem proposta e sistemas existentes. Uma maior quantidade de exemplos de detecção de olhos geradas neste experimento, agrupados pela estratégia de combinação de mapas de evidência utilizada, são apresentados no Apêndice C.



Figura 5.9: Detecções de olhos na base Caltech produzidas pela abordagem proposta.

5.6 Comentários finais

Para avaliar a eficiência do sistema foi realizado um experimento considerando 10 imagens de cada base de face utilizada nos experimentos descritos nas seções anteriores (a saber: IMM, Caltech e JAFFE). Neste experimento, foi observado o comportamento da implementação do sistema proposto em relação à utilização da memória e o tempo de processamento. Os dados produzidos são apresentados na Tabela 5.5. Na primeira coluna, indica-se a base de faces utilizada; na segunda, as dimensões das imagens contidas em cada base; na terceira coluna, o valor máximo de memória requerida; a quarta coluna, o tempo gasto com a etapa de detecção de olhos (que compreende os procedimentos realizados do módulo de compensação de iluminação até a combinação de mapas) e na quinta coluna, o tempo total gasto

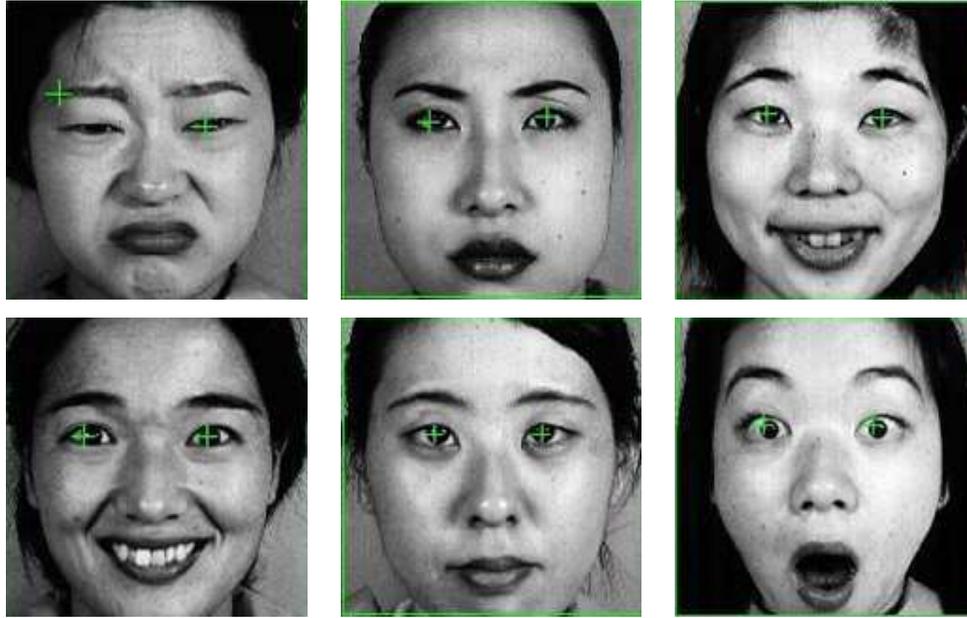


Figura 5.10: Detecções de olhos na base JAFFE produzidas pela abordagem proposta.

no processamento (incluindo as etapas de leitura e escrita de arquivo e detecção de faces). Este valores foram obtidos utilizando um computador com processador AMD Athlon XP 2400+ com frequência de operação de 2.0GHz e 512MB de memória RAM com sistema operacional Windows XP. Em todos os cenários do experimento, foi utilizada a compensação de iluminação, a melhoria de brilho e contraste através da expansão linear de histograma e a combinação dos mapas de evidência através da regra do produto. Esta configuração foi escolhida por apresentar a maior taxa de acerto nos experimentos realizados da Seção 5.3 a 5.5.

Tabela 5.5: Tempos de execução e memória requerida para a detecção de olhos em 10 imagens utilizando a abordagem proposta.

Base	Dimensões (px)	Memória (MB)	Detecção de olhos (s)	Tempo total (s)
IMM	640 × 480	15	54	64
Caltech	896 × 592	26	100	130
JAFFE	256 × 256	9	32	37

De acordo com os dados apresentados no Tabela 5.5, a memória requerida e o tempo de processamento são proporcionais à dimensão da imagem que será processada. Outra

informação fornecida por esta tabela, é a relação entre o tempo de execução da detecção de olhos que corresponde, em média, a cerca de 80% do tempo de processamento total.

É importante ressaltar que a implementação da abordagem proposta não considerou a eficiência, em relação a utilização de memória e tempo de processamento, como requisito necessário.

Os principais fatores (definidos através de inspeção visual das detecções produzidas nos experimentos com as bases IMM, Caltech e JAFFE) que influenciaram os erros nas detecções de olhos com a abordagem proposta nesta dissertação foram: sombra do nariz sobre o olho (Figura 5.11(a)), imagens com pouca iluminação (Figura 5.11(b)), uso de óculos (Figura 5.11(c)), oclusão parcial dos olhos pelos cabelos (Figura 5.11(d)), sobrancelhas/cílios muito escuros (Figura 5.11(e)) e rotação do plano da face (Figura 5.11(f)).

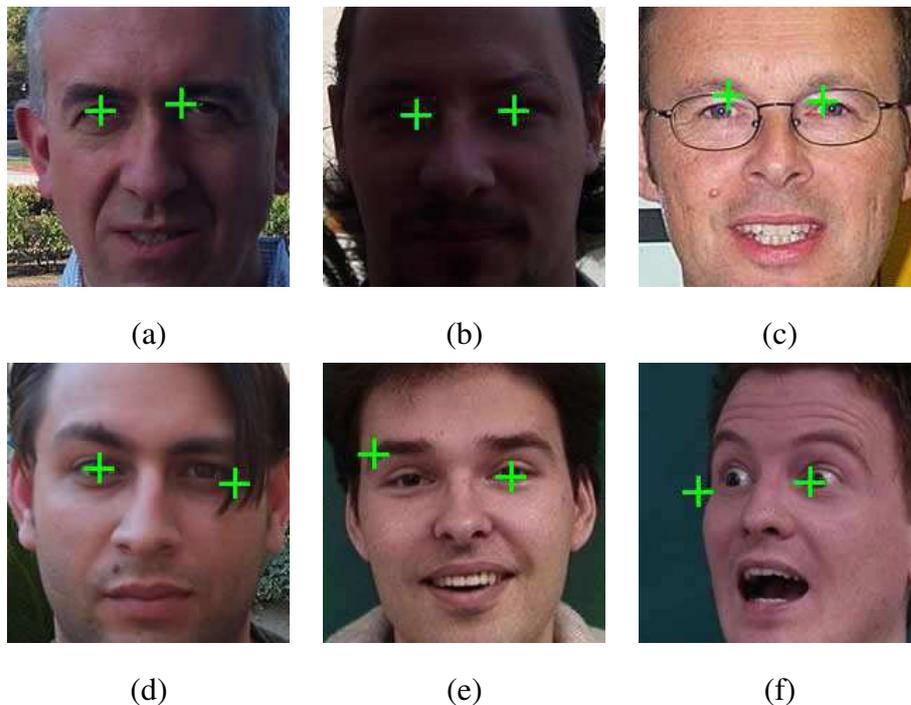


Figura 5.11: Exemplos de erros nas detecções de olhos.

Apesar das dificuldades para encontrar detectores de olhos disponíveis publicamente, foram encontradas duas implementações, Rowley et al [RBK99] e Fasel et al [FFM05], as quais foram comparadas com uma versão preliminar da abordagem proposta (limitada a utilização do módulo de geração de mapa baseado em técnicas de aprendizagem). Os resultados desta comparação se encontra no artigo *A Learning-based Eye Detector Coupled with Eye Candidate Filtering and PCA Features* publicado no *XX Brazillian Symposium on*

Computer Graphics and Image Processing - SIBGRAPI 2007 [LPG⁺07]. Nestes resultados, considerando uma tolerância de 9 *pixels* (quando este artigo foi publicado, o processo de normalização descrito na Seção 5.2 ainda não havia sido adotado), a taxa de acerto da versão preliminar da abordagem proposta e de Rowley et al [RBK99] estão acima dos 90%, enquanto o desempenho do detector de Fasel et al [FFM05] se encontra abaixo de 70%. Um trabalho futuro consiste em refazer este experimento comparativo, agora considerando a versão completa da abordagem proposta (com a combinação de mapas de evidência baseados na análise da aparência, em aprendizagem e em modelos).

5.7 Sumário

Os experimentos realizados com a implementação da abordagem proposta (Capítulos 3 e 4) são apresentados neste capítulo. Inicialmente, são descritas as características das bases de imagens de faces utilizadas nos experimentos (IMM, Caltech e JAFFE) e as dificuldades de acesso a bases empregadas em artigos que constam na revisão bibliográfica (Capítulo 2). Em seguida, é definida a metodologia de cálculo e a tolerância a erros considerada ao longo dos experimentos. O primeiro experimento comprova os ganhos de desempenho obtidos com o uso da compensação de iluminação lateral. Os resultados obtidos no segundo experimento demonstram que o emprego da expansão linear de histograma provoca melhoria de desempenho, enquanto a equalização de histograma provoca o efeito inverso. Finalmente, a implementação do sistema proposto é confrontada com o desempenho de trabalhos relacionados, apresentando taxa de correta detecção de olhos superiores, na maioria dos casos, ou ligeiramente inferiores, mas ainda aceitáveis, quando comparados com os resultados apresentados pelos trabalhos presentes revisão bibliográfica (Tabela 2.4).

Capítulo 6

Considerações finais

Neste capítulo, apresentam-se um sumário dos principais pontos abordados nesta dissertação, bem como as contribuições da pesquisa desenvolvida e propostas de trabalhos futuros, a partir dos problemas encontrados durante o desenrolar da pesquisa, os quais necessitam de estudos mais aprofundados.

6.1 Sumário da dissertação

A motivação para o desenvolvimento desta dissertação surgiu a partir das necessidades de algumas aplicações que utilizam a detecção de olhos em faces humanas como ponto de partida, as quais se destacam: a interação homem-máquina, a geração sintética de modelos em três dimensões (3D) para interação em ambientes virtuais, o monitoramento de atenção de condutores de veículos, a identificação de indivíduos através da face ou íris, a análise de expressões faciais e a definição de regiões de interesse na produção de vídeos em sistemas de vídeo conferência.

A abordagem proposta para a detecção de olhos em faces humanas é constituída de módulos para: a detecção de faces, utilizando a biblioteca OpenCV [Ope]; a compensação de iluminação lateral, através de filtragem homomórfica; a melhoria de brilho e contraste, com o uso de operações pontuais de equalização e expansão linear de histograma; a limitação do espaço de busca na área da face; a geração de mapas de evidência, baseados na análise da aparência, em técnicas de aprendizagem e em casamento de modelos e, finalmente, a combinação dos mapas de evidência produzidos, utilizando técnicas de combinação de

classificadores.

O sistema baseado na abordagem proposta passou por experimentos para avaliação de sua viabilidade e eficácia, além disto, o resultado do sistema proposto foi comparado com o desempenho de outros alguns sistemas de detecção de olhos existentes, publicados em artigos presentes na revisão bibliográfica. A conclusão obtida com a realização destes experimentos é discutida na próxima seção.

6.2 Considerações e contribuições do trabalho realizado

A partir dos resultados obtidos nos experimentos realizados no Capítulo 5, é possível concluir que objetivo de desenvolver um sistema para detecção de olhos em imagens com faces humanas proposto nesta dissertação foi alcançado. A implementação do sistema proposto atingiu o critério de sucesso esperado, superando o desempenho (em termos de taxa de acerto) das abordagens existentes. Os experimentos realizados comprovaram os benefícios do uso da compensação de iluminação lateral, da melhoria de brilho e contraste através da operação pontual de expansão linear de histograma e do uso da combinação de classificadores aplicados ao problema da detecção de olhos em faces humanas.

A enumeração seguinte apresenta as principais contribuições produzidas por este trabalho.

1. A constatação experimental dos benefícios obtidos com a utilização de técnicas de compensação de iluminação lateral na detecção de olhos
2. Uma comparação baseada em resultados experimentais do efeito da utilização de operações pontuais de expansão e de equalização de histogramas, para melhoria de brilho e contraste na resolução do problema de detecção de olhos em faces humanas;
3. A definição de uma metodologia para geração de espaço de busca para olhos em faces humanas, para um determinado detector de faces;
4. A proposição inédita de um sistema híbrido composto de classificadores baseados na análise de aparência, em técnicas de aprendizagem e em modelos para detecção de olhos com a utilização de diversas técnicas de combinação de resultados.

6.3 Trabalhos futuros

A seguir, são apresentadas sugestões de trabalhos futuros para complementação dos resultados produzidos nesta dissertação.

Durante a realização dos experimentos com uso da compensação de iluminação (Seção 5.3), foi perceptível a diferença de desempenho no aspecto de velocidade de execução nos diversos cenários testados. Como a compensação de iluminação foi feita através da filtragem homomórfica e esta agrega uma complexidade computacional consideravelmente alta, o processo de detecção de olhos com o uso da compensação de iluminação aumentou em cerca de 25% o tempo de processamento total. Esta constatação gera uma demanda por uma investigação a respeito da detecção de condições que indique a presença de iluminação lateral para permitir a aplicação de compensação apenas se necessário.

Um possível ponto de investigação para melhoria do desempenho da abordagem proposta consiste na realização de experimentos semelhantes ao descrito na Seção 5.4. Desta vez, comparando as operações de expansão e equalização de histograma, descritas nas Seções 3.3.1 e 3.3.2, com outros procedimentos para melhoria de brilho e contraste como expansão não-linear de histograma e processamento Retinex [FCM00].

Conforme discutido na Seção 3.8, com relação à combinação de classificadores, existem trabalhos que utilizam esquemas de combinação com uma estratégia de ponderação para cada classificador utilizado, produzindo combinações com pesos diferenciados que podem amplificar ou atenuar as respostas produzidas por um determinado classificador para determinadas situações. Uma lacuna a ser investigada é a otimização da combinação de pesos através do uso de algoritmos genéticos nos moldes da abordagem proposta por Pereira [Per07], para o problema de detecção de olhos.

Na etapa de geração de mapas de evidência, podem ser conduzidos testes para determinação de regiões de pele em espaços de cores além do RGB, estudos para adição de outras abordagens como a geração de mapas de simetria (descritos no trabalho de Reisfeld [RWY95]) para detecção de olhos, utilização de outros extratores de características e máquinas de aprendizagem, como os descritos no livro de Haykin [Hay98], aumento da base de imagens (utilizada no treinamento do classificador baseado em aprendizagem) através da extração de exemplos de olhos de bases de imagens de face disponíveis publicamente, apli-

cação de abordagem alternativa para situações onde existe a presença de óculos e utilização de modelos variáveis na avaliação de função de energia da abordagem baseada em modelos.

Uma proposta a ser conduzida é a elaboração de um experimento comparativo de desempenho, seguindo a metodologia aplicada por Leite et al [LPG⁺07]. Desta vez, considerando a versão completa da abordagem proposta e os detectores de olhos de Rowley et al [RBK99] e de Fasel [FFM05] (Seção 5.5).

Bibliografia

- [AWT06] T. Akashi, Y. Wakasa, and K. Tanaka. Performance of eye detection by genetic algorithms. In *Proceedings of International Conference of Computational Intelligence*, pages 94–99, 2006.
- [AZVK05] A. Amir, L. Zimet, A. S. Vincentelli, and S. Kao. An embedded system for an eye detection sensor. *Computer Vision and Image Understanding - Special Issue on Eye Detection and Tracking*, 98:104–123, 2005.
- [Bis95] C. Bishop. *Neural Networks for Pattern Recognition*. Oxford, 1995.
- [BK00] J. Bilmes and K. Kirchhoff. Directed graphical models of classifier combination: Application to phone recognition. In *Proceedings of International Conference on Spoken Language Processing*, pages 1–1, 2000.
- [BKP05] G. R. Bradski, A. Kaehler, and V. Pisarevsky. Learning-based computer vision with intel’s open source computer vision library. *Intel Technology Journal - Compute-Intensive, Highly Parallel Applications and Uses*, 9, 2005.
- [BKR03] T. N. Bhaskar, F. T. Keat, S. Ranganath, and Y. V. Venkatesh. Blink detection and eye tracking for eye localization. In *Proceedings of Conference on Convergent Technologies for Asia-Pacific Region - TENCON*, pages 821–824, 2003.
- [Bov05] A. Bovik. *Handbook of image and video processing*. Elsevier, 2005.
- [CG05] C. S. C. Cunha and H. M. Gomes. People detection in still images based on a skin filter and body part evidence. In *Proceedings of Brazilian Symposium on Computer Graphics and Image Processing*, pages 1–2, 2005.

- [CG07] F. H. Cardoso and H. M. Gomes. A probabilistic approach to skin detection. In *Proceedings of Brazillian Symposium on Computer Graphics and Image Processing - SIBGRAPI*, 2007.
- [CSH⁺00] P. Corr, D. Stewart, P. Hanna, J. Ming, and F.J. Smith. Discrete chebyshev transform - a natural modification of the dct. In *Proceedings of International Conference on Pattern Recognition - ICPR*, pages 1142–1145, 2000.
- [CTY⁺03] C. C. Chiang, W. Tai, M. Yang, Y. Huang, and C. Huang. A novel method for detecting lips, eyes and faces in real time. *Real-Time Imaging*, 9:277–287, 2003.
- [DGK06] K. Delac, M. Grgic, and T. Kos. Sub-image homomorphic filtering technique for improving facial identification under difficult illumination conditions. In *Proceedings of International Conference on Systems, Signals and Image Processing*, pages 95–98, 2006.
- [DLSG04] T. D’Orazio, M. Leo, P. Spagnolo, and C. Guaragnella. A neural system for eye detection in a driver vigilance application. In *Proceedings of International IEEE Conference on Intelligent Transportation Systems*, pages 320–325, 2004.
- [FCM00] B. Funt, F. Ciurea, and J. McCann. Retinex in matlab. In *Proceedings of the IS&T/SID Color Imaging Conference: Color Science, Systems and Applications*, pages 112–121, 2000.
- [FER] FERET. The Facial Recognition Technology Database. http://www.itl.nist.gov/iad/humanid/feret/feret_master.html. Acessado em 29/06/2008.
- [FFM05] I. Fasel, B. Fortenberry, and J. R. Movellan. A generative framework for real-time object detection and classification. *Computer Vision and Image Understanding*, 2005.
- [FM04] A. Fathi and M. T. Manzuri. Eye detection and tracking in video streams. In *Proceedings of IEEE International Symposium on Communications and Information Technology*, volume 2, pages 1258–1261, 2004.

- [FRG] FRGC. Face Recognition Grand Challenge. <http://www.frvt.org/FRGC/>. Acessado em 29/06/2008.
- [GEZMT99] T. Goto, M. Escher, C. Zanardi, and N. Magnenat-Thalmann. MPEG-4 based animation with face feature tracking. In *Proceedings of Computer Animation and Simulation*, pages 89–98, 1999.
- [Gro05] R. Gross. *Handbook of Face Recognition*. Springer-Verlag, 2005.
- [GW92] R. C. Gonzalez and R. E. Woods. *Digital Image Processing*. Addison-Wesley, 1992.
- [Hay98] S. Haykin. *Neural Networks: A Comprehensive Foundationm, 2nd Edition*. Prentice Hall, 1998.
- [HHI] HHI. MPEG7 content set from Heinrich Hertz Institute. <http://www.darmstadt.gmd.de/mobile/hm/projects/MPEG7/Documents/N2466.html>. Acessado em 29/06/2008.
- [HKN02] H. Han, T. Kawaguchi, and R. Nagata. Eye detection based on grayscale morphology. In *Proceedings of IEEE Region 10 Conference on Computers, Communications, Control and Power Engineering*, volume 1, pages 498–502, 2002.
- [HSLW98] W. Huang, Q. Sun, C. Lam, and J. Wu. A robust approach to face and eyes detection from images with cluttered background. In *Proceedings of International Conference on Pattern Recognition*, volume 1, pages 110–113, 1998.
- [ISB03] Y. Ivanov, T. Serre, and J. Bouvrie. Error weighted classifier combination for multi-modal human identification. Technical report, Computer Science and Artificial Intelligence Laboratory - Massachusetts Institute of Technology, 2003.
- [JKF01] O. Jesorsky, K. J. Kirchberg, and R. W. Frischholz. Robust face detection using the Hausdorff distance. *Lectures Notes in Computer Science*, 2091:212–227, 2001.

- [JR02] M. J. Jones and J. M. Rehg. Statistical color models with application to skin detection. *International Journal of Computer Vision*, 46(1):81–96, 2002.
- [JWDF05] Q. Ji, H. Wechsler, A. Duchowski, and M. Flickner. Editorial. *Computer Vision and Image Understanding - Special Issue on Eye Detection and Tracking*, 98:1–3, 2005.
- [JY01] Q. Ji and X. Yang. Real time visual cues extraction for monitoring driver vigilance. *Lecture Notes In Computer Science*, 2095:107–124, 2001.
- [JYS+06] L. Jin, X. Yuan, S. Satoh, J. Li, and L. Xia. A hybrid classifier for precise and robust eye detection. In *Proceedings of International Conference on Pattern Recognition*, volume 4, pages 731–735, 2006.
- [KMB07] P. Kakumanu, S. Makrogiannis, and N. Bourbakis. A survey of skin-color modeling and detection methods. *Pattern Recognition*, 40:1106–1122, 2007.
- [KRR02] R. T. Kumar, S. K. Raja, and A. G. Ramakrishnan. Eye detection using color cues and projection functions. In *Proceedings of International Conference on Image Processing*, volume 3, pages 337–340, 2002.
- [KT04] S. Kawato and N. Tetsutani. Detection and tracking of eyes for gaze-camera control. *Image and Vision Computing*, 22:1031–1038, 2004.
- [LAKG98] M. Lyons, S. Akamatsu, M. Kamachi, and J. Gyoba. Coding facial expressions with gabor wavelets. In *Proceedings of International Conference on Automatic Face and Gesture Recognition*, pages 200–205, 1998. Acessado em 29/06/2008.
- [LM02] R. Lienhart and J. Maydt. An extended set of haar-like features for rapid object detection. In *Proceedings of International Conference on Image Processing*, pages 900–903, 2002.
- [LPG+07] B. B. Leite, E. T. Pereira, H. M. Gomes, L. R. Veloso, C. E. N. Santos, and J. M. Carvalho. A learning-based eye detector coupled with eye candidate filtering and pca features. In *Proceedings of Brazillian Symposium on Computer Graphics and Image Processing - SIBGRAPI*, pages 187–194, 2007.

- [LY04] D. Lin and C. Yang. Real-time eye detection using face-circle fitting and dark-pixel filtering. In *Proceedings of IEEE International Conference on Multimedia and Expo*, volume 2, pages 1167–1170, 2004.
- [MBZ02] T. Morris, P. Blenkhorn, and F. Zaidi. Blink detection for real-time eye tracking. *Journal of Network and Computer Applications*, 25:129–143, 2002.
- [MDWW01] Y. Ma, X. Ding, Z. Wang, and N. Wang. Robust precise eye location under probabilistic framework. In *Proceedings of IEEE International Conference Automatic Face and Gesture Recognition*, pages 339–344, 2001.
- [MGS07] J. G. R. Maia, F. C. Gomes, and O. Souza. Automatic eye localization in color images. In *Proceedings of Brazillian Symposium on Computer Graphics and Image Processing - SIBGRAPI*, pages 195–202, 2007.
- [Mic] Microsoft. MSDN Academic Alliance. [http://msdn2.microsoft.com/pt-br/academic/default\(en-us\).aspx](http://msdn2.microsoft.com/pt-br/academic/default(en-us).aspx). Acessado em 29/06/2008.
- [Nis05] S. Nissen. Neural networks made simple. *Software 2.0*, 2:14–19, 2005.
- [Ope] OpenCV. Open Computer Vision Library. <http://www.intel.com/research/-mrl/research/opencv/>. Acessado em 29/06/2008.
- [Par97] J. R. Parker. *Algorithms for Image Processing and Computer Vision*. John Wiley & Sons, 1997.
- [PCRK05] K. Peng, L. Chen, S. Ruan, and G. Kukharev. A robust algorithm for eye detection on gray intensity face without spectacles. *JCS & T*, 5:127–132, 2005.
- [Per07] E. T. Pereira. Atenção visual bottom-up guiada por otimização via algoritmos genéticos, 2007. Dissertação de mestrado.
- [PZB98] H. Peng, C. Zhang, and Z. Bian. Human eyes detection using hybrid neural method. In *Proceedings of International Conference on Signal Processing Proceedings*, volume 2, pages 1088–1091, 1998.

- [RBK99] H. A. Rowley, S. Baluja, and T. Kanade. *Neural Network-Based Face Detection*. PhD thesis, Carnegie Mellon University, 1999.
- [RWY95] D. R., H. Wolfson, and Y. Yeshurun. Context-free attentional operators: The generalized symmetry transform. *International Journal of Computer Vision*, 14(2):119–130, 1995.
- [Sca] Human Scan. BioID Face Database. <http://www.bioid.com>. Accessed on 29/06/2008.
- [SEL03] M. B. Stegmann, B. K. Ersbøll, and R. Larsen. FAME – a flexible appearance modelling environment. *IEEE Trans. on Medical Imaging*, 22(10):1319–1331, 2003.
- [SHT01] S. A. Samad, A. Hussain, and A. Teoh. Eye detection using hybrid rule based approach and contour mapping. In *Proceedings of International Symposium on Signal Processing and its Applications*, volume 2, pages 631–634, 2001.
- [SJST07] S. J. Schmugge, S. Jayaram, M. C. Shin, and L. V. Tsap. Objective evaluation of approaches of skin detection using roc analysis. *Computer Vision and Image Understanding*, 108:41–51, 2007.
- [Smi97] S. W. Smith. *The Scientist and Engineer’s Guide to Digital Signal Processing*. California Technical Publishing, 1997.
- [Smi02] L. I. Smith. A tutorial on principal components analysis. Technical report, Computer Science Department - University of Otago, 2002.
- [TB05] F. H. C. Tivive and A. Bouzerdoum. A fast neural-based eye detection system. In *Proceedings of International Symposium on Intelligent Signal Processing and Communication Systems*, pages 641–644, 2005.
- [VJ01] P. Viola and M. J. Jones. Rapid object detection using a boosted cascade of simple features. In *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 511–518, 2001.

- [VJ04] P. Viola and M. J. Jones. Robust real-time face detection. *International Journal of Computer Vision*, 57:137–154, 2004.
- [Web] M. Weber. Caltech Frontal Image Database. <http://www.vision.caltech.edu/html-files/archive.html>. Acessado em 29/06/2008.
- [WGJW05] P. Wang, M. B. Green, Q. Ji, and J. Wayman. Automatic eye detection and its validation. In *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, volume 3, pages 164–164, 2005.
- [WT05] J. Wu and M. M. Trivedi. A binary tree for probability learning in eye detection. In *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, volume 3, pages 170–170, 2005.
- [WY05] J. Wang and L. Yin. Eye detection under unconstrained background by the terrain feature. In *Proceedings of IEEE International Conference on Multimedia and Expo*, pages 1528–1531, 2005.
- [XH06] Z. Xingming and Z. Huangyuan. An illumination independent eye detection algorithm. In *Proceedings of International Conference on Pattern Recognition*, volume 1, pages 392–395, 2006.
- [YC05] D. H. Yoo and M. J. Chung. A novel non-intrusive eye gaze estimation using cross-ratio under large head motion. *Computer Vision and Image Understanding - Special Issue on Eye Detection and Tracking*, 98:25–51, 2005.
- [Yip05] B. Yip. Face and eye rectification in video conference using artificial neural network. In *Proceedings of IEEE International Conference on Multimedia and Expo*, pages 690–693, 2005.
- [ZHG04] Z. Zhi-Hua and X. Geng. Projection functions for eye detection. *Pattern recognition*, 37(5):1049–1056, 2004.
- [ZL00] L. Zhang and P. Lenders. Knowledge-based eye detection for human face recognition. In *Proceedings of International Conference on Knowledge-Based*

Intelligent Engineering Systems and Allied Technologies, volume 1, pages 117–120, 2000.

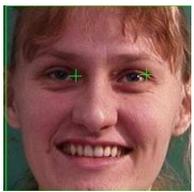
[ZMV+] A. Zell, G. Mamier, M. Vogt, N. Mache, R. Hubner, K. Herrmann, T. Soyez and M. Schmalzl, T. Sommer, A. Hatzigeogiou, S. Doering, and D. Posselt. Stuttgart Neural Network Simulator.

Apêndice A

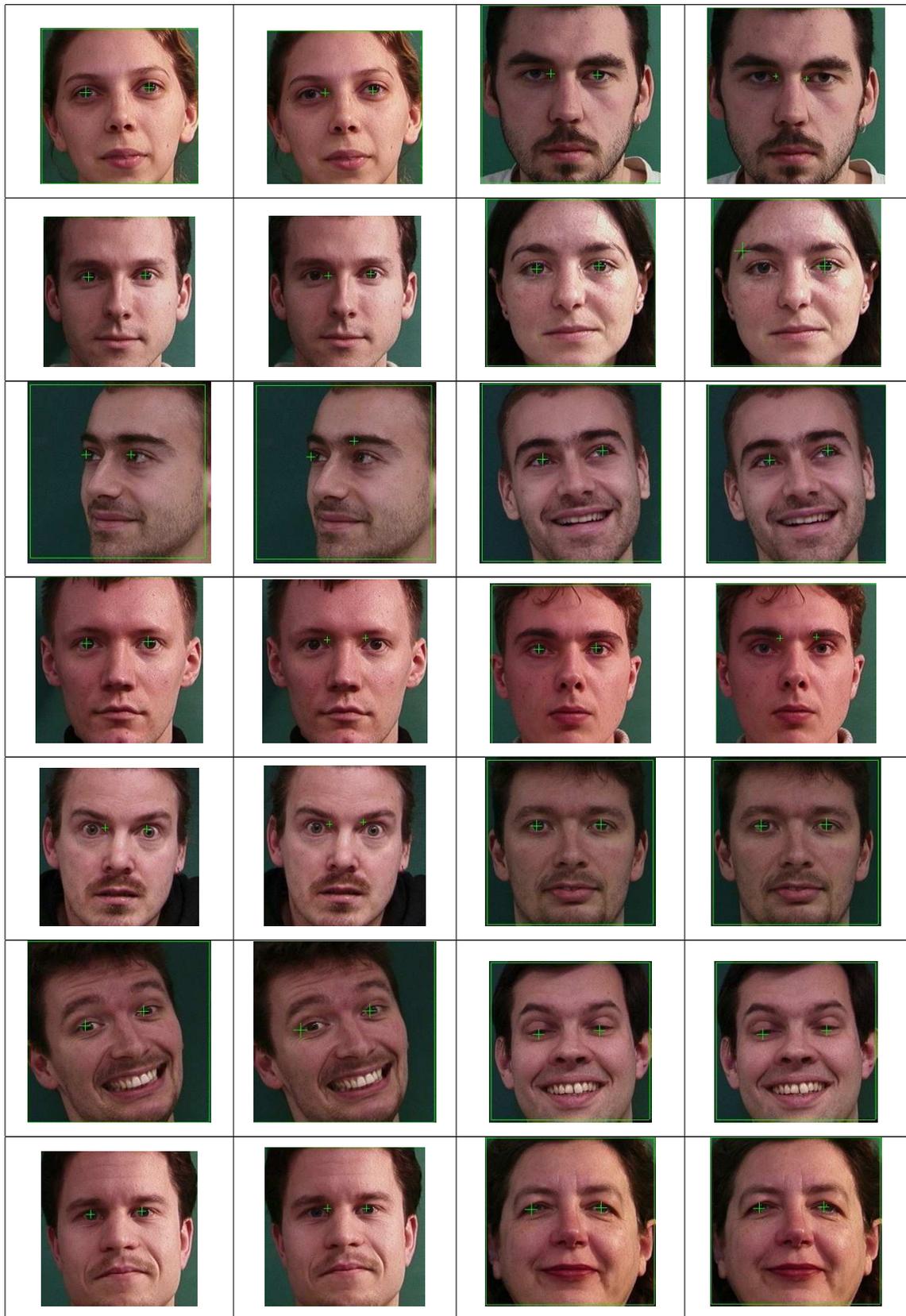
Resultados da avaliação do desempenho com uso da compensação de iluminação

Neste apêndice, apresentam-se exemplos de imagens com detecções de olhos produzidas no experimento de avaliação do desempenho com uso de compensação de iluminação descritos na Seção 5.3. As imagens são apresentadas na Tabela A.1, em cujas primeira e terceira colunas apresentam-se as imagens com as detecções obtidas a partir do uso da compensação da iluminação e em cujas segunda e quarta colunas ilustram-se as detecções, relativas às faces contidas nas imagens da primeira e terceira colunas, sem a compensação da iluminação.

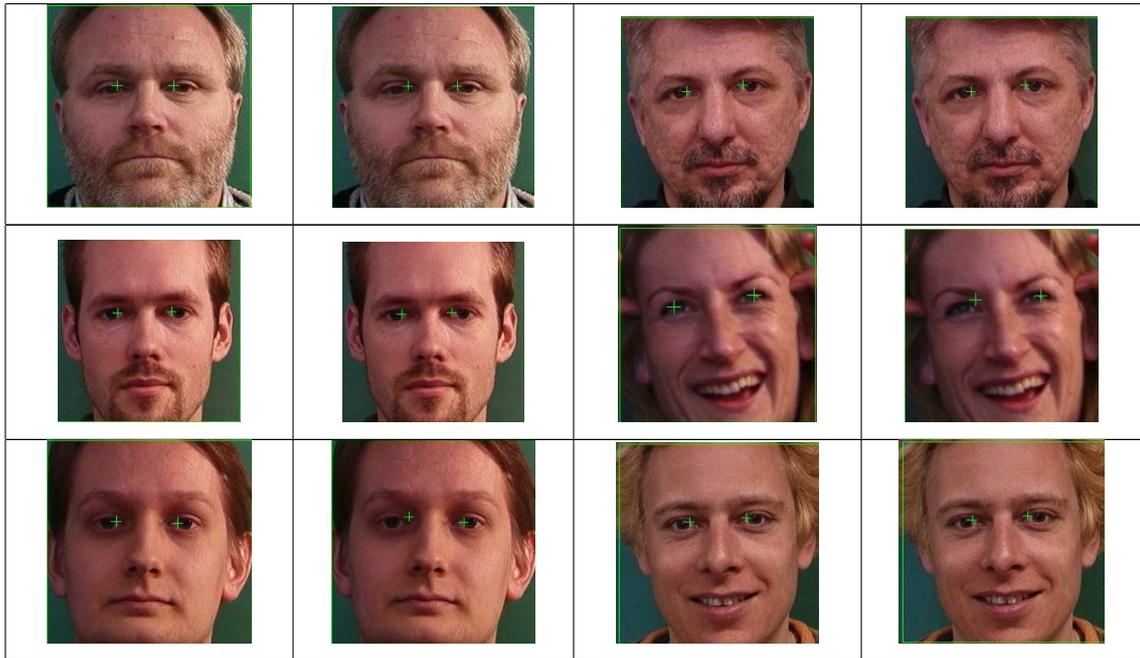
Tabela A.1: Resultados do experimento com uso da compensação de iluminação.

Com compensação	Sem compensação	Com compensação	Sem compensação
			
			

continua na próxima página



continua na próxima página



Apêndice B

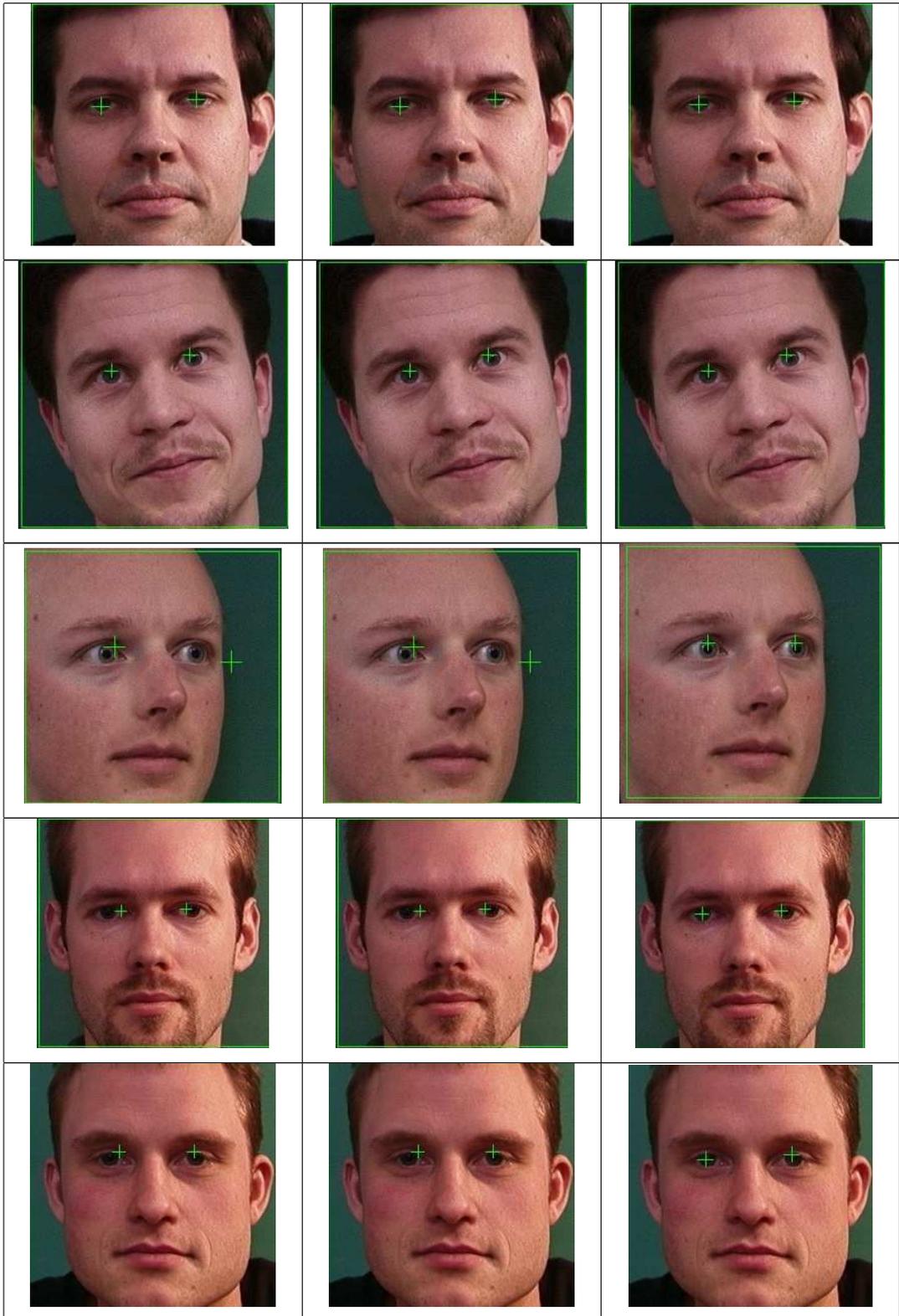
Resultados do comparativo entre operações pontuais para melhoria de brilho e contraste

Alguns exemplos de imagens com detecções de olhos produzidas no comparativo entre operações pontuais para melhoria de brilho e contraste apresentado na Seção 5.4 são ilustrados na Tabela B.1. As imagens com as detecções obtidas com a utilização da equalização de histograma, expansão seguida de equalização de histograma e expansão linear de histograma estão distribuídas na primeira, segunda e terceira coluna, respectivamente.

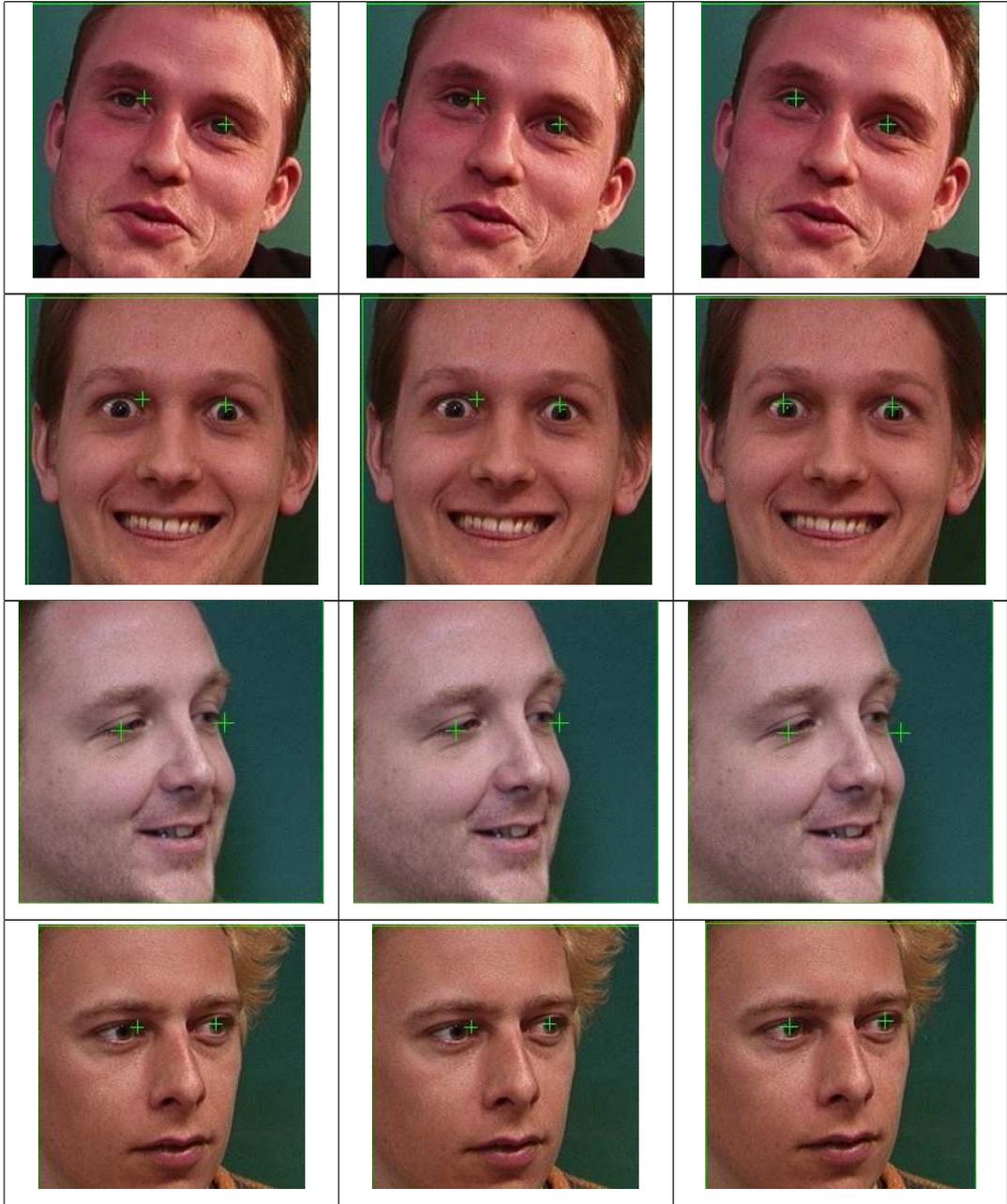
Tabela B.1: Resultados do experimento comparativo para melhoria de brilho e contraste.

Equalização	Expansão e equalização	Expansão
		

continua na próxima página



continua na próxima página



Apêndice C

Resultados apresentados pela abordagem proposta utilizando as bases de faces Caltech e JAFFE

Resultados de detecções de olhos produzidos no experimento comparativo entre a abordagem proposta e sistemas existentes, descrito na Seção 5.5, são apresentados nas Tabelas C.1 e C.2. Em ambas as tabelas, cada coluna corresponde a uma estratégia de combinação de mapa de evidência: média; *ranking* e produto, respectivamente.

Tabela C.1: Exemplos de detecções de olhos na base Caltech.

Média	<i>Ranking</i>	Produto
		
		

continua na próxima página

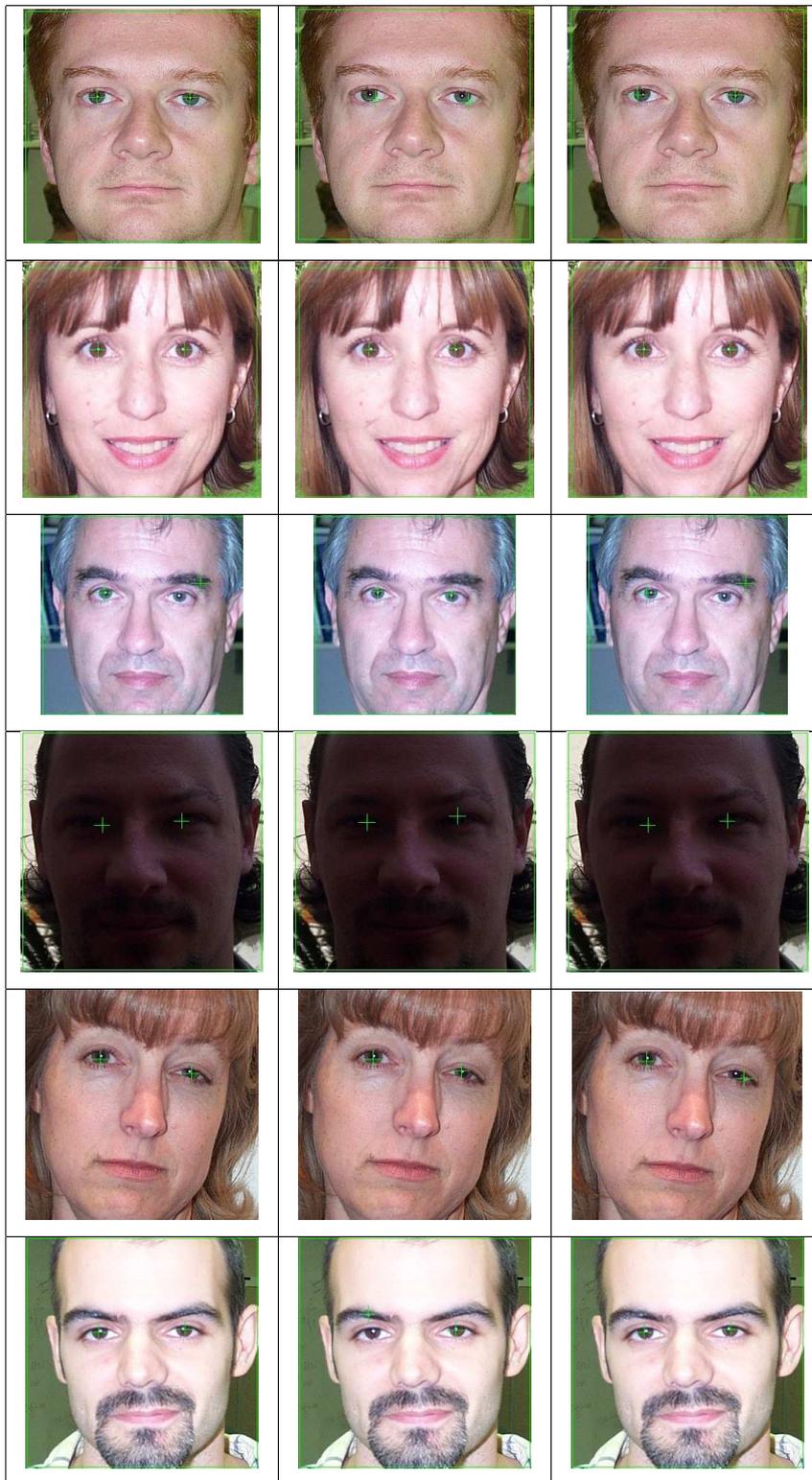
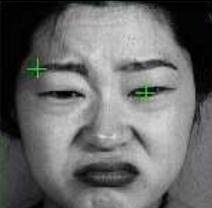
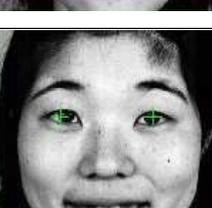
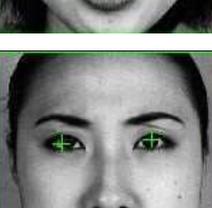
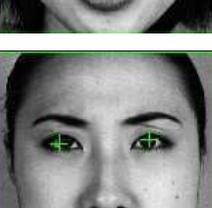
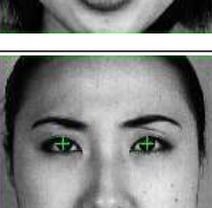
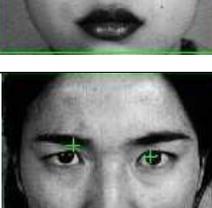
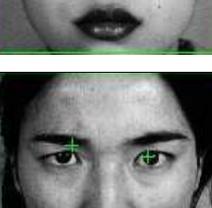
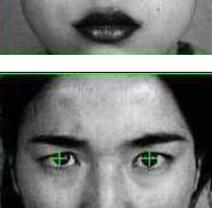


Tabela C.2: Exemplos de detecções de olhos na base JAFFE.

Média	<i>Ranking</i>	Produto
		
		
		
		
		
		

continua na próxima página

