

PONTIFÍCIA UNIVERSIDADE CATÓLICA DO RIO GRANDE DO SUL
FACULDADE DE INFORMÁTICA
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA COMPUTAÇÃO

**DETECÇÃO DE COMPONENTES FACIAIS
BASEADO EM MODELOS DE COR
E ANTROPOMETRIA**

JULIANO LUCAS MOREIRA

Dissertação de Mestrado apresentada como requisito para obtenção do título de Mestre em Ciência da Computação pelo Programa de Pós-graduação da Faculdade de Informática. Área de concentração: Ciência da Computação.

Orientadora: Profa. Dra. Soraia Raupp Musse

Porto Alegre, Brasil
2010

Dados Internacionais de Catalogação na Publicação (CIP)

M838d Moreira, Juliano Lucas
 Detecção de componentes faciais baseado em modelos de cor
 e antropometria / Juliano Lucas Moreira. – Porto Alegre, 2010.
 58 f.

 Diss. (Mestrado) – Fac. de Informática, PUCRS.
 Orientador: Prof^ª. Dr^ª. Soraia Raupp Musse.

 1. Informática. 2. Computação Gráfica. 3. Animação por
 Computador. 4. Bases de Dados. I. Musse, Soraia Raupp.
 II. Título.

CDD 006.6

**Ficha Catalográfica elaborada pelo
Setor de Tratamento da Informação da BC-PUCRS**



Pontifícia Universidade Católica do Rio Grande do Sul
FACULDADE DE INFORMÁTICA
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA COMPUTAÇÃO

TERMO DE APRESENTAÇÃO DE DISSERTAÇÃO DE MESTRADO

Dissertação intitulada "Detecção de Componentes Faciais Baseado em Modelos de Cor e Antropometria", apresentada por Juliano Lucas Moreira, como parte dos requisitos para obtenção do grau de Mestre em Ciência da Computação, Sistemas Interativos e de Visualização, aprovada em 17/03/10 pela Comissão Examinadora:

Prof. Dra. Soraia Raupp Musse -
Orientadora

PPGCC/PUCRS

Prof. Dr. João Batista Souza de Oliveira -

PPGCC/PUCRS

Prof. Dr. Cláudio Rosito Jung -

UFRGS

Homologada em 29/06/10, conforme Ata No. 011, pela Comissão Coordenadora.

Prof. Dr. Fernando Gehm Moraes
Coordenador.

PUCRS

Campus Central

Av. Ipiranga, 6681 - P32 - sala 507 - CEP: 90619-900

Fone: (51) 3320-3611 - Fax (51) 3320-3621

E-mail: ppgcc@pucrs.br

www.pucrs.br/facin/pos

“Não é fácil ter paciência diante dos que têm excesso de paciência.”

Carlos Drummond de Andrade

Agradecimentos

A minha orientadora Soraia, pela dedicação e paciência durante a orientação.

A minha namorada Mônica pela compreensão e troca de experiência e angústias com relação ao mestrado.

A minha mãe Núbia, vó Maria e irmã Mariana, pelo apoio nos momentos difíceis.

Aos grandes amigos que fiz nestes 2 anos de mestrado, junto ao Laboratório de pesquisa VHLAB (Adriana, Henry, Rafael Hocevar, Marcos, Rossana, Rafael Rodrigues, Humberto, Leandro, Júlio, Marcelo Paravisi, Marcelo Cohen, ...).

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

DETECÇÃO DE COMPONENTES FACIAIS BASEADO EM MODELOS DE CORES E ANTROPOMETRIA

RESUMO

Existem diferentes situações em que a detecção de componentes faciais desempenha um papel importante, como por exemplo: no reconhecimento facial, interação humano computador, teleconferências, análise de expressões faciais e animação. Esta dissertação apresenta, primeiramente, uma revisão bibliográfica dos trabalhos referentes à detecção dos componentes faciais. Após, é exposto o modelo desenvolvido para a detecção de componentes faciais, sendo este baseado em modelos de cor e antropometria. Ao final, são mostrados resultados referentes a detecção dos componentes em bases de dados públicas e experimentos realizados na animação de avatares 3D.

Palavras-chave: Detecção Componentes Faciais; Base de Dados IMM; Animação Avatar 3D.

DETECTION OF FACIAL COMPONENTS BASED ON MODELS OF COLOR AND ANTHROPOMETRY

ABSTRACT

There are different situations in which the detection of facial components plays an important role, for example, facial recognition, human computer interaction, teleconferencing, analysis of facial expressions and animation. This master thesis presents a literature review of work on the facial features detection. Then, the model developed for the facial features detection is described, which is based on models of color and anthropometry. At the end we show results of the proposed model in feature detection applied in public databases, as well as experiments in 3D avatars animation.

Keywords: Facial Features Detection; IMM Database, 3D Avatar Animation.

Lista de Figuras

1.1	Exemplo de resultado do trabalho de Dutreuve <i>et al.</i> (14)	16
2.1	Template para olhos proposto por Yuille et al. (51). Duas parábolas descrevem os limites das pálpebras e uma circunferência descreve a íris. O rastreamento é feito através do ajuste dos parâmetros a , b e c das parábolas, P_1 , P_2 do centro da esclera, r e x_c da íris à imagem.	17
2.2	Demonstração das 19 características identificadas pelo método de (43), em duas imagens distintas	18
2.3	Distribuição de pontos baseados no padrão MPEG4, utilizado por Tong <i>et al.</i> (41).	19
2.4	Exemplo da extração de componentes faciais (27).	20
2.5	Exemplos da extração de contornos feita a partir dos pontos característicos localizados (47)	22
2.6	Modelo Facial proposto por Sohail and Bhattacharya (37) para localização das regiões faciais.	23
3.1	Exemplo de faces detectadas no módulo de Detecção de Faces.	26
3.2	Exemplo de face detectada no módulo de Detecção de Faces, salientando as regiões de interesse(ROI) identificadas por antropometria para os olhos direito e esquerdo, e sobrancelhas direita e esquerda.	27
3.3	Estrutura do modelo de Detecção de Íris.	28
3.4	Mostra dos 14 pontos estipulados pelo padrão MPEG-4 de animação facial (31) para olhos e sobrancelhas.	29
3.5	Arquitetura interna do modelo de Detecção das Características Faciais.	30
3.6	Demonstração da transformação entre os espaços de cor, a fim de obter o vermelho invertido.	31
3.7	Demonstração da transformação exponencial aplicada ao canal de cor vermelho invertido.	32
3.8	Demonstração dos resultados obtidos com a binarização para as regiões da sobrancelha direita e esquerda, e para as regiões do olho direito e esquerdo.	33
3.9	Exemplo de resultados obtidos para a busca por <i>blobs</i>	34
3.10	Exemplo de <i>blobs</i> válidos obtidos.	35
3.11	Ilustração dos pontos característicos localizados.	36
3.12	Exemplos das características faciais localizadas.	36
3.13	Arquitetura do modelo proposto para detecção de características faciais.	38
3.14	Estrutura do módulo de Entrada de Vídeo.	39

4.1	Ampliação da região do olho, mostrando a dificuldade em especificar um único ponto para o canto do olho em uma imagem de resolução de 640x480. O quadrado branco tem 5 pixels de lado. Qualquer ponto no interior desse quadrado poderia ser considerado uma resposta correta para detecção do canto do olho.	42
4.2	Imagem com informações de <i>Ground Truth</i> numeradas, referentes a base de dados IMM (30).	43
4.3	Imagem com os 8 pontos em comum entre as aplicações.	44
4.4	Gráfico contendo os 4 pontos referentes aos olhos analisados no modelo. . .	45
4.5	Gráfico contendo os 4 pontos referentes às sobrancelhas analisados no modelo.	46
4.6	Diferença entre as localizações de componentes faciais do <i>Ground Truth</i> da base de dados IMM para os pontos localizados na aplicação de Milborrow (28).	47
4.7	Exemplo de resultados para os três modelos testados, seguindo a ordem: as imagens da coluna da esquerda são referentes ao modelo proposto, central referem-se ao aplicativo Cristinacce (10) e as imagens da coluna da direita são resultados do aplicativo Milborrow (28)	49
4.8	MPEG4 <i>Face Parameter Units</i> para os olhos. Estas unidades são utilizadas para mapear o deslocamento das pálpebras de um usuário real para um avatar (32).	50
4.9	Exemplo de animação de avatar 3D a partir de imagens reais.	51

Lista de Tabelas

2.1	Proporções das distâncias (D2, D3, D4 e D5) para D1 medido a partir de pessoas de diferentes <i>landmarks</i>	24
4.1	Porcentagem das distâncias menores que 5 pixels, da posição encontrada da íris em relação a informação de <i>ground thuth</i>	40
4.2	Comparação dos percentis de valores de d obtidos através da aplicação do modelo proposto em diferentes espaços de cor.	41
4.3	Médias de erro obtida nas aplicações.	44
4.4	Desvio padrão do erro comparado com a aplicação utilizada.	45

LISTA DE SIGLAS

IHC	Interação Humano Computador
PDA	Performance Driven Animation
DTs	<i>Deformable Templates</i>
ASM	<i>Active Shape Models</i>
PDM	<i>Point Distribution Model</i>
AAM	<i>Active Appearance Models</i>
KLT	Kanade-Lucas-Tomasi
PCA	<i>Principal Component Analysis</i>
Fps	<i>Frames per second</i>
ROI	Region of interesting
SAD	Sum of Absolute Difference
IR	Inverted Red
OpenCV	Intel Open Source Computer Vision Library
RGB	<i>Red, Green and Blue</i>
HSV	<i>Hue, Saturation and Value</i>
YCbCr	<i>Luminance, Blue Chrominance an Red Chrominance</i>
FPU	<i>Face Parameter Units</i>
ESO	<i>Eye Separation</i>
IRISDO	<i>Iris Diameter</i>
FAP	<i>Facial Animation Parameters</i>
VHLAB	<i>Virtual Humans Animation and Simulation</i>

Sumário

Lista de Figuras	8
Lista de Tabelas	10
Lista de Abreviaturas	11
1. Introdução	14
1.1 Problema	14
1.2 Motivação	15
1.3 Objetivos	15
1.4 Estrutura da Dissertação	16
2. Trabalhos Relacionados	17
2.1 Máscaras Deformáveis	17
2.2 <i>Wavelets</i> de Gabor	18
2.3 Modelos de Forma Ativos - <i>Active Shape Models (ASM)</i>	19
2.4 Modelos de Aparência Ativos - <i>Active Appearance Models (AAM)</i>	20
2.5 Fluxo Óptico	21
2.6 Outras Abordagens	21
2.7 Contexto deste trabalho no estado-da-arte	23
3. Modelo Proposto	25
3.1 Detecção da Face	25
3.2 Detecção da Íris	25
3.3 Detecção das Características Faciais	28
3.3.1 Definição do Canal de Cor	29
3.3.2 Transformação Exponencial	29
3.3.3 Binarização da Imagem	31
3.3.4 Busca por <i>Blobs</i>	32
3.3.5 Seleção e Validação do <i>Blob</i>	33
3.3.6 Localização das Características	35
3.3.7 Filtragem dos Dados	36
3.4 Protótipo	37

3.4.1	Entrada de Vídeo	38
4.	Resultados	40
4.1	Detecção da Íris	40
4.2	Diferentes Espaços de Cor	40
4.3	Avaliação do Modelo	43
4.4	Animação de Avatares 3D	50
5.	Considerações Finais	52
5.1	Trabalhos Futuros	52
	Referências Bibliográficas	54

1. Introdução

A detecção de componentes faciais baseada em vídeo representa, atualmente, uma importante área de pesquisa, podendo ser aplicada em inúmeras áreas da ciência e tecnologia, tais como as citadas por (5; 47; 27; 4; 39):

- Reconhecimento de faces;
- Interação Humano Computador (IHC) ;
- Teleconferências;
- Animação;
- Análise de expressões faciais;
- Aplicações para deficientes físicos (leitura de lábios).

Também podem ser utilizadas em aplicações que necessitam de alto grau de precisão na identificação das características faciais, que é o caso da análise de expressões faciais para fins de animação de personagens (*avatares* 3D), como o que ocorre no filme *Avatar*¹, que são personagens animados a partir de expressões realizadas por pessoas reais.

Nas próximas seções serão discutidos o problema, motivação, objetivos e a estrutura desta dissertação.

1.1 Problema

O problema sob investigação desta pesquisa é a seguir caracterizado: dada uma imagem extraída de vídeo, encontrar a posição de uma seqüência de *landmarks* (pontos de referência) referentes aos componentes faciais que representam os olhos e sobrancelhas de uma pessoa. Estes pontos localizados podem ser usados como características faciais, para a extração de expressões faciais. Aplicações diretas podem ser em Interação humano computador (IHC) e na animação de avatares 3D (27; 23; 4).

Existem vários métodos para a extração dos componentes faciais (23; 44; 16), mas a maioria possui limitações, tais como:

- Não funcionar em tempo real;
- Ser sensível a variações de iluminação;

¹<http://www.avatarmovie.com/index.html>

- Ser sensível a ruídos gerados pelo dispositivo de câmera;
- Possuir dependência de orientação;
- Possuir dependência do espaço de cor utilizado;
- Não permitir que gestos sobreponham características (ex. mão sobrepondo um dos olhos).

Algumas destas limitações deverão ser tratadas nesta pesquisa.

A questão de pesquisa deste trabalho é endereçar alguns destes aspectos, provendo processamento em tempo real de imagens capturadas por webcam.

1.2 Motivação

Existem diversas situações em que a detecção de componentes faciais desempenha um papel importante, como as citadas anteriormente: IHC, reconhecimento facial, análise de expressões faciais, animação, dentre outras. A principal motivação na escolha deste tema para a dissertação foi o desafio de detectar de uma forma precisa e rápida os componentes da região dos olhos, tendo em vista as limitações existentes na área.

Os olhos e sobrancelhas são fontes representativas dos sentimentos humanos (15). Assim sendo, o algoritmo proposto pode ser útil em diferentes aplicações. Uma das possíveis aplicações seria o monitoramento da fadiga de motoristas ou operadores de máquinas perigosas (22; 12). Se o algoritmo captar que o usuário permanece por muito tempo com os olhos fechados poderá ser disparado um sinal sonoro para despertá-lo.

Na área de cinema e jogos, surgiu a motivação de fazer o mapeamento dos componentes faciais para *avatares* 3D, para assim, a partir destas animações, gerar avatares mais convincentes e interativos.

Na Figura 1.1, pode-se observar um exemplo de aplicação de animação dirigida por performance (*Performance Driven Animation - PDA*), que consiste na animação de personagens 3D a partir de pontos pré-definidos em pessoas. Nesta figura as expressões faciais do usuário são capturadas por algoritmos de visão computacional e mapeados para um *avatar* 3D (14).

Neste exemplo, a precisão dos dados capturados tem um grande impacto na qualidade da animação gerada. Este é um desafio da área.

1.3 Objetivos

De forma mais específica, são objetivos desse trabalho:

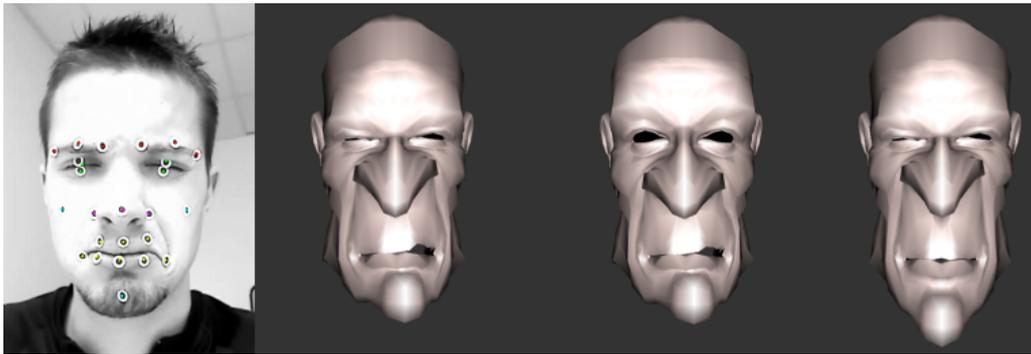


Figura 1.1: Exemplo de resultado do trabalho de Dutreuve *et al.* (14)

- Realizar um estudo sobre as técnicas utilizadas na literatura para a detecção de componentes faciais;
- Definir uma arquitetura para o modelo de detecção de componentes faciais;
- Desenvolver o modelo de detecção de componentes faciais;
- Gerar o mapeamento dos componentes faciais para avatares 3D;
- Comparar os resultados obtidos com outras técnicas.

1.4 Estrutura da Dissertação

No Capítulo 2, são apresentados diversos trabalhos científicos relacionados ao tema desta dissertação, os quais serviram como referencial teórico para o desenvolvimento do modelo proposto no Capítulo 3. O Capítulo 3 apresenta o modelo proposto para a detecção de componentes faciais, sendo dividido em diversos módulos para uma melhor elucidação do seu funcionamento.

Os resultados obtidos nos testes do modelo são expostos no Capítulo 4. Já no Capítulo 5 são expostas algumas considerações sobre o trabalho realizado, destacadas as principais contribuições e sugeridos alguns trabalhos futuros.

2. Trabalhos Relacionados

Neste capítulo será apresentada uma revisão bibliográfica dos trabalhos referentes à detecção de componentes faciais, como boca, narinas, olhos e sobrancelhas. Serão descritas as diferentes técnicas utilizadas pelos autores a fim de identificar tais características. Estas técnicas serão expostas nas Seções subsequentes.

2.1 Máscaras Deformáveis

Yuille e colegas (51) propuseram as *Máscaras Deformáveis* (DTs - *Deformable Templates*). Tais máscaras, ao invés de serem definidas pelas intensidades dos níveis de cinza dos pixels, são definidas por funções parametrizáveis que descrevem características do objeto tais como contornos e picos e vales nas intensidade dos pixels da imagem. A função de comparação passa a ser uma função de energia que deve ser minimizada. Tais funções são descritas por potenciais, cujas forças são construídas de tal forma a representar certas propriedades dos objetos de busca.

Como exemplo, considere a máscara representando os olhos mostrada na Figura 2.1, apresentada pelos autores em (51). A máscara deformável que representa os olhos é descrita por duas parábolas representando as pálpebras, uma circunferência representando a íris e dois pontos correspondendo ao centro das duas partes da esclera localizadas entre a íris e as pálpebras.

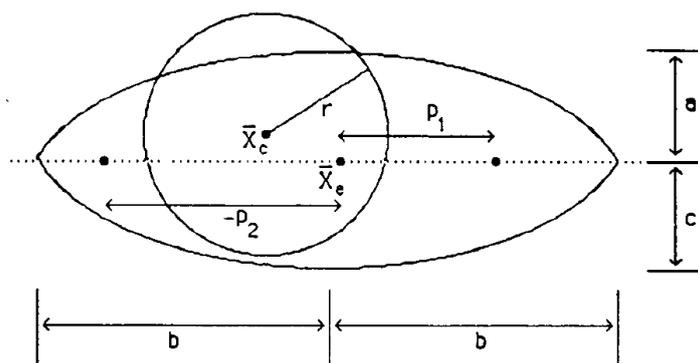


Figura 2.1: Template para olhos proposto por Yuille et al. (51). Duas parábolas descrevem os limites das pálpebras e uma circunferência descreve a íris. O rastreamento é feito através do ajuste dos parâmetros a , b e c das parábolas, P_1 , P_2 do centro da esclera, r e x_c da íris à imagem.

2.2 Wavelets de Gabor

A transformada *Wavelet* é utilizada para a obtenção de uma análise multi-resolução, tanto no domínio espacial quanto no domínio de freqüências de sinais (no caso de imagens, sinais bi-dimensionais).

Wei e colegas propuseram o uso de *wavelets* de Gabor para a identificação inicial de componentes faciais em (45). Nesse sistema a aquisição da imagem é feita através de captação de infra-vermelho, o que facilita a localização exata das pupilas. Foram utilizadas 18 *wavelets*, com 3 coeficientes de escala ($m = 3$) e 6 coeficientes de rotação ($n = 6$). Após a identificação dos pontos dos componentes faciais, o algoritmo proposto faz o rastreamento da posição da face e dos componentes faciais baseando-se no movimento das pupilas e utilizando filtros de Kalman. Esta abordagem utiliza Infra-vermelho (*Infra-Red*) para realizar o rastreamento das características faciais.

Tong e colaboradores (41) utilizaram *wavelets* de Gabor e gradientes de níveis de cinza para detecção automática dos pontos principais da face no quadro inicial, baseando-se na distância entre a matriz de coeficientes de Gabor obtida durante um treinamento com banco de dados de face anotados e a matriz de coeficientes de Gabor da imagem do quadro inicial. Assim como em (45), foram utilizadas 18 *wavelets* de Gabor.

No trabalho de Vukadinovic e Pantic (43), é utilizada uma versão adaptada do algoritmo de Viola e Jones (42) para detectar a face, e posteriormente são demarcadas regiões de interesse para busca das componentes faciais. Nestas regiões são aplicados filtros de Gabor para identificar as características faciais, sendo identificadas 19 características faciais, como pode ser observado na Figura 2.2.



Figura 2.2: Demonstração das 19 características identificadas pelo método de (43), em duas imagens distintas

2.3 Modelos de Forma Ativos - *Active Shape Models (ASM)*

O principal objetivo dos autores ao propor os Modelos de Forma Ativos (9), (ASM) foi construir modelos que pudessem ser deformados apenas de maneiras permitidas, específicas do objeto que estão representando.

Tong e colaboradores (40), (41), utilizam *Active Shape Model (ASM)*. Nesse trabalho, os autores utilizaram 26 pontos para construção do modelo de distribuição de pontos (*Point Distribution Models - PDM*), distribuídos conforme a Figura 2.3. Para detecção dos pontos, os autores utilizaram *wavelets* de Gabor (modelo descrito na Seção 2.2).

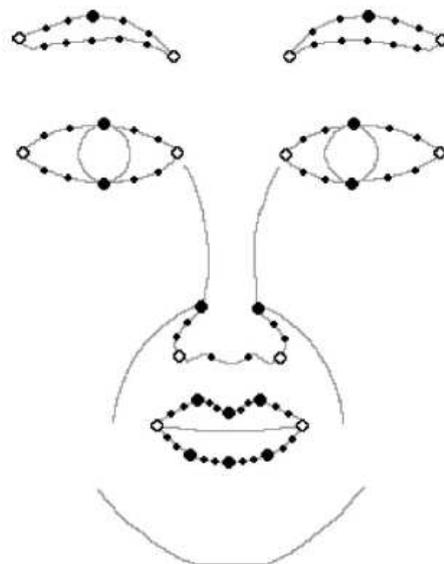


Figura 2.3: Distribuição de pontos baseada no padrão MPEG4, utilizado por Tong *et al.* (41).

Nesse trabalho, Tong *et al.* desenvolveram um modelo hierárquico que, a partir dos resultados encontrados por ASM, determina estados para componentes faciais como olhos fechados, olhos abertos, boca aberta, etc.

Zuo e colaboradores (53) propõem a combinação de ASM com a aplicação de uma técnica de casamento descrita como *Haar-wavelet*. Esta abordagem destaca-se pela velocidade para identificação, que segundo o autor é de 30 – 70ms em imagens de 300x300 pixels.

O trabalho de Mahoor e Abdel-Mottaleb (27), propõem um melhoramento na técnica de ASM, que consiste em identificar o centro dos olhos e boca para a inicialização da ASM e através destas coordenadas, aplica novos filtros na imagem para salientar as características da boca, para com isso conseguir fazer uma identificação de pontos mais correta nesta região.

São demonstrados em (27) resultados referentes ao reconhecimento de faces utilizando essa abordagem e o ASM padrão, e pode ser observada uma melhora no casamento do

modelo com as características faciais. Na Figura 2.4, pode ser observado um destes resultados citados.

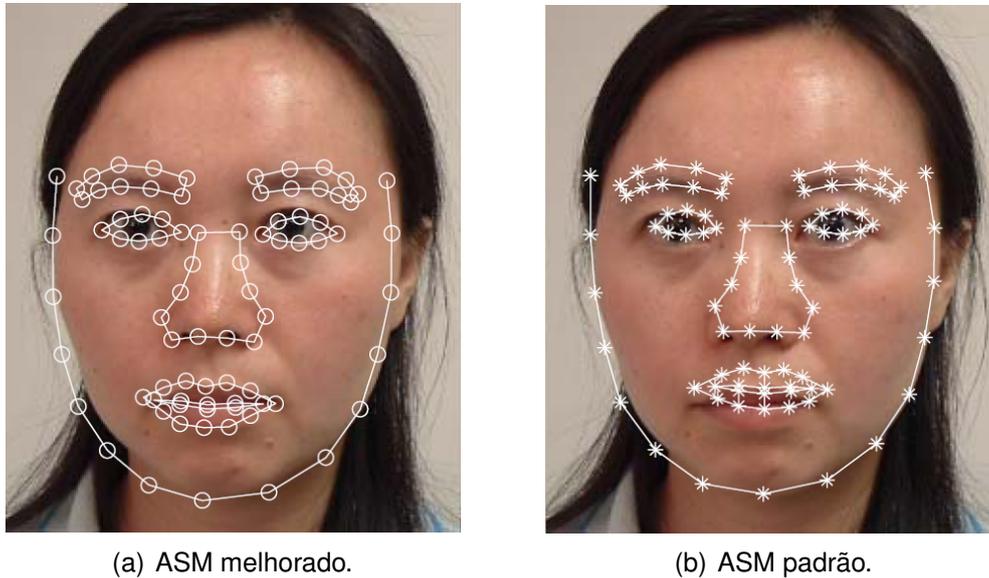


Figura 2.4: Exemplo da extração de componentes faciais (27).

Haj e colaboradores (17) utilizaram ASM para detecção de componentes faciais, uma vez tendo sido encontrada a face em um quadro de vídeo. Nesse trabalho, os autores utilizaram dois PDMs ao invés de um, associando dois ASMs a esses pontos. Esse passo é utilizado na inicialização das posições de uma malha 3D (Candide-3 (20)), a qual é usada no rastreamento do movimento das sobrancelhas, lábios e pálpebras.

2.4 Modelos de Aparência Ativos - *Active Appearance Models (AAM)*

Os mesmos autores que propuseram o ASM, desenvolveram um outro modelo para detecção de formas, o qual chamaram *Active Appearance Models*, ou Modelos de Aparência Ativos (8).

Dornaika e Davoine (13) propuseram a obtenção simultânea tanto da posição da cabeça quanto da deformação da face. Para o rastreamento da posição e ângulo da cabeça, os autores utilizaram filtros de partículas e para obter as deformações dos componentes faciais, utilizaram AAM.

Zalewski e Gong (52) utilizaram AAM para classificação de estados como alegria, neutralidade, raiva, tristeza, surpresa, medo e repulsa baseados na expressão facial. Para tanto, na fase de treinamento, os autores preferiram utilizar diferentes imagens de uma mesma pessoa com diferentes expressões, a fim de aumentar a robustez. Além disso, os autores estenderam o AAM para incorporar implicitamente variação de pose (rotação e

deslocamento da cabeça) na distribuição estatística.

2.5 Fluxo Óptico

O principal objetivo da técnica de Fluxo Óptico é estimar movimento. A idéia principal é calcular uma aproximação do campo de velocidades a partir do movimento da intensidade dos pixels da imagem em uma seqüência de vídeo (34). Variações na intensidade de iluminação afetam grandemente a estimativa do fluxo óptico. Existem técnicas para aumento da robustez, como a proposta por Kanade-Lucas-Tomasi (KLT) (26).

DeCarlo e Metaxas (11) propuseram o uso de fluxo óptico para rastreamento de faces, levando em consideração as restrições impostas por um modelo deformável de face. A proposta dos autores consiste em projetar os pontos rastreados na imagem no modelo 3D para melhor estimativa da pose da cabeça e da deformação dos componentes faciais.

He e colegas (18) propuseram um modelo para rastreamento de características faciais baseado no algoritmo KLT. Para aumentar ainda mais a robustez do algoritmo, eles utilizaram restrições impostas pelas próprias características da face como distância entre as pupilas, comprimento e altura da boca, comprimento do nariz e comprimento dos olhos. Com essas restrições, os autores dizem ter alcançado um melhor desempenho do que simplesmente utilizando KLT.

Hsieh e pesquisadores, utilizaram fluxo óptico com restrições para reconhecimento de faces de indivíduos com diferentes expressões em (19).

2.6 Outras Abordagens

Algumas abordagens tinham como foco inicial a detecção de pele e posteriormente o tratamento das imagens, como é o caso dos trabalhos de Chang *et al.* (5) e Wu *et al.* (47). Chang (5) utiliza variação de limiares na escala de cromaticidade de imagem para identificar o ponto central dos olhos, nariz e boca, e Wu *et al.* (47) utilizam *active contour model* para estimar a região dos olhos, nariz e boca, como pode ser observado na Figura 2.5.

Em Xue *et al.* (49), é proposta uma abordagem nova para a época, a qual propõe a utilização de *Bayesian Shape Model* (25) para extrair os componentes faciais e juntamente com dados obtidos através de PCA (*Principal Component Analysis*). O modelo visa ajustar as formas ao modelo e extrair a face para gerar uma animação da mesma em um modelo.

Já Yen e Nithianandan (50) propõem um método baseado em Algoritmos genéticos para localizar a melhor região elíptica que venha a fazer parte dos olhos, nariz ou boca. Este método foi testado em imagens com iluminação natural e alguma adição de ruído.

No trabalho de Cheddad *et al.* (6), é proposta uma abordagem diferente das demais: são utilizados diagramas de *Voronoi* e triangularização de *Delaunay* para identificação de face e extração das características faciais.



Figura 2.5: Exemplos da extração de contornos feita a partir dos pontos característicos localizados (47)

Panning *et al.* (33) utilizam *Haar Classifiers* (46) para detectar as componentes faciais, sendo desenvolvido um treinamento para classificadores dos componentes desejados. Utilizando estes classificadores treinados são identificados 13 componentes faciais, e a partir destas componentes é feita uma análise através de redes neurais para classificar 5 tipos de expressões faciais (raiva, medo, alegria, surpresa e desgosto).

Uma outra abordagem interessante é proposta por Xu *et al.* (48), exclusivamente para a detecção dos cantos dos olhos. Nesse trabalho é construído um modelo para identificação dos cantos com base na comparação dos pontos achados com os pontos vizinhos, sendo um dos parâmetros para validação do ponto o ângulo que este forma com a pálpebra.

Chen e colegas (7) utilizaram o método de agrupamento de dados *k-means* para encontrar os contornos das sobrancelhas e da parte inferior da face (queixo e bochechas). Os autores utilizaram o algoritmo proposto em trabalhos anteriores para encontrar os cantos dos olhos, a parte superior da pálpebra e região da sobrancelha. Como a sobrancelha pode não ser homogênea, os autores sugeriram que a região extraída fosse dividida em três sub-regiões horizontais. Em cada uma dessas sub-regiões, foi aplicado o método de agrupamento de dados *k-means* sobre o valor do nível de cinza dos pixels.

Cristinacce e Cootes (10) propuseram um rastreador por seleção de *templates* de imagens de face rotuladas manualmente, que usa distância euclidiana. Os autores afirmam terem chegado a resultados semelhantes aos obtidos em sua implementação de AAM em bases de imagens públicas e até mesmo superiores em imagens de faces em pessoas dentro de carros, embora a implementação do modelo proposto seja mais simples.

Sohail e Bhattacharya (37) desenvolveram um modelo de face baseado em estatísticas

antropométricas para detecção dos 18 pontos-chave mais importantes. Nesse modelo, a distância entre as duas pupilas obtida através de técnicas de detecção e classificação de objetos proposta por Fasel e colegas (16) serve como principal parâmetro de medida para a localização do centro dos outros componentes faciais. Os autores utilizaram operações morfológicas, binarização e detecção de bordas para obtenção de características que foram submetidas a um conjunto de regras geométricas e baseadas na intensidade dos pixels para localização de cada componente facial que os autores consideraram relevante. Por exemplo, para detecção dos olhos, binariza-se a região próxima à localização da pupila, determina-se componentes conexos utilizando um elemento estruturante de 4 pixels e, para o olho direito, determina-se que o canto direito do componente conexo é o canto direito do olho.

Os sete pontos de referência utilizados pelo método e as distâncias calculadas podem ser observadas na Figura 2.6, e as proporções obtidas podem ser observadas na Tabela 2.1.

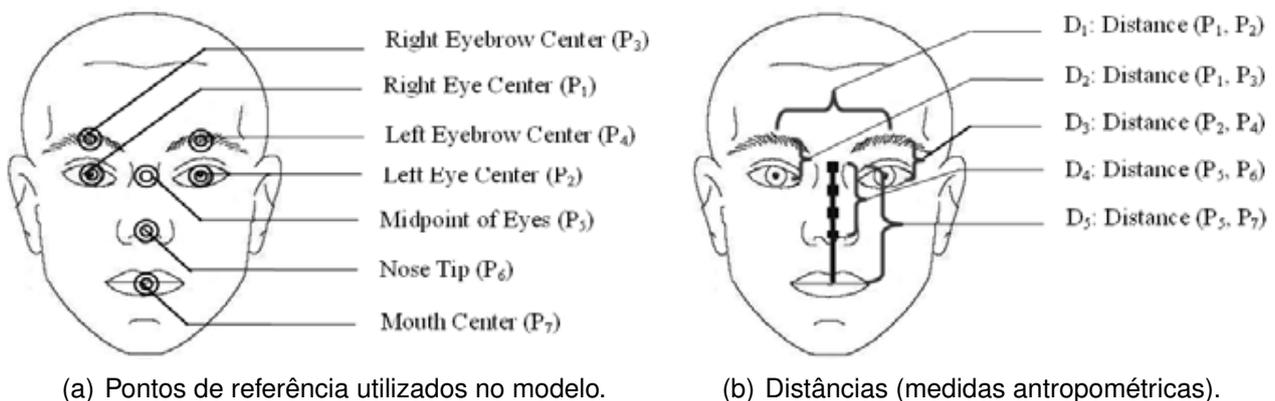


Figura 2.6: Modelo Facial proposto por Sohail and Bhattacharya (37) para localização das regiões faciais.

Existem inúmeros outros trabalhos que se propõe a extrair as características faciais, sendo a maioria voltados para a área de reconhecimento facial. Nesta seção foram citados apenas os trabalhos que tiveram maior afinidade com a proposta do modelo a ser desenvolvido nesta dissertação.

2.7 Contexto deste trabalho no estado-da-arte

Este trabalho utiliza uma abordagem um pouco diferente das citadas nas seções anteriores. Devido a necessidade da aquisição de imagens a partir de webcams comuns e do pressuposto que o modelo deve ser executado em tempo real (tratar as imagens de entrada a uma taxa de quadros por segundo próxima de 30 fps), partiu-se para uma abordagem

Tabela 2.1: Proporções das distâncias (D2, D3, D4 e D5) para D1 medido a partir de pessoas de diferentes *landmarks*.

Proporção	Descrição	Constante
D2/ D1	Proporção da distância entre o centro do olho direito e o centro da sobrancelha direita à distância entre os centros dos olhos	≈ 0.33
D3/ D1	Proporção da distância entre o centro do olho esquerdo e o centro da sobrancelha esquerda à distância entre os centros dos olhos	≈ 0.33
D4/ D1	Proporção da distância entre o ponto médio entre dos centros dos olhos e a ponta do nariz à distância entre os centros dos olhos	≈ 0.60
D5/ D1	Proporção da distância entre o ponto médio entre os centros dos olhos e o centro da boca à distância entre os centros dos olhos	≈ 1.10

voltada à performance.

Assim, a hipótese definida leva em consideração a detecção de componentes faciais da região dos olhos baseada em modelos de cor e antropometria, a fim de reduzir o custo computacional comparado com outros métodos.

Foram utilizados *Haar Classifiers* para a identificação da face na imagem, algoritmo de Viola e Jones (42), similar aos utilizados por Panning e colaboradores (33), e as medidas antropométricas calculadas por Sohail e Bhattacharya (37), bem como algumas outras abordagens descritas no decorrer deste trabalho.

Este modelo destaca-se dos demais pela sua simplicidade, e funcionar em webcams.

3. Modelo Proposto

Neste capítulo é apresentado o modelo proposto para a detecção de características faciais da região dos olhos. Este modelo consiste em duas partes distintas: um módulo desenvolvido no âmbito do laboratório de pesquisa¹ para detecção da íris e o modelo desenvolvido para a detecção de componentes faciais.

Inicialmente será detalhada a abordagem desenvolvida para detecção da face e identificação das regiões de interesse. Na Seção 3.2 será detalhado o modelo para a identificação da íris, na Seção 3.3 é explicado o modelo desenvolvido para detecção da região dos olhos e sobrancelhas. Já na Seção 3.4, é apresentado o protótipo desenvolvido para a detecção das características faciais.

3.1 Detecção da Face

Este módulo tem o objetivo de detectar uma face na imagem de entrada. Para este fim, utilizou-se o método proposto por Viola e Jones (42), que caracteriza-se pela codificação de informações previamente treinadas, para efetuar a detecção. Neste módulo foi utilizado um treinamento específico para a identificação de poses frontais. Este método utiliza características *Haar-like* (42) para codificar a existência de contrastes entre regiões retangulares adjacentes na imagem, a partir dessas características é possível identificar diferentes características faciais humanas.

Utilizou-se a implementação proveniente da biblioteca *OpenCV*, especificamente o algoritmo *HaarDetectObjects*, que provê a implementação do método proposto por Viola e Jones (42), Este algoritmo retorna um retângulo, onde o rosto está inscrito. Exemplos de imagens resultantes deste algoritmo podem ser observados na Figura 3.1, onde são demonstradas as imagens originais e a saída obtida após a execução deste módulo.

A identificação das regiões de interesse para os olhos e sobrancelhas foi feita utilizando medidas antropométricas apresentadas por (37), a fim de estimar onde provavelmente estariam situadas estas características faciais. A Figura 3.2 mostra um exemplo destas regiões de interesse.

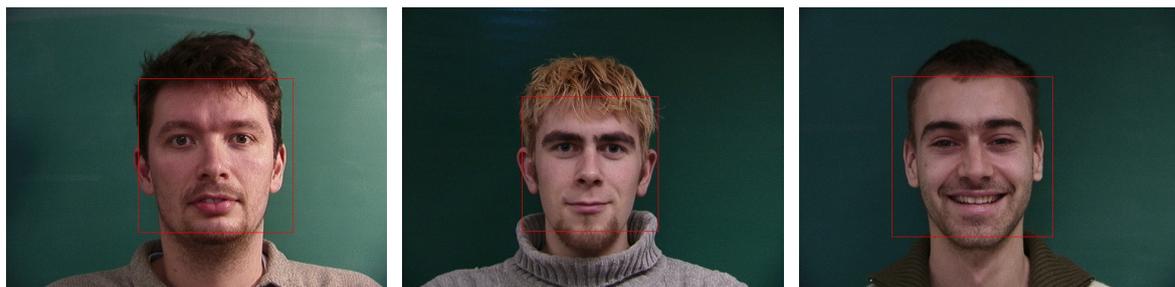
3.2 Detecção da Íris

Uma contribuição deste trabalho foi a colaboração no desenvolvimento de um detector de íris (21), baseado em *template matching*, que é uma técnica bastante conhecida em processamento de sinais, para identificação da íris. As fases deste modelo podem ser

¹VHLAB - <http://www.inf.pucrio.br/~smusse/Lab/vhlab.html>



(a) Imagens obtidas a partir do módulo de Entrada de Vídeo



(b) Imagens com o retângulo da face identificada evidente.



(c) Imagens apenas das regiões de interesse de faces (ROI).

Figura 3.1: Exemplo de faces detectadas no módulo de Detecção de Faces.

visualizadas na Figura 3.3, sendo seu funcionamento brevemente descrito abaixo (detalhes podem ser encontrados em (21)):

- Inicialização do *Template*: é necessário indicar qual *template* deve ser carregado e para que proporção de face ele foi criado, indicada pelo tamanho do raio da face detectado pelo algoritmo de detecção de faces. Alguns *templates* foram testados, sendo um deles escolhido, a partir dos valores obtidos nos resultados demonstrados na Seção 4.1.
- Mudança de tamanho do *template*: conforme o tamanho do raio da face detectada, o *template* sofre aumento ou diminuição de tamanho, isto ocorre devido à necessidade do objeto a ser procurado ter o mesmo tamanho do objeto alvo.
- *Template Matching*: nesta etapa, o *template* é deslocado sobre a imagem dos olhos adquirida no módulo de detecção de face. Posteriormente, é calculada a Soma das

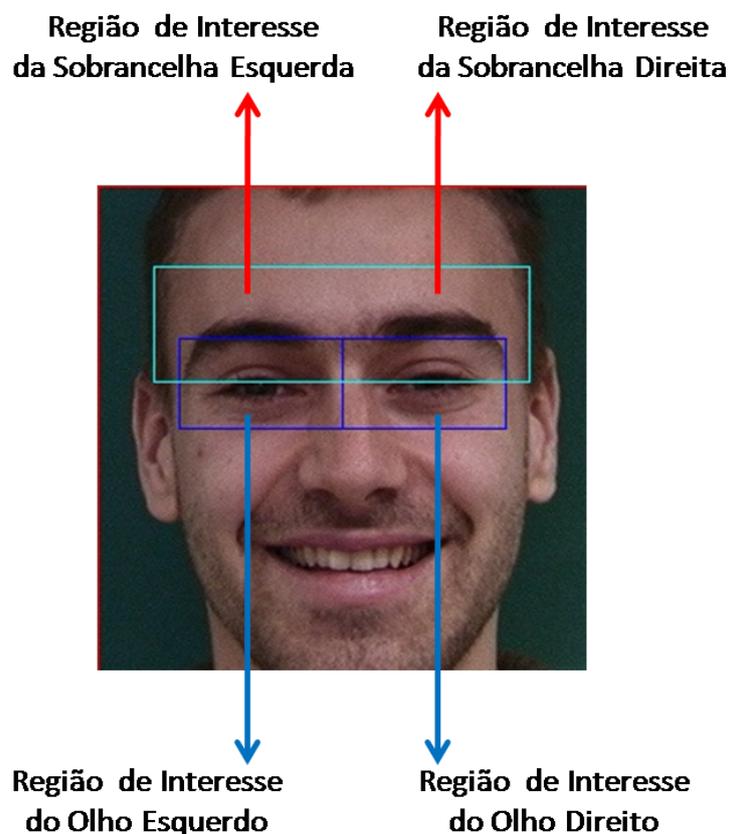


Figura 3.2: Exemplo de face detectada no módulo de Detecção de Faces, salientando as regiões de interesse(ROI) identificadas por antropometria para os olhos direito e esquerdo, e sobrancelhas direita e esquerda.

Diferenças Absolutas (*Sum of Absolute Difference - SAD*) para todos os deslocamentos onde o *template* não possui a cor vermelha, porque nestes locais os valores não são calculados. A posição onde o erro for mínimo deve indicar a posição da íris.

- Retorno do posicionamento da íris: tendo todos os erros calculados anteriormente, nesta etapa é assumido que o erro mínimo computado refere-se à posição da íris na imagem de entrada, sendo esta posição retornada pelo módulo.

Este módulo é optativo na execução do modelo apresentado na Seção 3.3. Conforme sua utilização, ocorrem alterações na fase subsequente descrita na Seção 3.3.5, que é responsável pela seleção e validação dos olhos detectados. A sua utilização implica em uma perda de velocidade na execução do modelo, devido a este método utilizar a técnica de *template matching*.

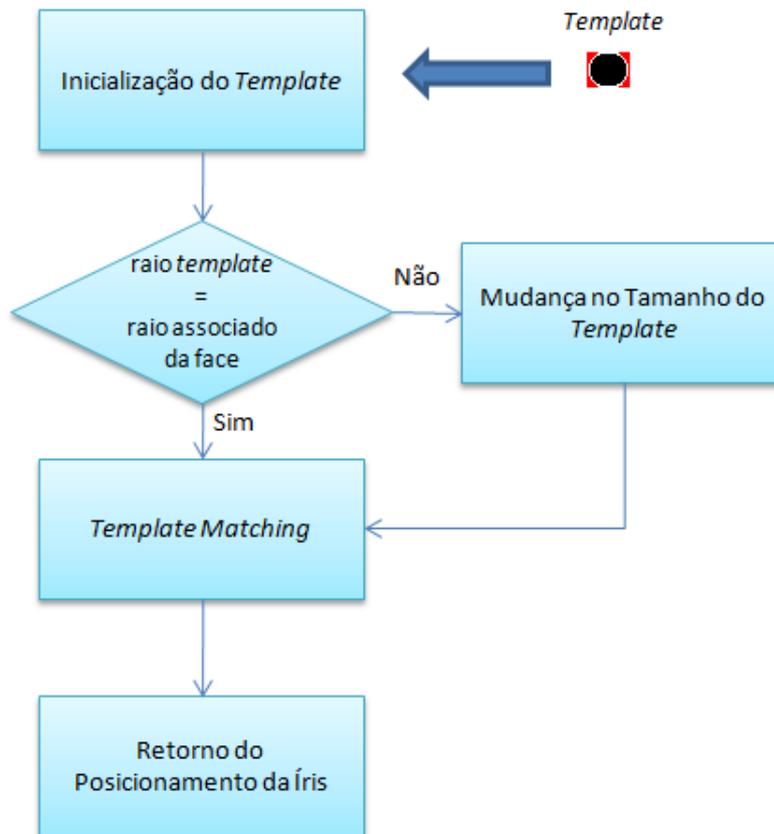


Figura 3.3: Estrutura do modelo de Detecção de Íris.

3.3 Detecção das Características Faciais

Nesta seção será exposto o modelo desenvolvido para identificação das características faciais da região dos olhos.

Um dos objetivos deste trabalho é prover informação para a animação de avatares, tomou-se a decisão de utilizar o padrão MPEG-4 de animação facial (31). Este modelo tem o intuito de identificar 14 componentes faciais, que são representados como localizações neste padrão, para os olhos e sobrancelhas. A Figura 3.4 ilustra os componentes faciais que este trabalho visa identificar. Como primeira aplicação deste modelo, foi animado um avatar, com o intuito de demonstrar o movimento dos olhos. Esta aplicação será apresentada no Capítulo 4.

A Figura 3.5 mostra a arquitetura interna do modelo. Para um melhor entendimento do modelo, os módulos responsáveis pela detecção de características faciais serão explicados separadamente, seguindo a ordem do fluxo de execução, nas próximas seções.

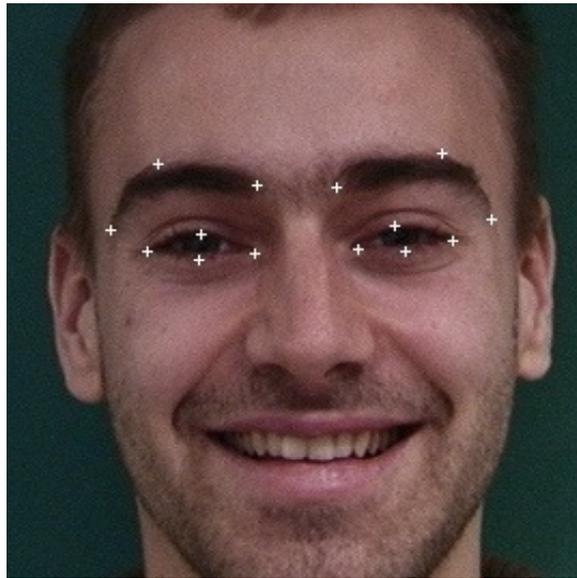


Figura 3.4: Mostra dos 14 pontos estipulados pelo padrão MPEG-4 de animação facial (31) para olhos e sobrancelhas.

3.3.1 Definição do Canal de Cor

Com o intuito de definir o melhor espaço de cor para segmentar as características faciais que o modelo objetiva, adotou-se uma metodologia de variação de contraste da imagem. A hipótese do modelo é de que deve-se destacar os pixels que não correspondem à pele na imagem. Pode-se destacar que os pixels da pele tem um elevado nível de intensidade no canal de cor vermelho (24).

Partindo da hipótese que os contornos dos olhos e sobrancelhas têm baixa intensidade no canal vermelho em comparação com a pele, utilizou-se o canal vermelho invertido (para cada pixel calculamos 255 menos o valor do canal vermelho do pixel). Outros espaços de cor foram testados, como poder ser visto no Capítulo 4.

Deve ser salientado que esta inversão no canal de cor vermelho foi realizada para proporcionar uma melhor visualização dos contrastes na imagem pelos seres humanos, mas não há mudança na intensidade dos pixels.

Na Figura 3.6, é possível observar esta operação de uma forma mais intuitiva.

3.3.2 Transformação Exponencial

Este módulo tem o objetivo de realçar a diferença entre a intensidade dos pixels no espaço de cor vermelho invertido (IR) . Para este fim, utilizou-se o operador exponencial para aumentar a diferença entre a intensidade dos pixels no canal vermelho invertido (29).

O operador exponencial consiste em calcular a seguinte transformação para cada pixel (i, j) da imagem de entrada I :

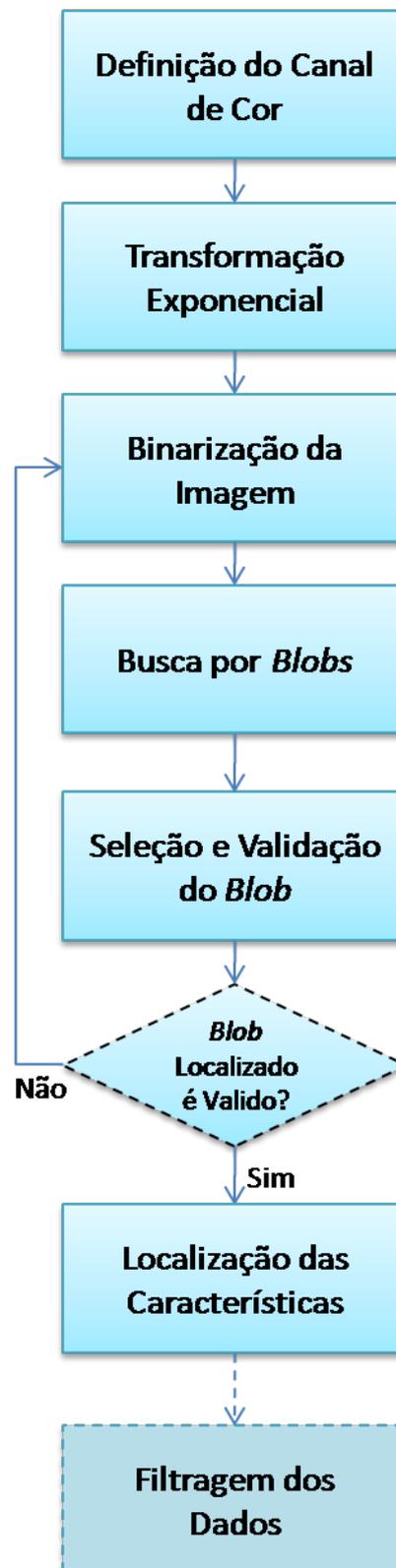


Figura 3.5: Arquitetura interna do modelo de Detecção das Características Faciais.

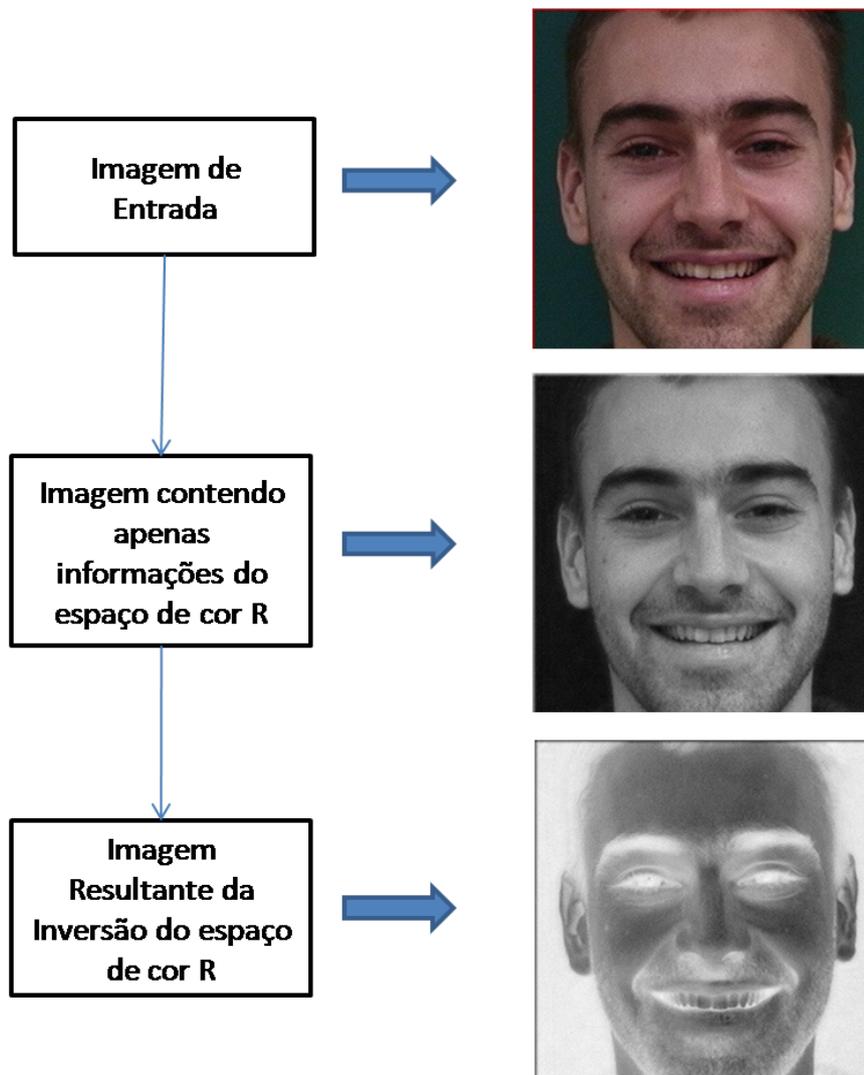


Figura 3.6: Demonstração da transformação entre os espaços de cor, a fim de obter o vermelho invertido.

$$I'_{i,j} = \exp \left[I_{i,j} \cdot \frac{\ln 255}{255} \right] \quad (3.1)$$

A constante expressa na fórmula foi escolhida empiricamente, a partir da análise dos resultados gerados pela transformação.

Os resultados desta transformação podem ser vistos na Figura 3.7.

3.3.3 Binarização da Imagem

Este módulo é responsável pela transformação da região de interesse do quadro de entrada em uma imagem binária. Para realizar a binarização da imagem para uma posterior busca por *Blobs* (componentes conexos), realiza-se a busca por pixels que contêm

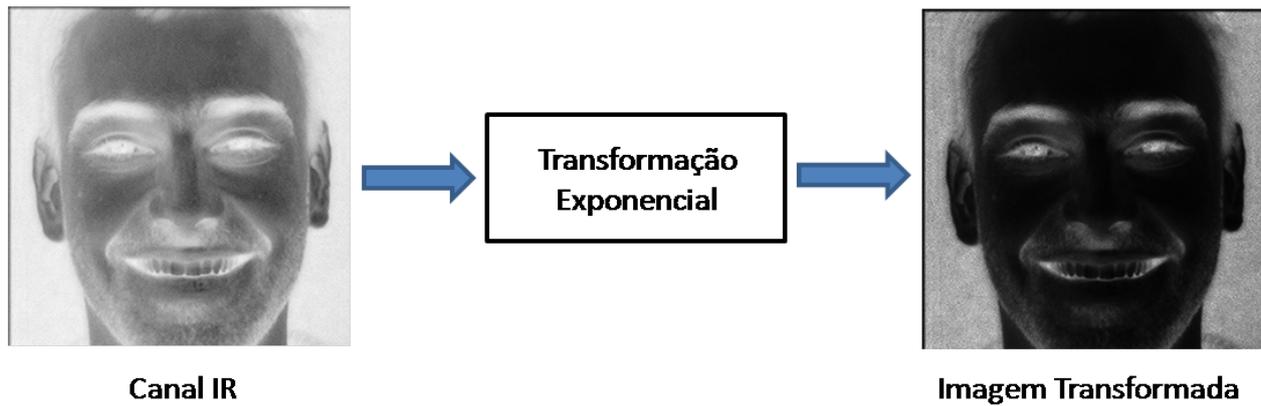


Figura 3.7: Demonstração da transformação exponencial aplicada ao canal de cor vermelho invertido.

intensidades que representam *outliers* em relação a distribuição de intensidades da imagem vermelho invertida como um todo. Desta forma, a imagem binária Bw conterá branco onde estejam representados os *outliers* e os restante será preenchido com preto, como pode ser visto em:

$$Bw_{i,j} = \begin{cases} 1, & \text{if } I'_{i,j} > \bar{I}' + Z\sigma \\ 0, & \text{senão} \end{cases} \quad (3.2)$$

onde, \bar{I}' e σ representam a média e o desvio padrão da intensidade dos pixels de IR, respectivamente, e Z é um valor de controle, o qual seleciona quais pixels vão fazer parte da imagem binária Bw . Quanto maior o valor de Z , maior o número de pixels que serão selecionados. Este parâmetro deve ser calibrado para atingir os melhores resultados. Em nossos experimentos, o valor ideal para Z é 0,9.

Este módulo é chamado separadamente para cada região de interesse. A Figura 3.8, mostra os resultados obtidos após a utilização do módulo de binarização para cada região de interesse que o modelo se propõe a identificar.

3.3.4 Busca por *Blobs*

Neste módulo são utilizadas as bibliotecas do projeto cvBlob², para identificar os componentes conexos na imagem, conhecidos como *blobs*, utilizando vizinhança de 8 pixels (36).

Sempre que uma imagem for binarizada este módulo será chamado para localizar os *blobs*, para uma posterior identificação dos mesmos nos módulos subseqüentes. Pode ser observado na Figura 3.9, um exemplo do resultado que a busca por *Blobs* gera, para a região da sobrancelha direita 3.9-a e do olho direito 3.9-b.

²Website do Projeto se encontra em: <http://code.google.com/p/cvblob/>

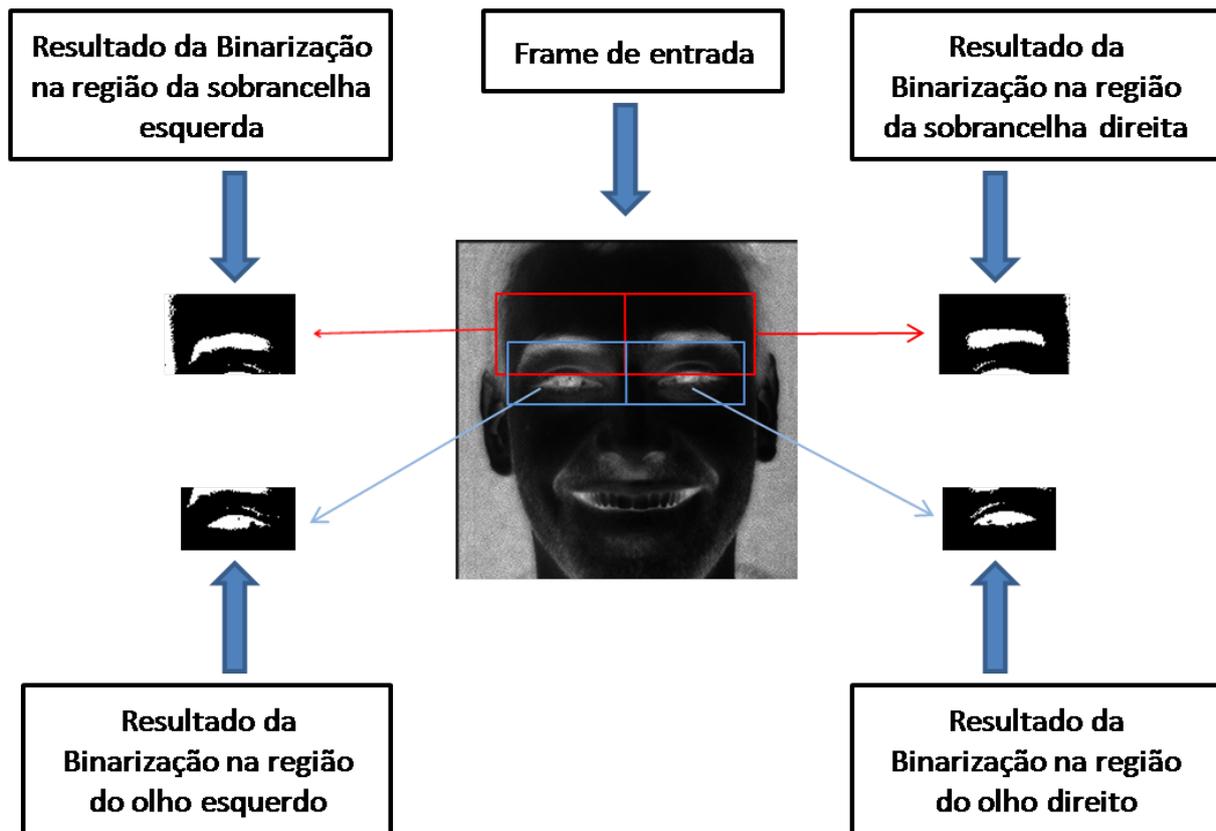


Figura 3.8: Demonstração dos resultados obtidos com a binarização para as regiões da sobrancelha direita e esquerda, e para as regiões do olho direito e esquerdo.

3.3.5 Seleção e Validação do *Blob*

Após a localização dos *blobs* é necessário identificá-los e validá-los. Para identificar os *blobs* são utilizadas duas abordagens distintas, uma para a identificação do *blob* da sobrancelha e outra para a identificação do *blob* do olho. E para o estágio de validação do *blob* da sobrancelha e do *blob* do olho é utilizada apenas uma relação antropométrica de proporção.

Na Figura 3.10, pode ser observado um exemplo do resultado do módulo de Seleção de *Blobs*.

A identificação do *blob* da sobrancelha, Figura 3.10-a, é feita da seguinte maneira: calcula-se a distância que cada *blob* está do ponto central da região de interesse da sobrancelha, o *blob* que tiver a menor distância relativa a este ponto é considerado como sendo o *blob* da sobrancelha.

Já para a identificação do *blob* do olho, Figura 3.10-b, foram utilizadas duas abordagens distintas:

- Utilização do ponto central do olho (íris), obtido através do algoritmo de identificação de íris mostrado na Seção 3.2, como ponto de referência para o cálculo das distâncias

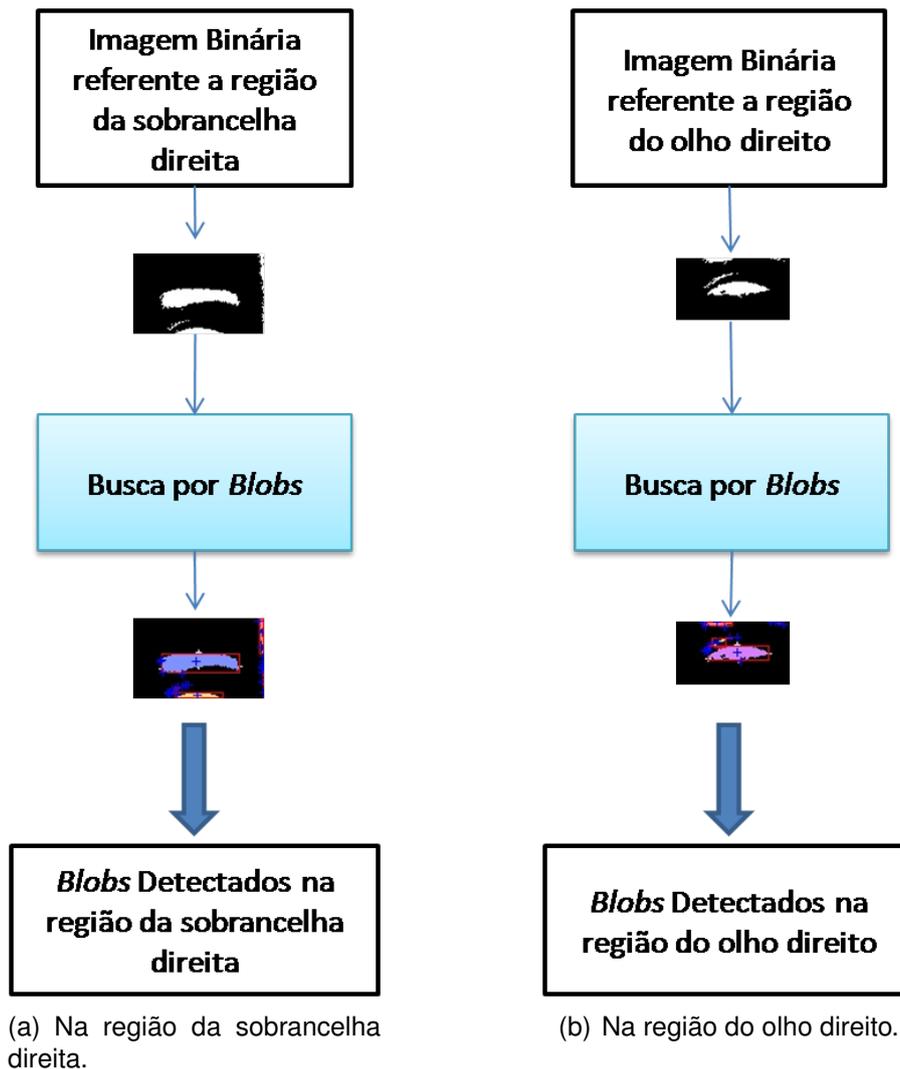


Figura 3.9: Exemplo de resultados obtidos para a busca por *blobs*.

absolutas.

- Utilização do *blob* situado mais abaixo na imagem, com área maior que 10% da ROI, devido ao tamanho dos *blobs* válidos terem mostrado esta proporção nos testes realizados.

Estas duas abordagens apresentaram resultados semelhantes, nos testes realizados com vídeos foi utilizada a segunda abordagem, porque ao utilizar a detecção da íris foi constatado que ao ocorrer o movimento de piscar os olhos, esta abordagem retornava valores errôneos, fazendo com que a identificação do *blob* retornasse o *blob* errado.

No caso da segunda abordagem, é feita uma validação dos *blobs*, levando em consideração a área do mesmo. Se a área do *blob* tiver um tamanho superior a 10% da área ROI, ele será considerado como válido.

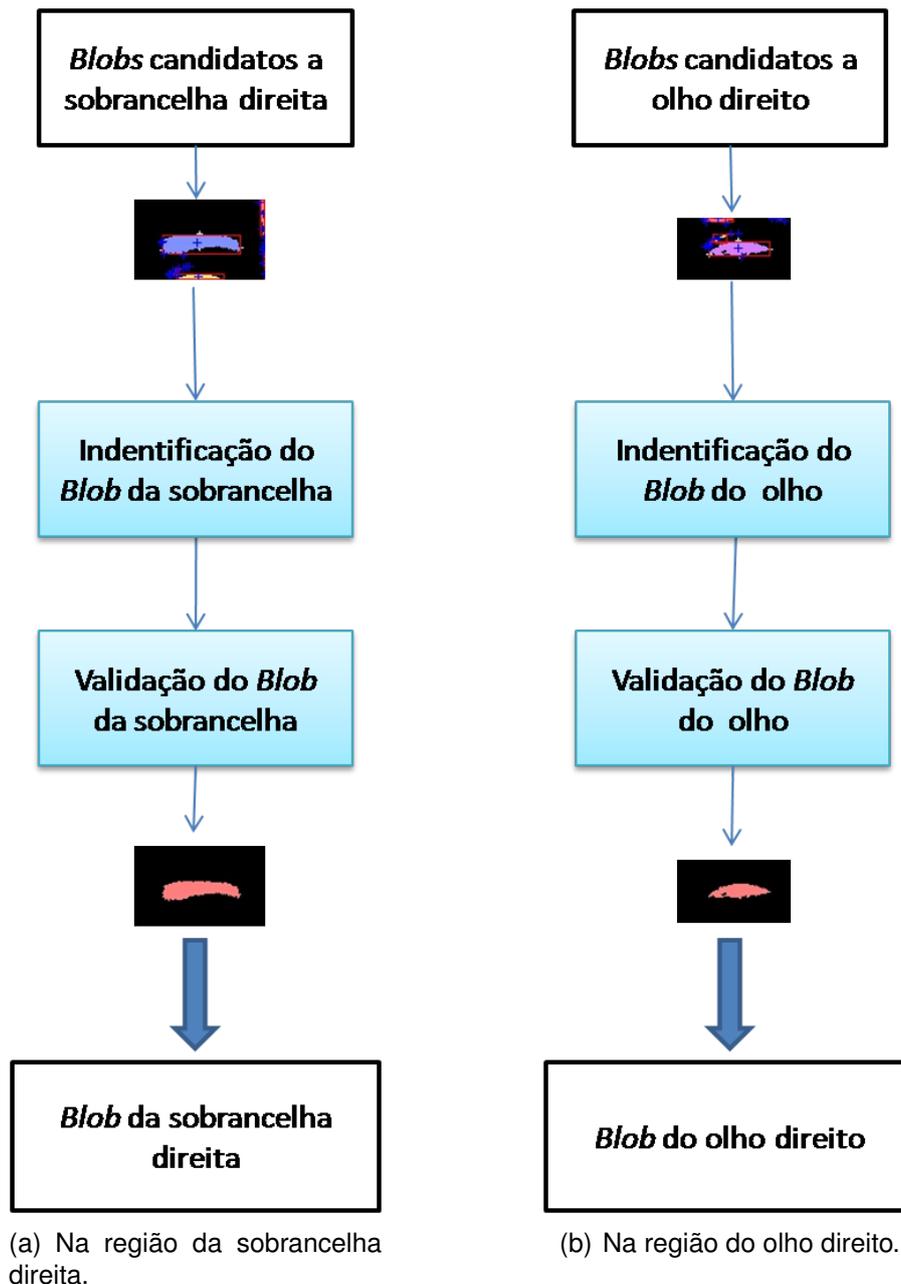


Figura 3.10: Exemplo de *blobs* válidos obtidos.

3.3.6 Localização das Características

Neste módulo é feita a localização dos cantos, situados nos limites horizontais e verticais dos *blobs*, identificados como olhos e sobrançelhas. Os pontos mais a esquerda e mais a direita representam o segmento de linha l_1 e, são considerados como pontos válidos, sendo eles denominados de p_1 e p_2 , respectivamente. Após é traçada uma linha normal ao ponto central de l_1 , o último pixel pertencente a esta linha no sentido superior à normal será considerado como o ponto p_3 e o inferior como ponto p_4 , representando assim os cantos

superior e inferior do *blob*.

A Figura 3.11 ilustra este processo e exemplos de componentes localizados podem ser vistos na Figura 3.12.

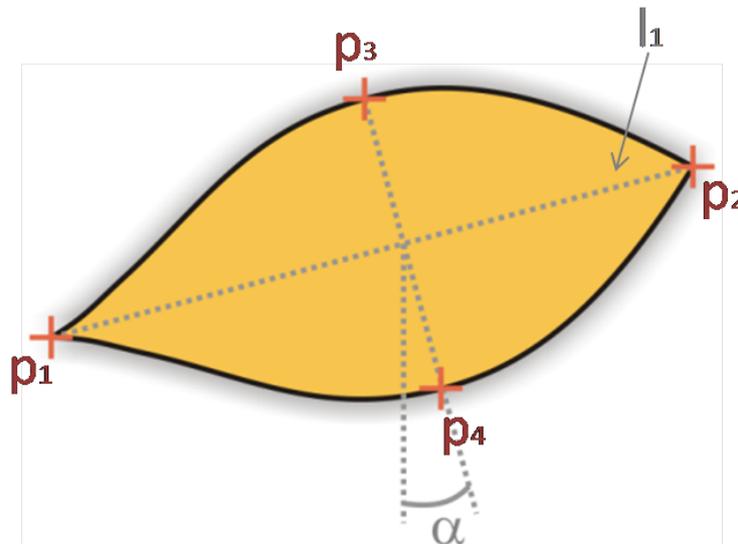


Figura 3.11: Ilustração dos pontos característicos localizados.



(a) Na região da sobrancelha direita.



(b) Na região do olho direito.

Figura 3.12: Exemplos das características faciais localizadas.

3.3.7 Filtragem dos Dados

Este módulo é utilizado apenas em seqüências de imagens com coerência temporal (vídeos). Tem o intuito de filtrar pequenas variações que possam ocorrer entre os quadros, em uma seqüência de imagens, decorrentes de ruídos na captura, compactação, etc. Com o objetivo de filtrar estas instabilidades, a posição do ponto característico é filtrada, utilizando um filtro linear no ponto atual $X = (x, y)$ levando em consideração a posição de n quadros anteriores, foram utilizados no experimento 3 quadros. Este filtro pode ser observado na equação:

$$X = \sum_{i=1}^n \mathbf{b}_i X_{n-i} - \sum_{i=2}^n \mathbf{a}_i X_{n-i} \quad (3.3)$$

onde a e b são matrizes de pesos. Na implementação desenvolvida foram utilizadas as matrizes de pesos $a = [1 \ 0 \ 0]$ e $b = [.33 \ .33 \ .33]$.

3.4 Protótipo

Para a execução do modelo foi desenvolvido um protótipo, utilizando a linguagem de programação C++, bem como a biblioteca *OpenCV* para a aquisição e manipulação das imagens de entrada para o modelo.

O fluxo de execução do modelo para a detecção das características faciais pode ser descrito, de uma forma sucinta, pelas seguintes etapas de execução:

1. São utilizados algoritmos para obter imagens de entrada, estas podem ser tanto seqüências de imagens, como imagens obtidas a partir de webcams comuns ou vídeos;
2. As faces são detectadas nas imagens de entrada, e são identificadas as regiões de interesse para a detecção das características faciais;
3. Se ocorreu a detecção correta de uma face, passa-se para a próxima etapa, senão o fluxo do programa retorna para a primeira etapa;
4. Seleciona-se a utilização ou não do detector de íris;
5. A partir das regiões de interesse estimadas na etapa de detecção da face, é utilizado um algoritmo para identificação das íris dos olhos;
6. Nesta etapa é executado o modelo proposto para a detecção das características faciais referentes aos olhos e sobrancelhas.

A Figura 3.13 apresenta uma visão geral das etapas de execução do modelo de detecção de características faciais, sendo estes módulos expostos no decorrer desta Seção.

Na seção 3.4.1 é exposto o módulo desenvolvido para a aquisição de imagens (1 na Figura 3.13), e na Seção 3.1 é explicado o módulo de detecção da face (2 na Figura 3.13). A arquitetura do protótipo pode ser observada na Figura 3.13. Os demais módulos do protótipo (5 e 6 na Figura 3.13) foram explicados nas Seções 3.2 e 3.3, respectivamente.

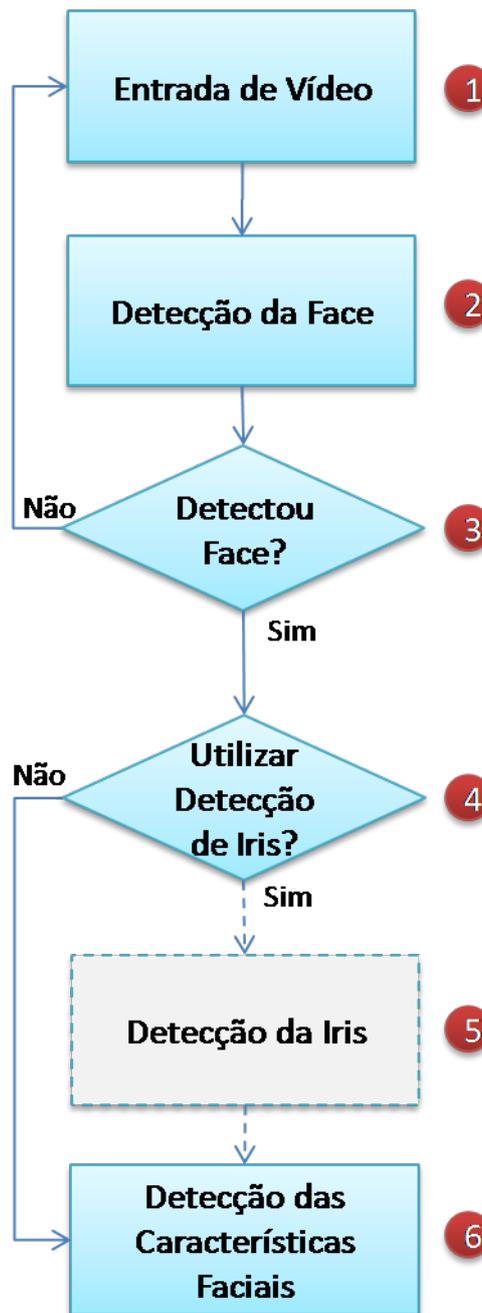


Figura 3.13: Arquitetura do modelo proposto para detecção de características faciais.

3.4.1 Entrada de Vídeo

Este módulo (1 da Figura 3.13) utiliza a biblioteca *OpenCV* para a aquisição de imagens. Essas podem ser oriundas de uma webcam ou de arquivos de vídeo. A estrutura do módulo de Entrada de Vídeo pode ser observada na Figura 3.14, e uma descrição do funcionamento deste módulo pode ser vista nos tópicos abaixo:

- Inicializar a captura de imagens: neste módulo ocorre a seleção do tipo de entrada de

dados, que será tratado pelo modelo. Estas imagens são provenientes de um dispositivo de vídeo, como uma webcam, ou podem ser adquiridas de vídeos previamente gravados.

- **Tratamento das Imagens Capturadas:** neste módulo são adquiridas as imagens do dispositivo de vídeo especificado no estágio anterior. Estas imagens são capturadas quadro a quadro, sendo gravadas em uma estrutura própria da biblioteca *OpenCV* para este fim. Esta função deve ser chamada ininterruptamente até o término do vídeo de entrada ou quando ocorrer uma interrupção da comunicação com a webcam.
- **Término do Programa:** Este módulo tem o objetivo de sinalizar aos demais módulos do modelo que a obtenção de imagens foi finalizada. Ele é executado sempre que ocorrem erros na captura de imagens, fim do vídeo de entrada ou por algum tipo de interrupção no fluxo do modelo.

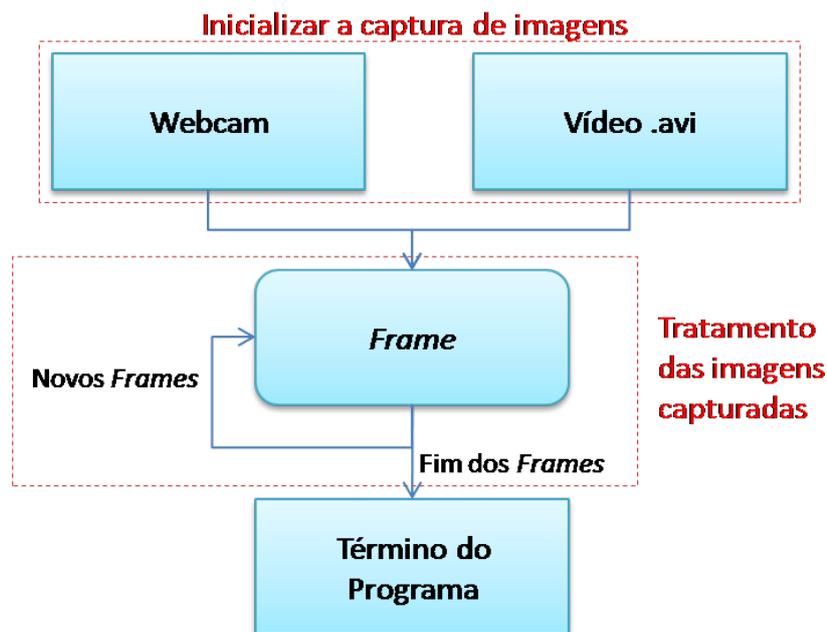


Figura 3.14: Estrutura do módulo de Entrada de Vídeo.

4. Resultados

Neste Capítulo são apresentados os resultados obtidos com o modelo proposto. São mostrados dois experimentos que utilizaram comparação dos dados obtidos com dados marcados por sujeitos e uma aplicação interativa que visa a animação de avatares 3D.

Inicialmente na Seção 4.1 serão mostrados os resultados obtidos a partir do modelo de detecção da Íris. Na Seção 4.2 são mostrados os resultados dos testes realizados em diferentes espaços de cor enquanto na Seção 4.3 são mostradas avaliações do modelo em comparação com outros modelos da literatura. Já na Seção 4.4 são mostrados alguns resultados obtidos na animação de olhos em avatares 3D.

4.1 Detecção da Íris

Para o modelo de Detecção da Íris foram realizados testes com as bases de dados BioID (1), IMM (38), Japanese Female (3) e Caltech (2), utilizando quatro *templates* diferentes. Estes *templates* são mostrados na Tabela 4.1, bem como os resultados obtidos a partir da execução de cada um deles. Nota-se que o melhor resultado é obtido a partir do *template* de número 3, o qual é formado por um círculo com uma borda branca ao seu redor, notando-se que a parte vermelha do *template* não é considerada no *matching*.

Tabela 4.1: Porcentagem das distâncias menores que 5 pixels, da posição encontrada da íris em relação a informação de *ground truth*.

Template	BioID(1)	IMM(38)	Japanese Female(3)	Caltech(2)
1 	62.5%	94.6%	99.5%	84.0194%
2 	80.2%	90.8%	100%	96.6102%
3 	81.5%	81.3%	99.5%	94.6731%
4 	75.4%	86.3%	99.5%	72.0698%

4.2 Diferentes Espaços de Cor

Para avaliar os resultados obtidos, foi medida a distância euclidiana d dos pontos detectados (x, y) , para os pontos tidos como reais na imagem (x_r, y_r) , de acordo com a Equação 4.1. Os pontos reais $Xr = (x_r, y_r)$ são localizados e determinados por uma pessoa que marca as localizações na imagem e armazena como valores corretos. Esta especificação é chamada de *Ground Truth*.

$$d = \sqrt{(x - x_r)^2 + (y - y_r)^2} \quad (4.1)$$

A localização por sujeitos destes pontos na imagem é muito imprecisa, devido à possibilidade destes componentes estarem situados em mais de um pixel da imagem, de acordo com a percepção da pessoa que está efetuando a localização dos pontos. Como pode ser visto na Figura 4.1, a localização exata do canto direito do olho direito pode estar situada em mais de um pixel, dependendo da resolução da imagem de entrada. Devido a este fato, foi estabelecido como erro aceitável d_{err} a distância de 5 pixel em uma face com raio de 240 pixels. Os resultados para a avaliação do canal de cor serão expressos também em termos de decis de d para que seja possível verificar o percentual de acertos para outros limiares de distância que não 5 pixels, visto que esse valor é arbitrário.

A Figura 4.1-a foi selecionada a partir da base de dados IMM (30), que consiste em uma base de dados pública de 40 pessoas distintas, em 6 posições diferentes, contendo informações de *Ground Truth*. Como a proposta do modelo é trabalhar com imagens coloridas, foram excluídas 3 pessoas desta base de dados, devido a estas estarem representadas em imagens em escala de cinza. Foram eliminadas também as imagens em que as pessoas não estavam na posição frontal, uma vez que o detector de face não foi treinado para localizar faces nestas posições. Após estas exclusões, resultaram 142 imagens de 37 pessoas diferentes.

Tabela 4.2: Comparação dos percentis de valores de d obtidos através da aplicação do modelo proposto em diferentes espaços de cor.

Espaço de Cor	d_1	d_2	d_3	d_4	d_5	d_6	d_7	d_8	d_9	d_{10}
RGB - <i>Red</i>	2	3	4	4	5	6	8	9	11	62
HSV - <i>Value</i>	2	3	4	4	5	6	7	9	12	62
Escala de Cinza	2	3	4	5	6	7	8	9	13	64
YCbCr - <i>Y</i>	2	3	4	5	6	7	8	8	13	64
RGB - <i>Green</i>	2	4	4	5	6	8	9	12	16	64
RGB - <i>Blue</i>	2	4	5	6	7	8	10	12	16	62
HSV - <i>Saturation</i>	3	4	6	8	10	12	16	18	21	65
YCbCr - <i>Cr</i>	3	4	5	6	7	10	15	41	55	81
YCbCr - <i>Cb</i>	12	17	21	24	28	32	38	46	56	81

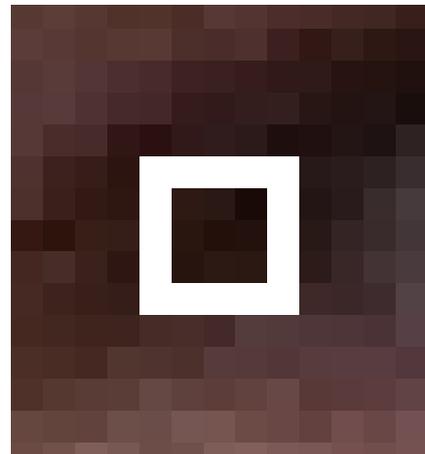
A Tabela 4.2, apresenta os decis de erros d , obtidos a partir da aplicação do modelo proposto no Capítulo 3, em vários canais de diferentes espaços de cor. Eles foram testados a fim de se verificar qual deles apresentaria o melhor resultado em termos de detecção das localizações de interesse. Como pode ser visto na Tabela 4.2, 90% dos valores de d são iguais ou menores que 11 pixels para o canal vermelho do espaço de cor RGB, 80% destas distâncias são iguais ou menores que 9 pixels neste mesmo canal, e assim por diante. Apesar da pequena diferença entre os resultados alcançados utilizando o canal vermelho do espaço de cor RGB e do canal V (*Value*) do espaço de cor HSV, foi escolhido o primeiro canal, porque não é necessário fazer transformações adicionais para obtê-lo



(a) Imagem original, mostrando faixa de pixels de erro aceita pelo modelo.



(b) Imagem do olho direito levemente ampliada, para salientar a região do canto direito do olho.



(c) Imagem ampliada da região do olho direito mostrando região onde se encontram possíveis cantos dos olhos.

Figura 4.1: Ampliação da região do olho, mostrando a dificuldade em especificar um único ponto para o canto do olho em uma imagem de resolução de 640x480. O quadrado branco tem 5 pixels de lado. Qualquer ponto no interior desse quadrado poderia ser considerado uma resposta correta para detecção do canto do olho.

e desta forma, não é agregado maior tempo computacional ao modelo. Como pode ser verificado na Tabela, a aplicação do modelo aos outros canais de cor apresentou resultados inferiores, em comparação com o RGB, portanto escolhido neste trabalho.

4.3 Avaliação do Modelo

Para avaliar a precisão do modelo proposto foi escolhida a base de dados pública IMM (30), com as mesmas imagens citadas na Seção 4.2. Esta base de dados contém 57 informações de *Ground Truth*. Estes pontos podem ser visto na Figura 4.2.



Figura 4.2: Imagem com informações de *Ground Truth* numeradas, referentes a base de dados IMM (30).

Foram encontrados na literatura duas implementações de modelos que identificam alguns componentes faciais que o modelo desenvolvido identifica. A primeira foi desenvolvida por Cristinacce e Coote (10) que utiliza basicamente seleção de *templates* e outra aplicação desenvolvida por Milborrow e Nicolls (28), que é uma implementação do modelo ASM.

Para a obtenção das informações dos pontos localizados pela aplicação de Cristinacce (10) foi necessário identificar manualmente os pontos localizados, marcando-os, um a um, a fim de gerar um arquivo de saída para comparação. Já para a aplicação disponibilizada

por Milborrow (28), os dados de saída do programa com as informações dos pontos, foram gravados diretamente em arquivo pela aplicação.

Com o objetivo de avaliar o desempenho do modelo proposto foram escolhidos os pontos em comum entre as três aplicações: modelo proposto, aplicação de Cristinacce (10) e aplicação de Milborrow (28). Estes pontos podem ser observados na Figura 4.3.



Figura 4.3: Imagem com os 8 pontos em comum entre as aplicações.

Foi efetuado o cálculo da medida euclidiana (mesmo explicado na Seção 4.2) para possibilitar a comparação dos modelos. Foram calculados a média dos erros e desvio padrão para os 8 pontos comuns entre os modelos, como pode ser observado na Figura 4.3. A Tabela 4.3 mostra as médias calculadas e a Tabela 4.4 os desvios padrões.

Tabela 4.3: Médias de erro obtida nas aplicações.

Aplicação	p_1	p_2	p_3	p_4	p_5	p_6	p_7	p_8
Modelo Proposto	3,36	6,83	7,42	3,30	20,43	11,25	13,35	23,82
Cristinacce (10)	6,28	7,23	6,90	5,65	12,16	7,37	8,90	9,15
Milborrow (28)	11,77	11,69	9,67	11,49	14,04	13,15	12,84	15,71

Com Base na Tabela 4.3 foram gerados dois gráficos, que podem ser visto na Figura 4.4

Tabela 4.4: Desvio padrão do erro comparado com a aplicação utilizada.

Aplicação	p_1	p_2	p_3	p_4	p_5	p_6	p_7	p_8
Modelo Proposto	2,12	2,54	3,65	2,04	13,19	11,18	12,39	12,63
Cristinacce (10)	4,44	4,27	3,70	3,55	6,18	3,82	5,77	5,09
Milborrow (28)	22,51	23,26	22,66	22,39	21,67	24,55	24,64	22,41

e 4.5. Como pode ser visto no gráfico da Figura 4.4, o modelo proposto apresentou melhores resultados nos componentes que representam os olhos. Já nos que dizem respeito a sobrancelhas, que são mostrados no gráfico da Figura 4.5, a aplicação de Cristinacce (10) apresentou melhores resultados.

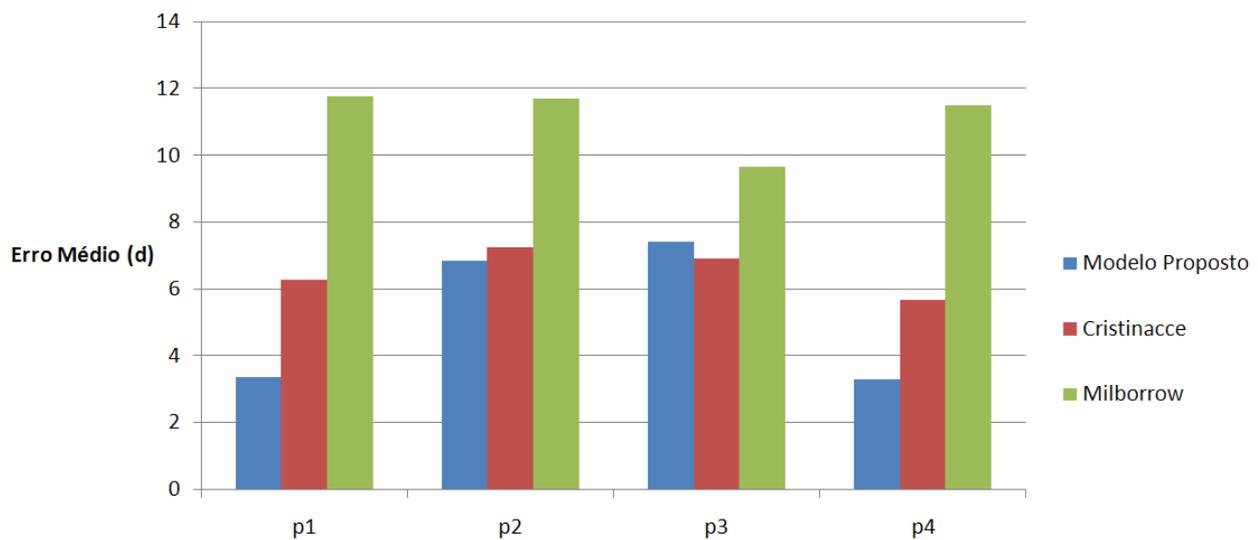


Figura 4.4: Gráfico contendo os 4 pontos referentes aos olhos analisados no modelo.

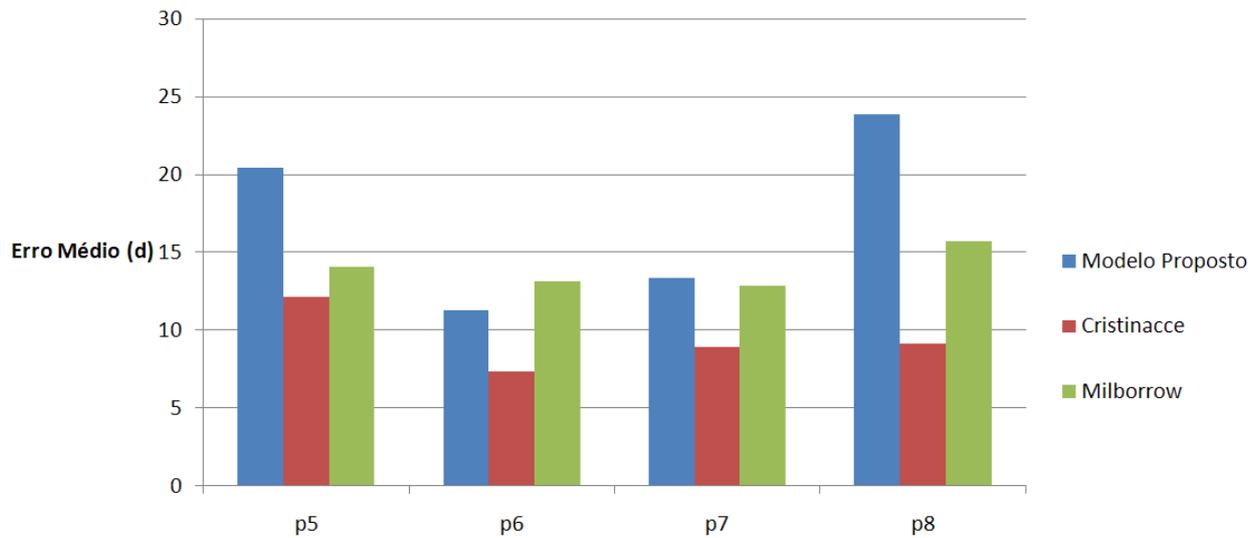
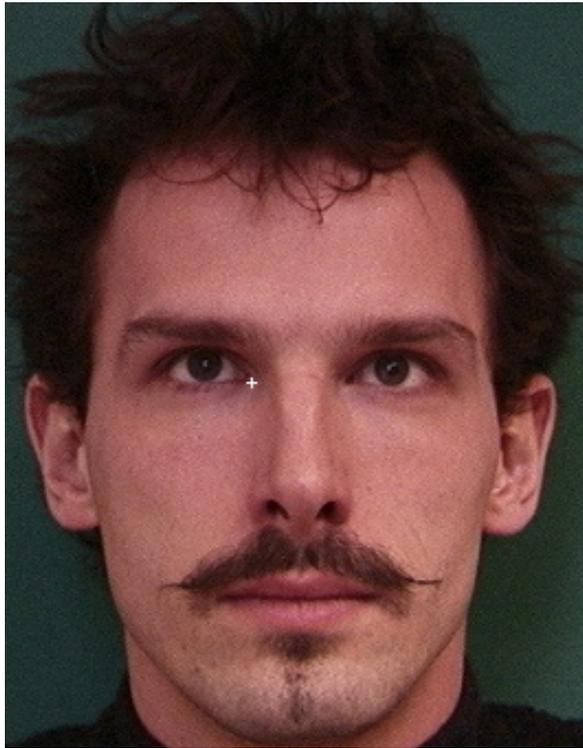


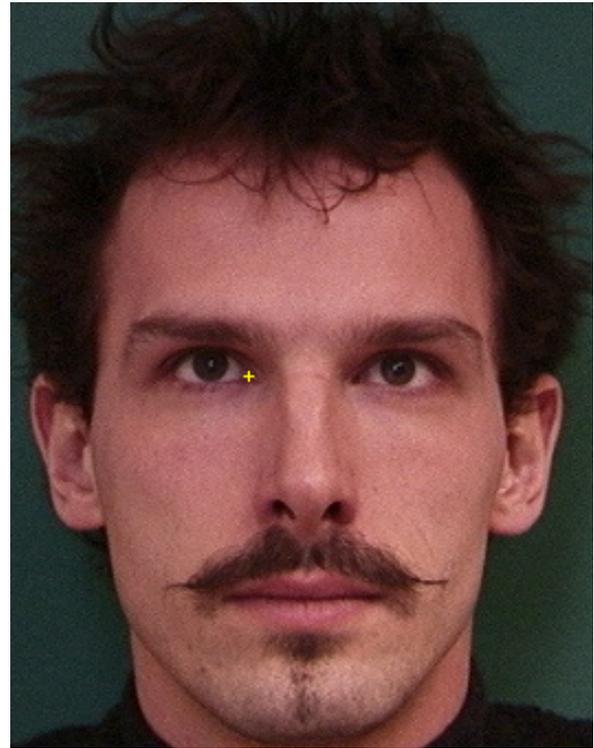
Figura 4.5: Gráfico contendo os 4 pontos referentes às sobrelhas analisados no modelo.

A partir destes resultados constatou-se que o modelo apresenta uma taxa de erro média entre 3 e 8 pixels para a detecção dos componentes relacionados ao olhos.

Um aspecto a observar é a dificuldade em visualmente detectar os dados a serem extraídos automaticamente dos vídeos. Por exemplo, na Figura 4.6, onde é comparado visualmente o ponto marcado como *Ground Truth* da base de dados IMM (Figura 4.6-a e o ponto localizado pela aplicação de Milborrow (28)(Figura 4.6-b. O erro entre as distâncias euclidianas da localização real da base de dados para a localização obtida pela aplicação foi de 3,6 pixels, nota-se, observando a imagem que tanto o ponto localizado como *Ground Truth* quanto o ponto localizado pela aplicação expressão corretamente a localização da componente desejada, diferem apenas na posição de alguns pixels, como pode ser visto na Figura 4.6-c.



(a) Ponto p_2 , marcado a partir do *Ground Truth*.



(b) Ponto p_2 identificado pela aplicação de Milborrow (28).



(c) Imagem ampliada da região do olho direito mostrando estes dois pontos marcados para salientar a distância entre eles.

Figura 4.6: Diferença entre as localizações de componentes faciais do *Ground Truth* da base de dados IMM para os pontos localizados na aplicação de Milborrow (28).

Alguns resultados dos testes realizados na base de dados IMM são mostrados na Figura 4.7, onde as imagens localizadas na coluna da esquerda referem-se a resultados do modelo proposto, as da coluna central são resultados da aplicação de Cristinacce (10) e as imagens

da coluna a direita são referentes a saída da aplicação de Milborrow (28):



continua...

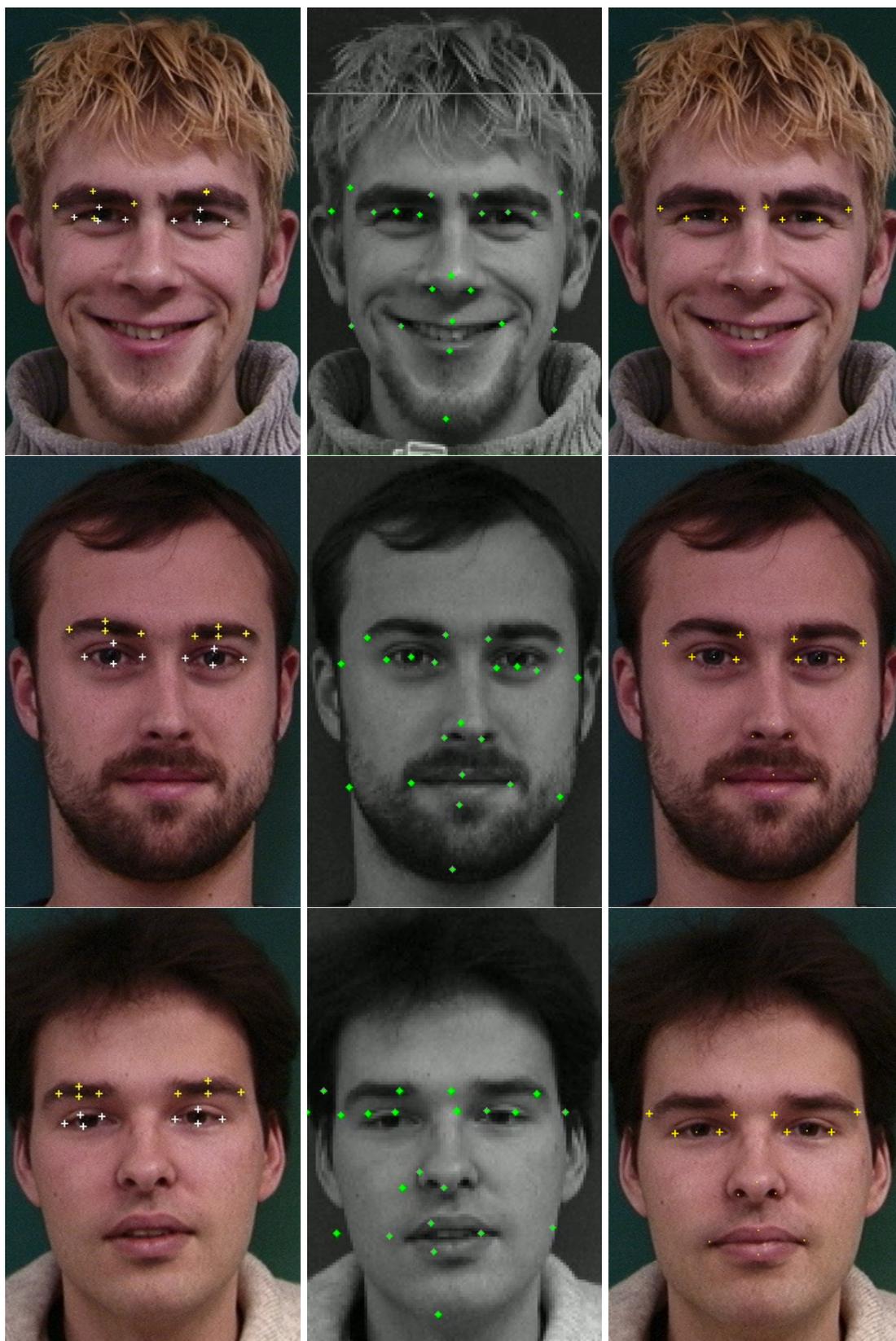


Figura 4.7: Exemplo de resultados para os três modelos testados, seguindo a ordem: as imagens da coluna da esquerda são referentes ao modelo proposto, central referem-se ao aplicativo Cristinacce (10) e as imagens da coluna da direita são resultados do aplicativo Milborrow (28)

Foi medido o desempenho dos três modelos, sendo calculado o tempo médio gasto para o processamento de cada quadro com 640x480 pixels nas imagens da base de dados IMM. O modelo proposto levou 37ms para calcular um quadro, Milborrow (28) teve um desempenho um pouco inferior, levando 66ms e Cristinacce (10) obteve o pior desempenho, levando 500ms para calcular um quadro.

4.4 Animação de Avatares 3D

Este experimento consiste em, a partir dos componentes faciais detectados referentes aos olhos direito e esquerdo, fazer o mapeamento, em relação ao tempo, dos pontos que representam cada olho para um avatar 3D. Foi definido um frame na imagem real para servir de referência para o modelo 3D do avatar. Como não há informações sobre a forma tridimensional dos pontos detectados nas imagens, neste teste foram feitas apenas variações bidimensionais no modelo, sendo que o usuário real deve estar a uma distância constante da câmera e deve ficar com a cabeça em posição frontal.

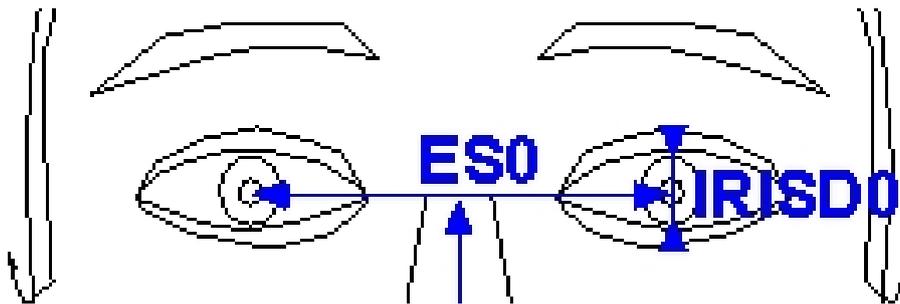


Figura 4.8: MPEG4 *Face Parameter Units* para os olhos. Estas unidades são utilizadas para mapear o deslocamento das pálpebras de um usuário real para um avatar (32).

Para variação de escala do modelo 3D, foram considerados os FPU (*Face Parameter Units*) (32) propostos no modelo MPEG4 padrão, mostrados na Figura 4.8. Nos experimentos foi solicitado ao usuário que ficasse com o rosto parado por 1 segundo no início da aplicação. Os primeiros 20 foram utilizados como quadros de treinamento, a fim de estabelecer um frame de referência para o modelo de animação. Nesta fase do treinamento, foi calculado o ESO e IRISDO, representados na Figura 4.8. Também foi estabelecida a posição média de cada ponto de referência da face, podendo ser chamados desta forma $\mu = \{(\bar{x}_1, \bar{y}_1), (\bar{x}_2, \bar{y}_2), \dots, (\bar{x}_8, \bar{y}_8)\}$.

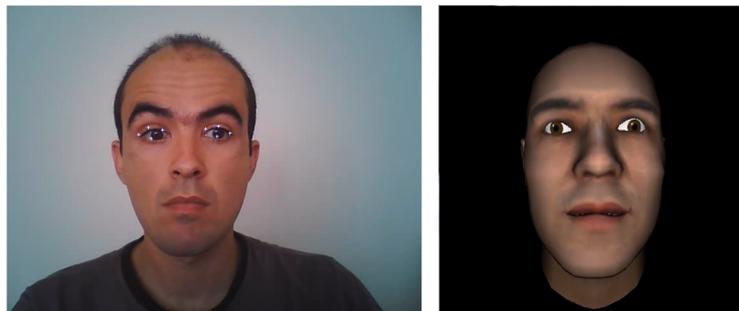
Os FAP's (*Facial Animation Parameters*) enviados para o modelo de animação facial, consistem em deslocamentos dos pontos de referência de μ e são calculados da seguinte forma:

$$Fap_i = \frac{\mu_i - X_i}{FAPU_{real}} \cdot FAPU_{avatar} \quad (4.2)$$

onde $FAPU_{real}$ são as *Facial Animation Units* medidas na fase de treinamento, e $FAPU_{avatar}$ são as *Facial Animation Units* do modelo do avatar 3D. As setas na Figura 4.8 indicam o sinal de cada Fap_i .

Os avatares utilizados neste experimento foram gerados com o pacote *FaceGen*¹, que preserva as proporções antropométricas de um rosto humano padrão. Assim sendo, o mapeamento dos pontos do usuário real para o avatar não necessitaram de nenhum ajuste adicional em funções de deformação. A Figura 4.9, mostra alguns resultados da animação do avatar 3D.

É importante salientar que não foram utilizadas informações das sobrancelhas para a animação do avatar 3D, foram utilizadas apenas as informações dos olhos.



(a) Resultado do mapeamento para os olhos abertos.



(b) Resultado do mapeamento para um olho piscando.



(c) Resultado do mapeamento para os olhos fechados.

Figura 4.9: Exemplo de animação de avatar 3D a partir de imagens reais.

¹<http://www.facegen.com/>

5. Considerações Finais

Nesta dissertação foi apresentado um modelo para detecção da íris, desenvolvido no âmbito de pesquisa do laboratório de pesquisa VHLAB , e foi apresentada uma abordagem para detecção de componentes faciais referentes a região dos olhos, baseada no modelo de cor vermelho invertido e relações antropométricas.

A partir dos resultados obtidos nos testes realizados no Capítulo 4, foi possível concluir que o modelo proposto identifica de uma forma correta os componentes faciais referentes aos olhos com um erro médio menor que o encontrado utilizando as aplicações de Cristinacce e Coote (10) e Milborrow e Nicolls (28), buscando as componentes referentes a olhos. Notou-se também a necessidade de trocar a abordagem para a busca por sobrancelhas, pois somente a execução do operador exponencial não foi suficiente para salientar as diferenças entre os pixels que representam as sobrancelhas.

Os erros na detecção da sobrancelha são agravados com a grande variação de iluminação que ocorre na base de dados IMM (30).

O modelo se distinguiu dos demais por seu desempenho computacional, as aplicações de Cristinacce (10) e Milborrow (28) tiveram um tempo de localização de componentes superior ao do modelo proposto, como pode ser observado na Seção 4.3 onde são expostos os tempos médios gastos por cada modelo para processar um quadro.

Pode-se também observar que o mapeamento descrito na Seção 4.4 expressou de uma forma correta a movimentação dos olhos no avatar 3D, apresentando animações que não condiziam com as faces reais, apenas quando algum componente não foi localizado na imagem.

O modelo proposto apresenta como principal contribuição o funcionamento em tempo real, bem como a sua precisão em comparação com os outros modelos e o fato de funcionar com uma simples webcam.

5.1 Trabalhos Futuros

Há algumas possibilidades de melhorias para o modelo desenvolvido. Uma delas poderia ser a utilização de máscaras deformáveis na Seção 2.1, para a localização dos componentes faciais, ao invés de fazer a localização dos componentes nos sentidos horizontal e vertical do *blob*.

Também poderia ser utilizada alguma técnica para minimizar a dependência do modelo quanto a variação de iluminação. Uma destas técnicas poderia ser a utilização do algoritmo Retinex (35) (foi testado no modelo de detecção da íris) para reduzir as variações de intensidade da imagem.

Uma proposta futura seria testar o modelo com um maior número de bases de dados de imagens públicas, e a comparação dos resultados nestas bases de dados ser feita com um maior número de modelos, inclusive comerciais como a FaceAPI¹.

¹www.seeingmachines.com/product/faceapi/

Referências

- [1] “The bioid face dataset”. Capturado em: <http://www.bioid.com/downloads/facedb/index.php>, Agosto 2008.
- [2] “The caltech faces dataset”. Capturado em: <http://www.vision.caltech.edu/html-files/archive.html>, Agosto 2008.
- [3] “Japanese female facial expression (jaffe) database”. Capturado em: <http://www.kasrl.org/jaffe.html>, Agosto 2008.
- [4] Bagherian, E.; Rahmat, R. W. O. “Facial feature extraction for face recognition: a review”. In *ITSim '08: International Symposium on Information Technology*, Aug. 2008, vol. 2, pp. 1–9.
- [5] Chang, T.; Huang, T.; Novak, C. “Facial feature extraction from color images”. In *Conference on Pattern Recognition, Conference B: Computer Vision & Image Processing Proceedings of the 12th IAPR International*, Oct 1994, vol. 2, pp. 39–43.
- [6] Cheddad, A.; Mohamad, D.; Manaf, A. A. “Exploiting voronoi diagram properties in face segmentation and feature extraction”. *Pattern Recogn.* 41, 12, 2008, 3842–3859.
- [7] Chen, Q.; Cham, W.-k.; Lee, K.-k. “Extracting eyebrow contour and chin contour for face recognition”. *Pattern Recogn.* 40, 8, 2007, 2292–2300.
- [8] Cootes, T.; Edwards, G. J.; Taylor, C. J. “Active appearance models”. In *European Conference on Computer Vision 1998*, 1998, vol. 2, Springer, pp. 484–498.
- [9] Cootes, T. F.; Taylor, C. J.; Cooper, D. H.; Graham, J. “Active shape models - their training and application”. *Computer Vision and Image Understanding* 61, 1995, 38–59.
- [10] Cristinacce, D.; Cootes, T. F. “Facial feature detection and tracking with automatic template selection”. In *FGR '06: Proceedings of the 7th International Conference on Automatic Face and Gesture Recognition*, Washington, DC, USA, 2006, IEEE Computer Society, pp. 429–434.
- [11] DeCarlo, D.; Metaxas, D. “Optical flow constraints on deformable models with applications to face tracking”. *International Journal of Computer Vision* 8, 2000, 99–127.
- [12] D’Orazio, T.; Leo, M.; Spagnolo, P.; Guaragnella, C. “A neural system for eye detection in a driver vigilance application”. In *The 7th International IEEE Conference on Intelligent Transportation Systems*, Oct. 2004, pp. 320–325.

-
- [13] Dornaika, F.; Davoine, F. "Head and facial animation tracking using appearance-adaptive models and particle filters". In *CVPRW '04: Conference on Computer Vision and Pattern Recognition Workshop*, 2004, pp. 153–153.
- [14] Dutreive, L.; Meyer, A.; Bouakaz, S. "Feature points based facial animation retargeting". In *VRST '08: Proceedings of the 2008 ACM symposium on Virtual reality software and technology*, New York, NY, USA, 2008, pp. 197–200.
- [15] Ekman, P.; Friesen, W. V.; Hager, J. C. "The facial action coding system", vol. 1. Weidenfeld & Nicolson, Texas, EUA, 2002.
- [16] Fasel, B.; Luetin, J. "Automatic facial expression analysis: a survey". *Pattern Recognition* 36, 2002, 259–275.
- [17] Haj, M. A.; Orozco, J.; Gonzalez, J.; Villanueva, J. J. "Automatic face and facial features initialization for robust and accurate tracking". In *19th International Conference on Pattern Recognition*, 2008, IEEE Computer Society, pp. 1–4.
- [18] He, K.; Wang, G.; Yang, Y. "Optical flow-based facial feature tracking using prior measurements". In *7th International Conference on Cognitive Informatics*, 2008, IEEE Computer Society, pp. 324–331.
- [19] Hsieh, C.-K.; Lai, S.-H.; Chen, Y.-C. "Expressional face image analysis with constrained optical flow". In *IEEE International Conference on Multimedia and Expo*, April 2008, pp. 1553–1556.
- [20] Hu, F.; Lin, Y.; Zou, B.; Zhang, M. "Individual 3d face generation based on candid-3 for face recognition". *Congress on Image and Signal Processing 1*, 2008, 646–650.
- [21] Jacques Jr., J. C. S.; Moreira, J. L.; Braun, A.; Musse, S. R.; Said, A. "A template-matching based method to perform iris detection in real-time using synthetic templates". In *ISM '09: Proceedings of the 2009 11th IEEE International Symposium on Multimedia*, Washington, DC, USA, 2009, IEEE Computer Society, pp. 142–147.
- [22] Ji, Q.; Yang, X. "Real time visual cues extraction for monitoring driver vigilance". In *ICVS '01: Proceedings of the Second International Workshop on Computer Vision Systems*, London, UK, 2001, Springer-Verlag, pp. 107–124.
- [23] Khan, A.; Ihtesham-UI-Islam. "A novel method for facial features extraction using bilateral filtering and wavelet transform". In *ICCIT '08. Third International Conference on Convergence and Hybrid Information Technology*, 2008, Nov. 2008, vol. 1, pp. 887–892.

-
- [24] Kourkoutis, L. G.; Panoulas, K. I.; Hadjileontia, L. “Automated iris and gaze detection using chrominance: application to human-computer interaction using a low resolution webcam”. In *19th IEEE Int. Conf. on Tools with Artificial Intelligence*, 2007, vol. 1, pp. 536–539.
- [25] Li, S. Z.; Lu, J. “Modeling bayesian estimation for deformable contours”. In *Proc. Seventh IEEE International Conference on Computer Vision*, Sept. 1999, vol. 2, pp. 991–996.
- [26] Lucas, B. D.; Kanade, T. “An iterative image registration technique with an application to stereo vision”. In *7th Intl Joint Conf on Artificial Intelligence 1981*, 1981, vol. 1, pp. 674–679.
- [27] Mahoor, M.; Abdel-Mottaleb, M. “Facial features extraction in color images using enhanced active shape model”. In *FGR 2006. 7th International Conference on Automatic Face and Gesture Recognition*, April 2006, p. 5.
- [28] Milborrow, S.; Nicolls, F. “Locating facial features with an extended active shape model”. *ECCV*, Capturado em: <http://www.milbo.users.sonic.net/stasm>, 2008.
- [29] Nixon, M.; Aguado, A. S. “Feature extraction and image processing”, 2nd ed. Academic Press, pages 424, 2008.
- [30] Nordstrøm, M. M.; Larsen, M.; Sierakowski, J.; Stegmann, M. B. “The IMM face database - an annotated dataset of 240 face images”. Tech. rep., Informatics and Mathematical Modelling, Technical University of Denmark, DTU, Richard Petersens Plads, Building 321, DK-2800 Kgs. Lyngby, may 2004.
- [31] Pandzic, I. S.; Forchheimer, R., Eds. “Mpeg-4 facial animation: The standard, implementation and applications”. John Wiley & Sons, Inc., New York, NY, USA, 2003.
- [32] Pandzic, R.; Fochheimer, R. “Mpeg-4 facial animation: The standard, implementation and applications”. Wiley, John and Sons, 2002.
- [33] Panning, A.; Al-Hamadi, A.; Niese, R.; Michaelis, B. “Facial expression recognition based on haar-like feature detection”. *Pattern Recognition and Image Analysis* 18, 3, Sept. 2008, 447–452.
- [34] Paragios, N.; Chen, Y.; Faugeras, O. “Handbook of mathematical models in computer vision”. Springer, 2006, ch. 15.
- [35] Park, Y. K.; Park, S. L.; Kim, J. K. “Retinex method based on adaptive smoothing for illumination invariant face recognition”. *Signal Process.* 88, 8, 2008, 1929–1945.

-
- [36] Ritter, G.; Wilson, J. "Handbook of computer vision algorithms in image algebra". CRC, Sept. 2000.
- [37] Sohail, A. S. M.; Bhattacharya, P. "Detection of facial feature points using anthropometric face model", vol. 31. Springer US, 2008, pp. 189–200.
- [38] Stegmann, M. B.; Ersbøll, B. K.; Larsen, R. "Fame - a flexible appearance modelling environment". *IEEE Trans. Med. Imaging* 22, 10, 2003, 1319–1331.
- [39] Stringa, L. "Eyes detection for face recognition". 365–382.
- [40] Tong, Y.; Wang, Y.; Zhu, Z.; Ji, Q. "Facial feature tracking using a multi-state hierarchical shape model under varying face pose and facial expression". *International Conference on Pattern Recognition* 1, 2006, 283–286.
- [41] Tong, Y.; Wang, Y.; Zhu, Z.; Ji, Q. "Robust facial feature tracking under varying face pose and facial expression". *Pattern Recognition* 1, 40, 2007, 3195–3208.
- [42] Viola, P.; Jones, M. "Robust real-time face detection". *International Journal of Computer Vision* 57, 2004, 137–154.
- [43] Vukadinovic, D.; Pantic, M. "Fully automatic facial feature point detection using gabor feature based boosted classifiers". In *Proc. IEEE International Conference on Systems, Man and Cybernetics*, Oct. 2005, vol. 2, pp. 1692–1698.
- [44] Wang, J. J.; Singh, S. "Video analysis of human dynamics - a survey". *Real-Time Imaging* 9, 2003, 321–346.
- [45] Wei, X.; Zhu, Z.; Yin, L.; Ji, Q. "A real time face tracking and animation system". In *Proc. Conference on Computer Vision and Pattern Recognition Workshop*, June, 2004, p. 71.
- [46] Wilson, P. I.; Fernandez, J. "Facial feature detection using haar classifiers". *J. Comput. Small Coll.* 21, 4, 2006, 127–133.
- [47] Wu, H.; Yokoyama, T.; Pramadihanto, D.; Yachida, M. "Face and facial feature extraction from color image". In *Proceedings of the Second International Conference on Automatic Face and Gesture Recognition*, Oct 1996, vol. 1, pp. 345–350.
- [48] Xu, C.; Zheng, Y.; Wang, Z. "Semantic feature extraction for accurate eye corner detection". In *Proc. 19th International Conference on Pattern Recognition ICPR 2008*, Dec. 2008, pp. 1–4.

- [49] Xue, Z.; Li, S.; Shen, D.; Teoh, E. K. "A novel bayesian shape model for facial feature extraction". In *ICARCV 2002. 7th International Conference on Control, Automation, Robotics and Vision*, Dec 2002, vol. 1, pp. 514–519.
- [50] Yen, G.; Nithianandan, N. "Facial feature extraction using genetic algorithm". In *CEC '02. Proceedings of the 2002 Congress on Evolutionary Computation*, 2002, vol. 2, pp. 1895–1900.
- [51] Yuille, A.; Cohen, D.; Hallinan, P. "Feature extraction from faces using deformable templates". In *Proceedings CVPR '89., IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 1989, vol. 1, pp. 104–109.
- [52] Zalewski, L.; Gong, S. "2d statistical models of facial expressions for realistic 3d avatar animation". In *Computer Vision and Pattern Recognition*, 2005, vol. 1, IEEE Computer Society, pp. 217–222.
- [53] Zuo, F.; de With, P. "Real-time facial feature extraction using statistical shape model and haar-wavelet based feature search". In *ICME '04. 2004 IEEE International Conference on Multimedia and Expo*, June 2004, vol. 2, pp. 1443–1446.