

CENTRO UNIVERSITÁRIO FEEVALE

LUIS ENRIQUE PETTINARI ACUÑA

DETECÇÃO FACIAL E SUAS CARACTERÍSTICAS

Novo Hamburgo, Novembro de 2006.

LUIS ENRIQUE PETTINARI ACUÑA

DETECÇÃO FACIAL E SUAS CARACTERÍSTICAS

Centro Universitário Feevale
Instituto de Ciência Exatas e Tecnológicas
Curso de Ciência da Computação
Trabalho de conclusão de curso

Professor Orientador: Prof. Marta R. Bez el Boukhari

Novo Hamburgo, Novembro de 2006.

A Regina e a Lorenzo

AGRADECIMENTOS

A Regina, minha esposa, pelo seu carinho e compreensão, e por ter me incentivado sempre a alcançar esta meta.

A meus fiéis torcedores, minha família (a de sangue), distante fisicamente, perto no coração.

A Marta, minha professora orientadora, por ter me brindado seu conhecimento, e por ter me apoiado nos momentos de desânimo. Em fim, a todos os professores que compartilharam comigo seu conhecimento e que tornaram possível este trabalho.

RESUMO

A face é um elemento fundamental na comunicação e interação entre os seres humanos. É, por este motivo, que existem, na atualidade, inúmeras pesquisas na área de análise de imagens de faces humanas, quais sejam: sistemas de reconhecimento facial, recuperação de imagens baseada em conteúdo, codificação de vídeo, vídeo conferência, vigilância de multidões, interfaces humano-computador inteligentes, etc. Na maioria delas tem-se como primeiro passo a detecção da face, quer dizer, identificar a presença de uma ou mais faces (se houver) na imagem, e determinar a localização e tamanho das mesmas. No problema específico da modelagem facial, devem-se também detectar as características faciais e seus pontos característicos.

O objetivo final deste trabalho é a construção de um sistema de detecção da face e suas características para modelagem 3D. Para tal foi realizado um estudo do padrão MPEG-4 e das diferentes técnicas utilizadas até hoje pelos pesquisadores, tanto para detecção da face, como para detecção das características da mesma. As mesmas técnicas podem ser usadas tanto para uma função como para a outra.

No sistema desenvolvido foi utilizada, para a detecção do rosto, a técnica da cor da pele, enquanto que a detecção das características faciais foi feita pela conjunção das técnicas: PCA, bordas e estrutural.

Palavras-chave: detecção de faces; detecção de características faciais; visão computacional; MPEG-4; PCA.

ABSTRACT

The human face is a fundamental element for the communication and the interaction between human beings. It is for this reason that today exist a number of researches in the human face image analysis, namely: face recognition systems, content-based image retrieval, video coding, video conferencing, crowd surveillance, intelligent human-computer interfaces, etc. The most of them have as their first step, the face detection, that is, to identify the presence of one or more faces in the image (if any), and to determine the localization and size of it. In the specific problem of facial modeling, it also must detect the facial features and its feature points.

The final target of this work is the construction of a system that detects a face and its features, aims to 3D modeling. For that, it was performed a research of the MPEG-4 standard and the different methods used, as for face detection, as for facial features detection. The same methods can be used for one or another function.

In the developed system it was used, for the face detection, the skin color method, whereas the face features detection was made by the conjunction of the methods: PCA, edges and structural.

Key words: face detection; facial feature detection; computer vision; MPEG-4, PCA.

LISTA DE FIGURAS

Figura 1-1 - Exemplo de faces com oclusão parcial por causa da iluminação.....	15
Figura 1-2 - Detecção da face dividido em abordagens.	18
Figura 2-1 - Pontos FAP e FDP.....	22
Figura 2-2 - Expressões básicas definidas por Ekman	24
Figura 2-3 - Definição das FAPUs	26
Figura 3-1 - Projeção horizontal e vertical de uma imagem.....	29
Figura 3-2 - Modelo abstrato da face imagem.....	30
Figura 3-3 - Exemplo de detecção da face pela cor da pele	33
Figura 3-4 - Aglomeração de cores no espaço de cores e aglomeração no espaço cromático.	34
Figura 3-5 – Processo de detecção da face pela cor da pele. De esquerda a direita, imagem original, pixels marcados como tendo cor de pele e por último, a imagem resultante após ter sido submetido a uma análise de componentes conectados.	34
Figura 3-6 - Contribuição à simetria dos gradientes nos pontos p_i e p_j	36
Figura 3-7 - (a) Imagem original; (b) Mapa de simetria isotrópica para um σ pequeno; (c) Binarização da imagem (b); (d) Imagens (a) e (c) sobrepostas; (e) Mapa de simetria isotrópica para um σ grande; (f) Binarização da imagem (e); (g) Imagens (a) e (f) sobrepostas; (h) Valores máximos de simetria radial marcados com uma cruz sobre a imagem original.	37
Figura 3-8 - Ilustração das sub-imagens relevantes para a transformada de Hough.	39
Figura 3-9 - Exemplo de características detectadas pelo método proposto por Nikolaidis.....	40
Figura 3-10 – Grade polar-logarítmica usada no algoritmo GAZE.....	41
Figura 3-11 - A grade polar-logarítmica aplicada sobre o olho direito de uma face na fase de treinamento.	42
Figura 3-12 - Representação gráfica do processo de busca sacádica. Os números representam os pontos de início, enquanto o X aponta a detecção do olho esquerdo.....	42
Figura 3-13 - Representação da face com os descritores genéricos e as respectivas relações espaciais.....	43
Figura 3-14 - Processo de detecção de uma face.....	43
Figura 3-15 - Pontos de interesse detectados usando o método de Förstner. Cruzes indicam padrões do tipo canto, enquanto que os círculos indicam padrões do tipo circular.	45
Figura 3-16 - Exemplo de um conjunto de padrões obtidos usando aglomeração por k-means de pequenas áreas da imagem.....	45
Figura 3-17 - O modelo da face e os grupos de componentes da face.	46
Figura 3-18 - Os modelos de características faciais.	47
Figura 3-19 - Processo de seleção de característica prévio.	47
Figura 3-20 - Molde da face de 14 x 16 pixels, composto por 16 regiões (blocos cinza) e 23 relações (setas).....	49
Figura 3-21 - Exemplo de extração do contorno da face.....	50

Figura 4-1 - Imagem média de um conjunto de 115 imagens (esquerda) e desvio de uma imagem do conjunto com a imagem média.	53
Figura 4-2 - Imagem de uma face com as características a serem localizadas marcadas em branco.	55
Figura 4-3 - Imagens originais (esquerda), PCA (centro) e ICA (direita).....	56
Figura 4-4 - (a) Modelo "canônico" da face, máscara para eliminar pixels próximos das bordas e modelo da face mascarado. (b) Macro-processo do algoritmo de detecção e (c) O algoritmo mais detalhado.....	58
Figura 4-5 - Na fila de cima, de esquerda a direita: Exemplos de padrões de faces; distribuição dos exemplos de faces; aproximação com aglomerados Gaussianos; centroides de cada uma das classes de padrões de faces. Na fila de baixo, idem para as classes de padrões de não faces.	59
Figura 4-6 - Exemplo de uma imagem com aparência de face.	60
Figura 4-7 - Algoritmo básico para detecção da face.	62
Figura 5-1 - Representação gráfica das amostras utilizadas na modelagem da cor da pele.	65
Figura 5-2 - Representação gráfica do modelo Gaussiano da cor da pele.....	66
Figura 5-3 - a) imagem original; b) detecção dos pixels com cor semelhante à cor da pele; c) limiarização da imagem (b).	66
Figura 5-4 - a) Imagem limiarizada; b) Perfil horizontal; c) Perfil vertical.	67
Figura 5-5 - Exemplo de detecção do rosto.	68
Figura 5-6 - Detectores de bordas de Sobel; a) sem limiarizar; b) limiarizado; c) com a imagem previamente equalizada por histograma, sem limiarizar; d) idem (c) mas limiarizado; e) imagem filtrada pelo filtro gaussiano; f) detector aplicado sobre a imagem filtrada (e).	69
Figura 5-7 – a) Detector de bordas de Robert Cross; b) Idem (a), com a imagem previamente equalizada pelo histograma; c) Detector de bordas de Canny; d) Idem (c) , com a imagem previamente equalizada pelo histograma.....	70
Figura 5-8 - Exemplos de amostras utilizadas na determinação do espaço vetorial do olho direito.....	71
Figura 5-9 - Imagem média das amostras capturadas para o olho direito.	71
Figura 5-10 – Representação dos <i>eigeneyes</i> do olho direito.	72
Figura 5-11 - Detalhe do resultado da detecção dos olhos.....	73
Figura 5-12 - Exemplos de amostras utilizadas na determinação do espaço vetorial do nariz.....	73
Figura 5-13 - Imagem média das amostras capturadas para o nariz.....	73
Figura 5-14 - Representação do <i>eigennozes</i> do nariz.	73
Figura 5-15 - Detalhe do resultado da detecção do nariz.	74
Figura 5-16 – a) Região de interesse da para detecção da boca; b)Detecção de bordas horizontais na região da boca; c) Perfil horizontal da imagem das bordas; d) Perfil vertical da imagem das bordas.	74
Figura 5-17 - Detalhe do resultado da detecção da boca.	75
Figura 5-18 - Detalhe da detecção dos pontos FDPs do nariz.....	76
Figura 5-19 - Detalhe da detecção dos pontos FDPs dos olhos.	77
Figura 5-20 - Detalhe da detecção dos pontos FDPs da boca.	78
Figura 5-21 – Regiões de busca da primeira implementação; a) Região de busca do olho direito; b) Região de busca do olho esquerdo; c) Região de busca do nariz; d) Região de busca da boca.....	81
Figura 5-22 – Regiões de busca da segunda implementação; a) Região de busca do nariz; b) Região de busca do olho direito; c) Região de busca do olho esquerdo; d) Região de busca da boca.....	83
Figura 5-23 - Fluxograma do processo de detecção da face e suas características.	83
Figura 5-24 - a) Representação gráfica do conjunto de amostras original; b) Representação gráfica do novo conjunto de amostras.	84

Figura 5-25 - a)Representação gráfica do modelo da cor da pele original; b)Representação gráfica do novo modelo da cor da pele.....	85
Figura 5-26 - Exemplos de resultados obtidos.	87
Figura 5-27 - exemplos de resultados obtidos com casos especiais.....	87

LISTA DE TABELAS

Tabela 2-1 - Grupos de FAPs	23
Tabela 2-2 - Visemas definidos no MPEG-4	23
Tabela 2-3 - Ações para cada expressão.....	24
Tabela 2-4 - Definição das FAPUs.....	26

LISTA DE ABREVIATURAS

3D	Tridimensional
AHT	Adaptive Hough Transform
DFFS	Distance From Feature Space
EM	Expectation Maximization
FA	Factor Analysis
FAP	Facial Animation Parameter
FAPU	Facial Animation Parameter Unit
FDP	Facial Definition Parameter
FIG	Facial Interpolation Graph
FIT	Facial Interpolation Table
GSL	GNU Scientific Library
ICA	Independent Components Analysis
MDL	Minimum Description Length
MPEG	Moving Picture Expert Group
NN	Neural Network
PCA	Principal Components Analysis
RPS	Refused Point Step

SUMÁRIO

Resumo	3
Abstract.....	6
Lista de Figuras	7
Lista de Tabelas	10
Lista de Abreviaturas.....	11
Sumário.....	12
1 Introdução.....	14
1.1 Objetivos Gerais	14
1.2 Objetivos Específicos	14
1.3 Justificativa.....	14
2 O Padrão MPEG-4.....	20
2.1 O que é o padrão MPEG-4	20
2.2 Parâmetros FDP.....	20
2.3 Parâmetros FAP.....	23
2.4 Unidades FAPU	25
2.5 Tabela FIT	26
3 Métodos baseados em características	28
3.1 Características de baixo nível.....	28
3.1.1 Bordas	28
3.1.2 Informações de níveis de cinza.....	29
3.1.3 Cor e textura	31
3.1.4 Outras características.....	34
3.2 Análise de características.....	37
3.2.1 Abordagem estrutural	37
3.2.2 Filtros de Gabor.....	40
3.2.3 Constelações	42
3.2.4 Modelos probabilísticos.....	46
3.3 Casamento de padrões	48
3.3.1 Padrões pré-definidos	49
3.3.2 Contornos ativos.....	49
4 Métodos baseados na aparência.....	52
4.1 Métodos de sub-espacos lineais.....	52
4.2 Métodos baseados em distribuição	57
4.3 Métodos baseados em discriminantes.....	61

5 Sistema Proposto	63
5.1 Detecção do rosto	64
5.1.1 Detecção da cor da pele	64
5.1.2 Determinação dos limites do rosto	66
5.2 Detecção dos olhos	68
5.2.1 PCA	70
5.3 Detecção do nariz	73
5.4 Detecção da boca	74
5.5 Detecção dos pontos FDPs	75
5.5.1 Pontos FDPs do nariz	75
5.5.2 Pontos FDPs do olho	76
5.5.3 Pontos FDPs da boca	77
5.6 O arquivo de configuração	78
5.7 Implementação.....	79
5.8 Testes	84
6 Conclusões.....	88
Referências Bibliográficas.....	90
Anexos.....	93
Anexo A Descrição dos pontos FDP.....	94
Anexo B Descrição dos pontos FAP.....	97
Anexo C Arquivo de configurações.....	101

1 INTRODUÇÃO

Neste capítulo serão apresentados os objetivos deste trabalho de conclusão de curso, uma justificativa para o desenvolvimento do mesmo, bem como apresentação geral das técnicas a serem empregadas.

1.1 Objetivos Gerais

O objetivo geral deste trabalho é desenvolver, com base no levantamento bibliográfico desenvolvido no TCI, um sistema para a detecção da face em uma imagem, e de suas características, demarcando na mesma os pontos FDPs do padrão MPEG4.

1.2 Objetivos Específicos

Os objetivos específicos deste trabalho são:

- Fazer um levantamento teórico sobre o padrão MPEG-4.
- Fazer um levantamento teórico sobre as principais técnicas de detecção da face.
- Fazer um levantamento teórico sobre as principais técnicas de extração de características faciais (olhos, nariz e boca).
- Desenvolver o módulo de detecção facial.
- Desenvolver o módulo de detecção dos olhos em uma face.
- Desenvolver o módulo de detecção do nariz em uma face.
- Desenvolver o módulo de detecção da boca em uma face.
- Desenvolver o módulo de detecção dos principais pontos FDPs na face.
- Apresentar os resultados dos módulos desenvolvidos.

1.3 Justificativa

Os seres humanos têm várias formas para comunicar-se, utilizando geralmente, em situações normais, mais de uma ao mesmo tempo. Não são só a boca e o ouvido, senão todo o corpo. Tirando a própria fala, poderia se dizer que a face fornece um maior conteúdo

informativo ao seu interlocutor; pois ela permite, entre outras coisas, perceber o estado de animo do interlocutor, a reação dele às nossas palavras, ou até mesmo, o simples fato de ver com quem se está falando, deixa à pessoa mais à vontade. Imagine uma interface homem-máquina que reconhecesse as expressões do usuário! Certamente seria uma grande fonte de informação sobre o que ele está sentindo e lhe permitiria tomar decisões ao respeito.

Para o ser humano o reconhecimento de uma face é uma tarefa trivial, já com poucos meses de vida é capaz de reconhecer as pessoas de seu entorno. Basta uma pessoa olhar para uma imagem para, quase instantaneamente, identificar se nela tem uma face, e reconhecê-la se for o caso. E isto sob as mais adversas condições (Figura 1-1).

Poder-se-ia pensar, então, que uma tarefa tão simples e rápida para o ser humano, não deveria apresentar dificuldade para os computadores modernos, com um poder de cálculo imensamente superior à mente humana. O problema é que o poder do computador está baseado em cálculos concretos, enquanto que a mente humana possui um poder de abstração enorme, ainda não alcançado por qualquer técnica computacional. A face humana é um objeto dinâmico e tem um alto grau de variabilidade na sua aparência, o que dificulta a sua detecção por meio das técnicas convencionais de casamento de padrões. Outros fatores que dificultam a tarefa são: variações de luminosidade da imagem, movimentos da cabeça, distância da câmera, existência de pelo facial, óculos, objetos ocultando parcial ou totalmente a face, diferentes expressões do rosto; riso, raiva, tristeza, olhos fechados, boca aberta, etc.



Figura 1-1 - Exemplo de faces com oclusão parcial por causa da iluminação.

Fonte: (<http://www.thesahara.net/paul.htm>).

A detecção da face é o primeiro passo em inúmeras aplicações tais como sistemas de reconhecimento facial, recuperação de imagens baseada em conteúdo, codificação de vídeo,

vídeo conferência, vigilância de multidões, e interfaces humano-computador inteligentes (HJELMÅS et al., 2001). Uma definição geral do problema de detecção de faces pode ser: dada uma imagem estática, ou de vídeo, detectar e localizar um número desconhecido (se houver) de faces. A solução do problema envolve segmentação e extração de faces, assim como das suas características faciais, de um fundo não controlado, independentemente da iluminação, orientação, ou distância à câmera.

A detecção de faces é uma área da visão computacional, a qual tem como principal objetivo a investigação de métodos automáticos para extração de informações contidas em imagens (CAMPOS, 2001).

Há décadas a ficção científica vem fantasiando com aparelhos de comunicação que permitem o contato visual entre os interlocutores. Naturalmente, os científicos têm procurado tornar a fantasia em realidade, e já desde a aparição do televisor tem apresentado protótipos de “vídeo-telefones”, porém não com a praticidade necessária, nem a um preço adequado. No entanto, a chegada massiva do computador aos lares, a aparição de câmeras de vídeo de estado sólido cada vez menores, e a alta integração de componentes eletrônicos, já tem tornado esta tecnologia possível. Pelo computador, através da internet, já é possível comunicar-se visualmente com um custo reduzido. Por outro lado, os aparelhos de telefonia celular já estão tecnicamente aptos para uma comunicação deste tipo, o que permitiria a sua massificação. Porém, a tecnologia atual de transmissão digital de imagens, consome uma largura de banda tal que se tornaria impraticável se massificada. A solução então é buscar novas tecnologias de codificação de vídeo que possuam um baixo consumo de banda.

Neste sentido, foi desenvolvida a norma MPEG-4 (ISO/IEC JTC1/SC29/WG11, 1998). Esta norma define a forma de descrever uma cena tridimensional de forma que possa ser reproduzida num receptor. Dentre os objetos descritos, recebe especial atenção, a face humana. Assim, o MPEG-4 define três tipos de dados faciais (ISO/IEC JTC1/SC29/WG11, 1998): Facial Definition Parameters (FDP); Facial Animation Parameters (FAP); e Facial Interpolation Parameters (FIT).

Estas informações, obtidas no padrão MPEG-4, poderão ser utilizadas para, no outro extremo da linha, reconstruir, ou melhor, sintetizar a imagem. Para isto existem hoje várias ferramentas, dentre as quais podem-se ressaltar a linguagem VRML (Virtual Reality Modeling Language) (ABÁSULO, 1998) e a biblioteca Java3D (SUN, 1999).

A extração de características da face pode ser usada não só na modelagem e transmissão de rostos, como também no reconhecimento de faces. A literatura é ampla para

estes dois tópicos, e entre os numerosos trabalhos podem-se diferenciar os métodos baseados em imagens estáticas e os baseados em seqüência de imagens. No entanto, pode se dizer que, em linhas gerais, estes últimos utilizam os primeiros, para a detecção inicial da face e suas características, utilizando a seguir técnicas de tracking¹, computacionalmente mais baratas, para conseguir a extração de características em tempo real.

Em (POPOVICI, 2004) pode-se encontrar uma relação das mais importantes técnicas utilizadas na extração de faces e suas características, classificadas pelo tipo de abordagem. Estas técnicas têm sido usadas tanto na extração da face, como na extração das suas características (boca, olhos, nariz, etc.), assim como também no tracking.

- Métodos baseados em características
 - Características de baixo nível: bordas, informação de níveis de cinza, cor e textura;
 - Análise de características: Abordagem estrutural, Filtros de Gabor, Constelações, Modelos probabilísticos;
- Casamento de padrões;
 - Padrões pré-definidos;
 - Contornos ativos: Snakes (cobras), Deformable Templates (padrões deformáveis), Smart Snakes (cobras inteligentes), Point Distribution Models (Modelos de Distribuição de Pontos);
- Métodos baseados na aparência;
 - Métodos de sub-espacos lineares: PCA, ICA, FA;
 - Métodos baseados em distribuição;
 - Métodos baseados em discriminante: NN, SVM, AdaBoost;
 - Métodos baseados na teoria da informação;

Hjelmås (2001) apresenta uma divisão das abordagens um pouco diferente, conforme mostrado na Figura 1-2.

¹ Seguimento de um determinado objeto através de uma seqüência de imagens

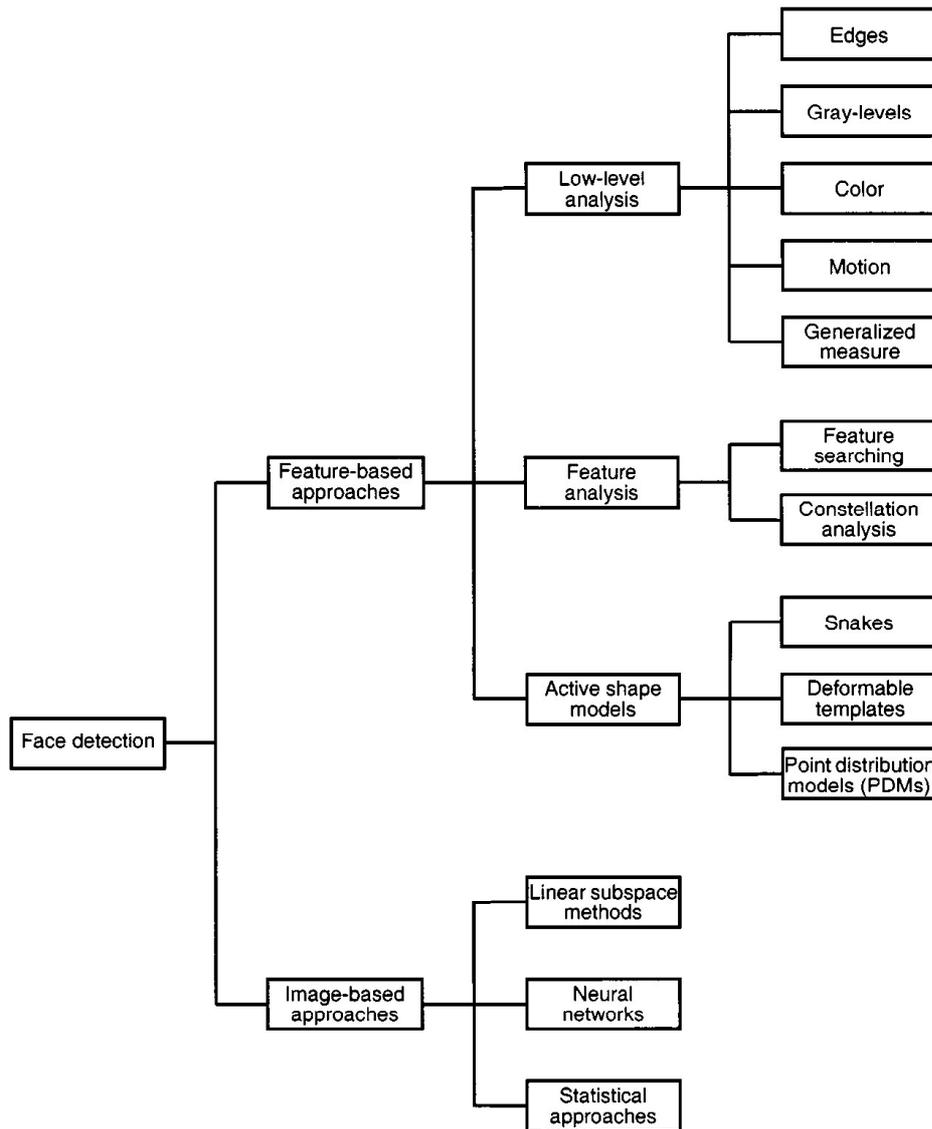


Figura 1-2 - Detecção da face dividido em abordagens.

Fonte: (HJELMÅS et al., 2001)

Para facilitar a apresentação dos métodos neste trabalho, optou-se por uma divisão mais simplificada, dividindo-se em duas principais abordagens: baseadas em características e baseadas em aparência. As técnicas baseadas em características fazem uso explícito do conhecimento da face e seguem a metodologia clássica de detecção, na qual as características de baixo nível são extraídas primeiro, para depois analisá-las em base ao conhecimento que se tem da face. Por outro lado, as técnicas baseadas na aparência visam detectar a face como um problema de detecção geral. Estas técnicas utilizam o conhecimento da face de forma implícita, o qual é aprendido a partir de um conjunto de imagens de treinamento. A detecção é feita diretamente, a partir da imagem de entrada, sem necessidade de extrair características, nem analisá-las (HJELMÅS et al., 2001).

Neste trabalho será utilizada a técnica da cor da pele para detecção da face, e as técnicas de PCA, bordas e estrutural para as características faciais.

A estrutura do trabalho é a seguinte: no segundo capítulo é abordado o padrão MPEG-4, em particular a parte que trata de faces humanas. São apresentados os pontos de controle para a calibração das faces, bem como para animação das mesmas. No terceiro capítulo são apresentadas algumas das técnicas mais utilizadas baseadas em características, enquanto que no quarto capítulo são apresentadas as técnicas baseadas na aparência. No quinto capítulo é apresentado o sistema proposto, sendo discutidas as técnicas implementadas para cada detecção efetuada. A técnica PCA é apresentada em detalhes, por ter sido a implementada na maior parte da detecção. Apresenta-se, também, neste capítulo as técnicas utilizadas para a detecção dos pontos FDPs das imagens, bem como testes executados com duas bases, uma extraída da internet e outra de imagens obtidas no Centro Universitário Feevale. Em seguida são apresentadas as conclusões e indicações de continuação deste trabalho.

2 O PADRÃO MPEG-4

2.1 O que é o padrão MPEG-4

O grupo MPEG (Moving Picture Experts Group) foi criado com o objetivo de desenvolver padrões de compressão de dados para armazenamento e distribuição de seqüência de imagens. Os padrões MPEG -1 e MPEG -2, desenvolvidos por este grupo, e amplamente utilizados na atualidade, são baseados em *frames*, quer dizer que as imagens correspondentes a cada *frame* são tratadas como uma unidade. Já o padrão MPEG-4 (ISSO/IEC JTC1/SC29/WG11, 1998), atendendo à crescente necessidade de integração entre imagens naturais (captadas por uma câmera) e sintéticas (geradas por computador), acrescenta uma abordagem baseada em objetos. Isto é, a cena é descrita pelos diferentes objetos que a compõem e as alterações que estes sofrem no decorrer do tempo. Por exemplo, a cena poderia ser composta por um carro em movimento, um fundo e o som do motor do carro, sendo que cada um destes seria um objeto independente.

Esta abordagem, além do objetivo principal, que é a integração entre imagens naturais e sintéticas, possibilita também um ganho enorme em consumo de banda, assim como escalabilidade, acesso universal, proteção, entre outras. Note-se que a cena, ao ser descrita em forma de objetos, pode ser sintetizada e apresentada ao usuário desde diferentes pontos de vista .

O padrão define objetos para modelar a figura humana, sendo um objeto para a cabeça e outro para o resto do corpo. Na definição do objeto *Face* são especificados quatro conjuntos de parâmetros: FDP, FAP, FAPU e FIT, os quais serão detalhados na seguintes seções.

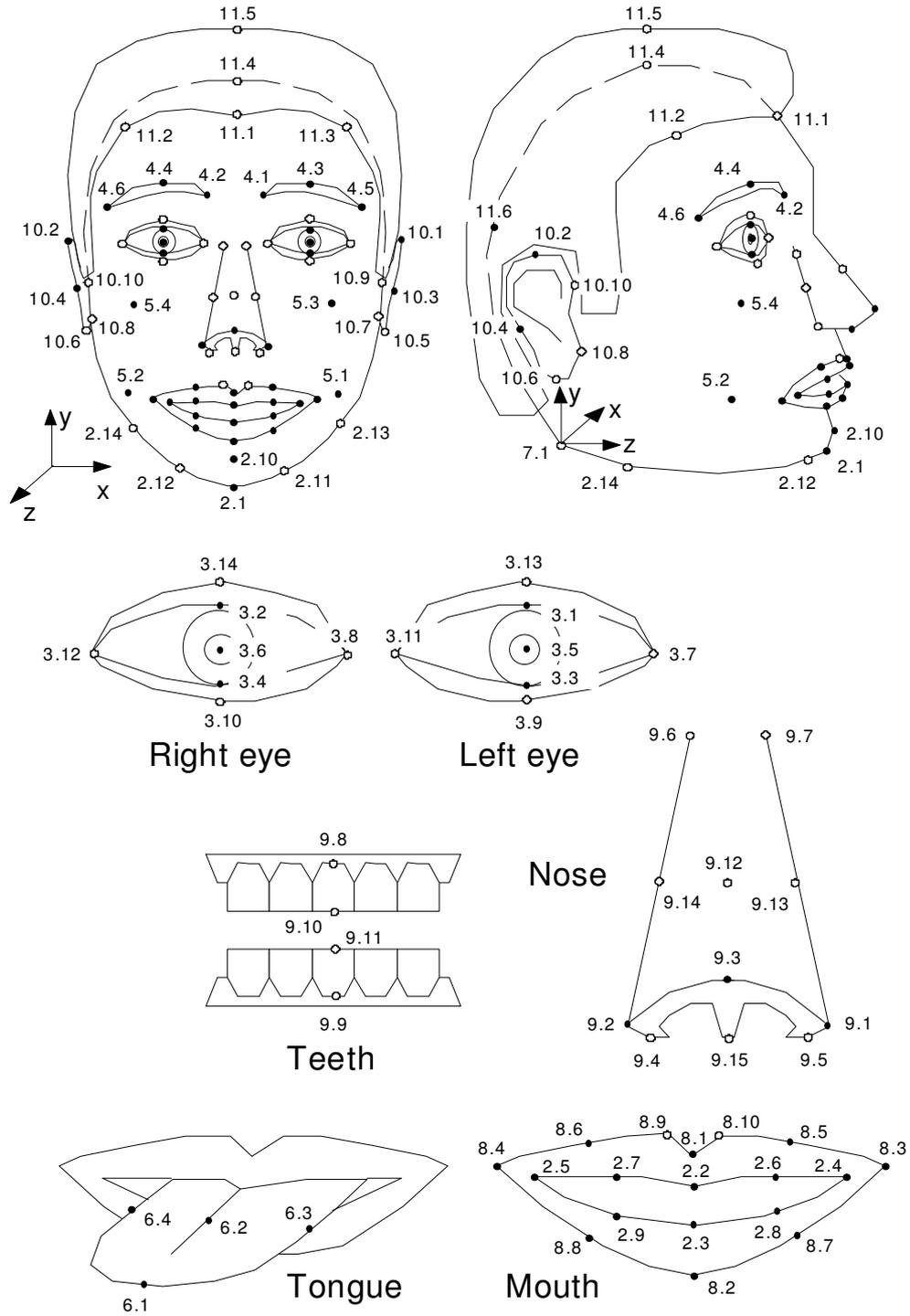
2.2 Parâmetros FDP

O padrão MPEG-4 define certos pontos característicos da face humana chamados de FDP (Facial Definition Parameter). Estes pontos permitem mudar a aparência da face sintetizada. Quer dizer, são os pontos que, aplicados num modelo de face 3D genérico,

permitem reproduzir a face original no receptor. O padrão define 84 pontos FDPs, sendo 18 na face, 18 na boca, 4 na língua, 4 nos dentes, 14 nos olhos, 11 no nariz, 6 nas orelhas, 6 nas sobrancelhas e 3 no cabelo. Na Figura 2-1 são apresentados os 84 pontos FDPs e no Anexo A encontra-se uma lista completa dos FDPs definidos no padrão, indicando a localização, descrição e simetria dos mesmos.

A obtenção destes pontos deve ser feita numa face neutra, que o padrão define como:

- cabeça reta;
- todos os músculos da face relaxados;
- olhar em direção ao eixo Z;
- pálpebras tangentes à íris;
- o diâmetro da pupila é um terço do diâmetro da íris;
- lábios em contato;
- a linha do interior dos lábios é horizontal e na mesma altura que os cantos da boca;
- a boca está fechada e os dentes superiores tocando os inferiores;
- a língua plana dentro da boca, horizontal, com a ponta tocando o limite entre os dentes superiores e inferiores



- Feature points affected by FAPs
- Other feature points

Figura 2-1 - Pontos FAP e FDP.
 Fonte: (ISO/IEC JTC1/SC29/WG11, 1998)

2.3 Parâmetros FAP

Os pontos FAPs (Facial Animation Parameter) são um subconjunto dos FDPs, que além de utilizados na definição da face, são utilizados também na animação da mesma. Estes permitem reproduzir no receptor os movimentos da maior parte dos componentes do rosto original. No padrão MPEG-4 são definidos 68 FAPs, sendo 2 de alto nível, visemas e expressões, e 66 de baixo nível, os quais determinam ações sobre pontos específicos da face.

Tabela 2-1 - Grupos de FAPs

Facial Animation parameters (FAPs)		
Grupo	Descrição	Nº de parâmetros
1	Visemas e expressões	2
2	Mandíbula, queixo, posições internas do lábio inferior, cantos da boca, lábios médios	16
3	Globos oculares, pupilas, pálpebras	12
4	Sobrancelha	8
5	Bochechas	4
6	Língua	5
7	Rotação da cabeça	3
8	Posições externas do lábio	10
9	Nariz	4
10	Orelhas	4

Fonte: (ISO/IEC JTC1/SC29/WG11, 1998)

Os FAPs estão divididos em 10 grupos (Tabela 2-1). No primeiro grupo encontram-se os visemas e as expressões. Os visemas (Tabela 2-2) são a representação gráfica dos fonemas e foram pensados para facilitar a animação da fala.

Tabela 2-2 - Visemas definidos no MPEG-4

Visema	Fonemas	Exemplo ²
0	Nenhum	Na
1	p, b, m	<u>put</u> , <u>bed</u> , <u>mill</u>
2	f, v	<u>far</u> , <u>voice</u>
3	T,D	<u>think</u> , <u>that</u>
4	t, d	<u>tip</u> , <u>doll</u>
5	k, g	<u>call</u> , <u>gas</u>
6	tS, dZ, S	<u>chair</u> , <u>join</u> , <u>she</u>

² Os exemplos de fonemas estão em inglês

Visema	Fonemas	Exemplo ²
7	s, z	<u>s</u> ir, <u>z</u> eal
8	n, l	<u>l</u> ot, <u>n</u> ot
9	R	<u>r</u> ed
10	A:	<u>c</u> ar
11	E	b <u>e</u> d
12	I	<u>t</u> ip
13	Q	top
14	U	<u>b</u> ook

Fonte: (ISO/IEC JTC1/SC29/WG11, 1998)

O FAP das expressões permite reproduzir no receptor, as seis expressões mais comumente usadas (Figura 2-2), alegria (joy), tristeza (sadness), raiva (anger), medo (fear), nojo (disgust) e surpresa (surprise). Estas expressões foram definidas em (EKMAN et al., 1979) e são amplamente aceitas pela comunidade científica. Na Tabela 2-3 são mostradas as ações específicas que devem ser tomadas para representar cada expressão.

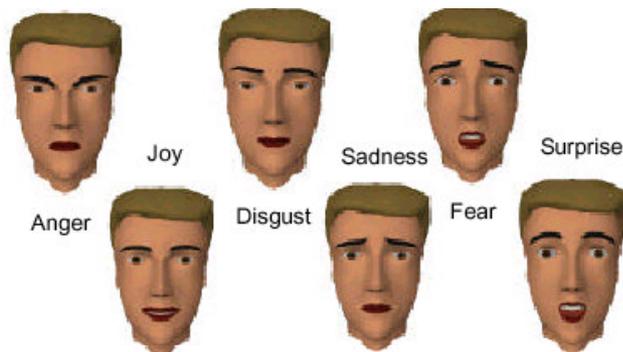


Figura 2-2 - Expressões básicas definidas por Ekman

Fonte: (EKMAN et al., 1979)

Tabela 2-3 - Ações para cada expressão

Expressão #	Nome	Descrição da expressão
0	nada	Nada
1	Alegria	As sobrancelhas estão relaxadas. A boca está aberta e os cantos da boca esticados em direção às orelhas.
2	Tristeza	As sobrancelhas internas estão dobradas para cima. Os olhos estão levemente fechados. A boca está relaxada.
3	Raiva	As sobrancelhas internas estão esticadas para baixo e juntas. Os olhos estão bem abertos. Os lábios estão apertados um contra o outro ou abertos, expondo os dentes.

Expressão #	Nome	Descrição da expressão
4	Medo	As sobrancelhas estão levantadas e esticadas juntas. As sobrancelhas internas estão dobradas para cima. Os olhos estão tensos e em atitude de alerta.
5	Nojo	As sobrancelhas estão relaxadas. O lábio superior está levantado e ondulado, freqüentemente de forma assimétrica.
6	Surpresa	As sobrancelhas estão levantadas. As pálpebras superiores estão totalmente abertas, e os de baixo relaxados. A mandíbula está aberta.

Fonte: (ISO/IEC JTC1/SC29/WG11, 1998)

A descrição dos FAPs, assim como suas medidas de animação se encontram no Anexo B. Como pode ser observado nesta tabela, para cada FAP, o padrão MPEG-4 está definido como segue:

- FAPU apropriada.
- Grupo FAP e pontos característicos.
- Direção do movimento positivo.
- Valores mínimos e máximos (em FAPUs).
- Movimento uni-direcional ou bi-direcional.

Valores que extrapolam os mínimos e máximos também são aceitos. Estes produzem movimentos impossíveis para o ser humano, mas são úteis para desenhos animados, por exemplo.

2.4 Unidades FAPU

O objetivo das FAPUs (Facial Animation Parameter Unit) é permitir a escalabilidade dos FAPs. Se estas últimas fossem expressas em unidades físicas, como milímetros ou pixels, o resultado obtido no receptor dependeria fortemente do tamanho da face. Para resolver este problema, o padrão define as FAPUs. Estas são distâncias entre objetos do rosto (Figura 2-3) e são utilizadas como unidades de medida para os FAPs, cada FAP utiliza uma destas unidades (Anexo B).

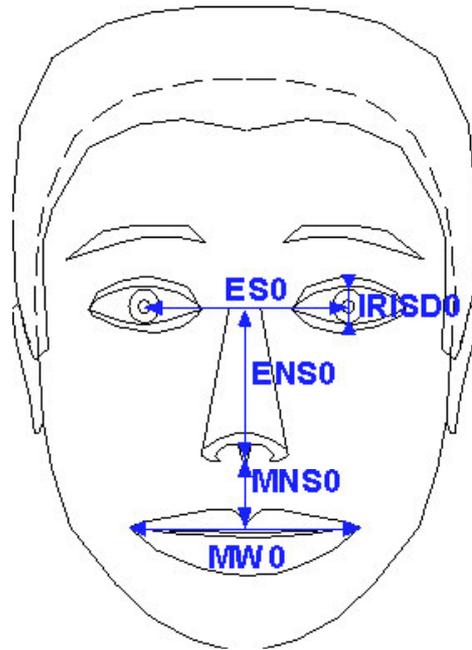


Figura 2-3 - Definição das FAPUs
Fonte: (ISO/IEC JTC1/SC29/WG11, 1998)

Tabela 2-4 - Definição das FAPUs

FAPU	FDPS	DESCRIÇÃO	VALOR FAPU
IRISD	$3.1.y - 3.3.y = 3.2.y - 3.4.y$	Diâmetro da íris	$IRISD = IRISD0 / 1024$
ES	$3.5.x - 3.6.x$	Separação dos olhos	$ES = ES0 / 1024$
ENS	$3.5.y - 9.15.y$	Separação entre os olhos e o nariz	$ENS = ENS0 / 1024$
MNS	$9.15.y - 2.2.y$	Separação entre a boca e o nariz	$MNS = MNS0 / 1024$
MW	$8.3.x - 8.4.x$	Largura da boca	$MW = MW0 / 1024$
AU		Unidade angular	$AU = 10^{-5} \text{ rad}$

Fonte: (ISO/IEC JTC1/SC29/WG11, 1998)

2.5 Tabela FIT

A tabela FIT (Facial Interpolation Table) permite que seja enviado um número menor de FAPs. A tabela consiste num conjunto de regras, especificadas pelo transmissor, e que permitem obter o conjunto de FAPs omitidos, a partir dos FAPs enviados. Por exemplo, pode ser enviado o FAP interno do lábio superior, o qual será usado no receptor para determinar o FAP externo do lábio superior. Estas regras são especificadas por uma FIG (Facial Interpolation Graph) e por um conjunto de funções polinomiais racionais que especificam quando e como o conjunto de FAPs recebidos será usado para determinar os FAPs omitidos.

Neste capítulo foi apresentado o padrão MPEG-4 utilizado para a modelagem e animação de faces. A seguir são apresentados os métodos de detecção de faces baseados em características.

3 MÉTODOS BASEADOS EM CARACTERÍSTICAS

Os métodos baseados em características são métodos que dependem fortemente do conhecimento prévio. Aliás, eles se baseiam no conhecimento prévio para extrair características de baixo nível e analisá-las para chegar a estruturas mais significativas. Podem-se dividir, então, em dois grupos; características de baixo nível e análise de características. Correspondem a este grupo os primeiros trabalhos de reconhecimento facial.

3.1 Características de baixo nível

Os métodos baseados em características de baixo nível visam detectar uma face e/ou seus componentes, valendo-se de características básicas dos mesmos, como forma, luminância, cor e textura. Na seqüência são descritas algumas das técnicas mais utilizadas.

3.1.1 Bordas

Esta técnica consiste em detectar uma borda na imagem causada por um determinado componente da face. Uma imagem qualquer costuma ter um grande número de bordas. Portanto, para o bom desempenho desta técnica, é necessário delimitar a área onde será procurada. Alguma outra característica detectada anteriormente pode servir de referência para prever a localização aproximada da característica que está sendo procurada.

Em (CRAW et al., 1992) é utilizado o detector de bordas de Canny (CANNY, 1986) para localizar o queixo, as bochechas, o cabelo, a mandíbula e a boca. O sistema descrito utiliza um padrão poligonal de 40 pontos que definem o contorno do rosto e do cabelo, sobrancelhas, olhos, nariz e boca. Como primeiro passo é encontrado o contorno da cabeça usando um padrão poligonal e em seguida são detectados os olhos usando um algoritmo muito parecido ao anterior, mas usando um modelo poligonal do olho. Neste ponto o sistema tem dois bons pontos de referência para a detecção do resto das características. Cada uma delas tem um algoritmo de procura especializado, o qual recebe como parâmetro a área onde deve ser procurada a característica e retorna a localização da mesma dentro da área.

3.1.2 Informações de níveis de cinza

As técnicas baseadas nas informações de níveis de cinzas trabalham, como o nome já sugere, com a intensidade de cada pixel, desconsiderando a informação de cor e utilizando 256 níveis de cinza. Esta técnica é muito usada na segmentação da face, e na detecção de olhos e narinas. As técnicas mais utilizadas dentro deste grupo são as projeções horizontal e vertical, e a detecção de aglomerados; ambas se apóiam em observações tais como que as sobrancelhas aparecem como áreas escuras sobre regiões circundantes mais claras.

Em (NIKOLAIDIS et al., 2000), o autor utiliza a análise de mínimos e máximos nas projeções horizontal e vertical para detectar os olhos, narinas e boca. A projeção horizontal é um vetor do tamanho da largura da imagem, onde o valor i -ésimo corresponde ao valor médio dos valores de cinza da i -ésima coluna. Analogamente, a projeção vertical é um vetor do tamanho da altura da imagem, onde o valor i -ésimo corresponde ao valor médio dos níveis de cinza da i -ésima fila (Figura 3-1). Os mínimos e máximos relevantes são determinados baseados no gradiente de cada mínimo/máximo em relação ao máximo/mínimo vizinho. A posição de cada uma das características é determinada baseando-se em diversas métricas, como, por exemplo, a posição relativa da característica dentro da face, o significado de um máximo entre dois mínimos, semelhança entre níveis de cinza, relação da distância entre mínimos e a largura da cabeça, entre outras. Finalmente, o melhor conjunto é determinado baseado na simetria vertical do mesmo.

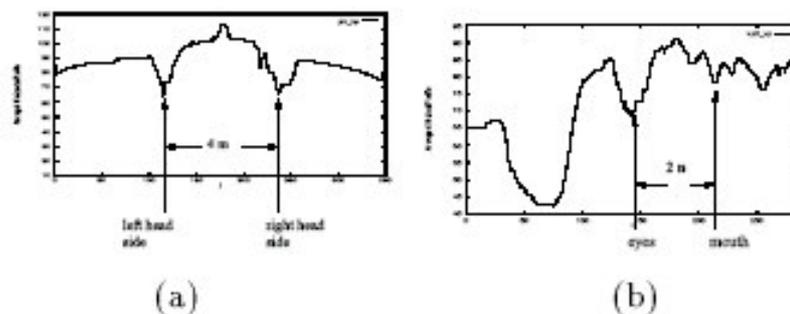


Figura 3-1 - Projeção horizontal e vertical de uma imagem
Fonte: (KOTROPOULOS et al., 1997).

Yang (1994) utiliza no seu trabalho uma técnica de imagens de multi-resolução ou *imagens mosaico*, onde a detecção da face é feita em baixa resolução, e a validação dos candidatos encontrados é feita detectando características faciais em resoluções mais altas. A detecção é feita utilizando um modelo abstrato de face (Figura 3-2), o qual é deslizado sobre a imagem verificando duas condições: uma área homogênea e clara sob as células de cor cinza claro (Figura 3-2), e uma região circundante, homogênea e escura, sob as células pretas do

modelo. Quando as duas condições forem confirmadas a área é etiquetada como candidata a face. O processo é repetido reduzindo sucessivamente a resolução da imagem. Na seguinte fase, as regiões candidatas devem ser validadas. Para tal, é feita uma re-amostragem da imagem, desta vez utilizando o dobro da definição na qual foi detectada. Utilizando as projeções horizontal e vertical, e baseando-se nas características geométricas da face, os autores montaram um conjunto de regras que devem ser cumpridas para o candidato ser confirmado como uma face. Por exemplo, os olhos são procurados como áreas escuras na metade superior da face, enquanto que o nariz e as narinas são procurados nas colunas entre os olhos e embaixo deles, como mínimos locais próximos da coluna central, e simétricos. A boca deve ser uma área homogênea de 2 ou 3 células e separada, no máximo, 3 células das narinas.

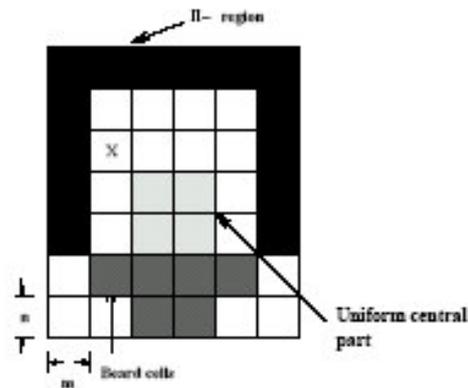


Figura 3-2 - Modelo abstrato da face imagem
Fonte: (KOTROPOULOS et al., 1997)

Em imagem (KOTROPOULOS et al., 1997), esta técnica é usada de forma diferente. Neste trabalho, a resolução na qual será detectada a face é determinada a partir do tamanho do rosto evitando o algoritmo de força bruta utilizado por Yang (1994). Como primeiro passo é calculado o tamanho de cada célula do mosaico, de tal forma que o rosto fique totalmente inserido dentro da área 4 x 4, central do modelo (área clara central na figura). Para tais efeitos são calculadas as projeções horizontal e vertical (Figura 3-1). Para o caso de uma imagem contendo uma face, o perfil horizontal apresenta dois mínimos locais significativos, correspondentes aos limites laterais da face, então, a dimensão horizontal de cada célula do modelo é calculada da seguinte forma:

$$lc = \frac{|mh2 - mh1|}{4}$$

Onde lc é a largura da célula e, $mh1$ e $mh2$ são os mínimos direito e esquerdo respectivamente. Já no perfil vertical, podemos identificar, depois da depressão principal

correspondente ao cabelo, um mínimo local significativo, correspondente aos olhos, e depois do máximo relativo, o seguinte mínimo relativo corresponde à boca. Desta forma, a dimensão vertical da célula é calculada como:

$$hc = \frac{|mv1 - mv2|}{2}$$

Onde hc é a altura da célula e, $mv1$ e $mv2$ são os mínimos correspondentes aos olhos e à boca respectivamente.

Neste momento, a imagem original é dividida em células de tamanho $lc \times hc$ e uma nova imagem será gerada, onde o valor de cada pixel da nova imagem será igual ao valor médio dos níveis de cinza dentro da célula correspondente. A nova imagem será de dimensões:

$$l' = \frac{l}{lc} \quad \text{e} \quad h' = \frac{h}{hc}$$

onde l e h são as dimensões da imagem original. Note-se que nesta nova imagem a largura da face é de 4 pixels, enquanto que a distância entre os olhos e a boca será de dois pixels. Na sequência deve-se verificar se a imagem pode ser considerada como um candidato à face, da mesma forma utilizada por Yang (1994). Por último, a validação da face também é feita de forma análoga a Yang (Ibidem), utilizando um conjunto de regras que devem ser verificadas para cada característica detectada.

3.1.3 Cor e textura

Os métodos baseados em cor e textura são usados principalmente na detecção de áreas candidatas à face. Eles se valem do fato de que a cor da pele dos seres humanos, independentemente da raça, forma um aglomerado denso no espaço de cores, o que facilita a sua classificação (Figura 3-4). Na literatura podem ser encontradas diversas abordagens para o aproveitamento desta característica: delimitar o espaço ocupado pelo aglomerado de cores, aproximação por distribuição Gaussiana, e modelagem estatística, entre outras.

Em seu trabalho, Senior (1999) utiliza a detecção pela cor da pele para obter regiões de interesse considerando o espaço Tonalidade-Cromaticidade-Intensidade (HCI – Hue-Chromaticity-Intensity) no qual o aglomerado de cores pode ser delimitado pelo cuboide definido por $H_{\min}, H_{\max}, C_{\min}, C_{\max}, I_{\min}$ e I_{\max} . Como a imagem é obtida em RGB, os limites do cuboide HCI são transformados em limites no espaço RGB. Para cada par (G,B) é determinado um limite máximo e um limite mínimo.

$$I_{\min}, I_{\max} \quad I = R + G + B$$

$$-\frac{\pi}{2} < H_{\min} < 0 < H_{\max} < \frac{\pi}{2}$$

$$C_{\min}, C_{\max}$$

Dados G , B e $M = \min(G, B)$, os limites de R são calculados por:

$$I_{\min} < R + G + B < I_{\max} \Rightarrow I_{\min} - G - B < R < I_{\max} - G - B$$

$$G + B + C_{\min} < R < G + B + C_{\max}$$

$$R < \max(M, C_{\max} + 3M - G - B)$$

$$R > \min(M, C_{\min} + 3M - G - B)$$

$$\text{Se } G > B \Rightarrow R > \frac{1}{2} \left(\frac{(G - B)\sqrt{3}}{\tan(H_{\min})} + G + B \right)$$

$$\text{Senão} \quad R < \frac{1}{2} \left(\frac{(G - B)\sqrt{3}}{\tan(H_{\max})} + G + B \right)$$

A partir destas formulas são geradas duas tabelas, uma com os valores mínimos de R e a outra com os valores máximos. Estas tabelas têm como entrada todas as combinações possíveis de valores de G e B , e como saída o correspondente valor de R_{\min} ou R_{\max} para o par (G, B) . Assim, para cada pixel da imagem de entrada, o valor de R é comparado com os limites indexados por G e B . Se estiver entre os limites, o pixel é etiquetado como tendo a cor da pele, e é feita a soma cumulativa dos pixels com tom da pele a esquerda do pixel atual e acima. No final, para cada região candidata é calculada a proporção de pixels com tom da pele e se esta proporção estiver por cima de um determinado limiar, a região é confirmada como candidata.

Já Yang (1996) utiliza um modelo Gaussiano para determinar se um pixel tem cor de pele. Primeiro os pixels são representados no espaço RGB normalizado, conhecido como espaço cromático ou de cores puras, onde:

$$r = \frac{R}{R + G + B} \quad \text{e} \quad g = \frac{G}{R + G + B}$$

O modelo é gerado, então, a partir de um conjunto de treinamento, de onde são calculadas as médias e a matriz de covariância, e modelado pela função Gaussiana $N = (m, \Sigma^2)$, com $m = (\bar{r}, \bar{g})$, onde:

$$\bar{r} = \frac{1}{N} \sum_{i=1}^N r_i \quad \bar{g} = \frac{1}{N} \sum_{i=1}^N g_i \quad \Sigma = \begin{bmatrix} \sigma_{aa} & \sigma_{ab} \\ \sigma_{ba} & \sigma_{bb} \end{bmatrix}$$

A detecção de regiões candidatas a face para uma determinada imagem de entrada é feita representando cada pixel no espaço cromático e comparando o mesmo com o modelo

Gaussiano obtendo uma imagem em tons de cinza, onde a sua intensidade corresponde à possibilidade de ter a cor da pele. Na Figura 3-3, na imagem da direita, pode-se ver o resultado da detecção, onde o pixel da imagem original é mantido se o valor do mapa de probabilidade for superior a um limiar, ou apagado no caso contrario.

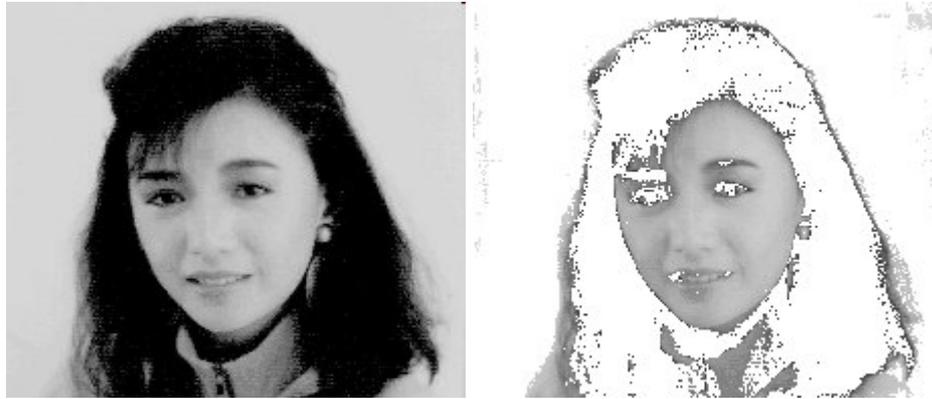


Figura 3-3 - Exemplo de detecção da face pela cor da pele
Fonte: (YANG et al., 1996)

Em (JEBARA et al., 1997) é apresentado um sistema para o seguimento de faces em seqüência de imagens, que tem como principal característica, a utilização de um modelo 3D, o qual é modelado a cada quadro, e que por sua vez, auxilia no seguimento das características da face no quadro seguinte, por meio da informação de posição da cabeça contida nele. O sistema tem uma fase de inicialização na qual é detectada a face e suas características. Esta fase é composta pelos seguintes passos: classificação da pele; transformada da simetria, normalização 3D e análise por eigenfaces³. O primeiro passo é a detecção de aglomerados de pixels pela classificação dos mesmos como sendo, ou não, de cor semelhante à cor da pele. A classificação é feita utilizando um modelo de distribuição probabilística da cor da pele (Figura 3-4), o qual foi criado utilizando o método Expectation Maximization⁴, sobre um conjunto de amostras da cor da pele de diversos indivíduos com diferentes tons de pele e sob diferentes condições de iluminação. O modelo é descrito pela equação:

$$p(x) = \sum_{i=1}^n \frac{mix_i \exp\left\{-\frac{1}{2}(x - \mu_i)^T \Sigma_i^{-1}(x - \mu_i)\right\}}{(2\pi)^{(d/2)} |\Sigma_i|^{1/2}}$$

³ Autovetores de um conjunto de faces.

⁴ Método probabilístico para obter estimadores de máxima verosimilhança em problemas com informação incompleta. Mais detalhes sobre o método de EM pode ser encontrado em Dempster (1977)

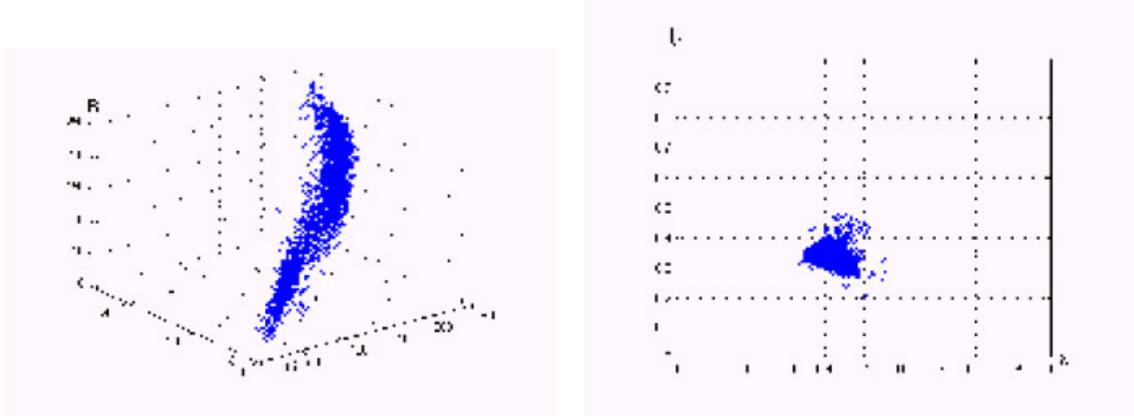


Figura 3-4 - Aglomeração de cores no espaço de cores e aglomeração no espaço cromático.
Fonte: (FERIS et al., 2000)

Para cada imagem obtida, cada pixel é avaliado usando o modelo; se estiver por cima de um determinado limiar, o pixel é etiquetado como “pele”. Por último é feita uma análise de componentes conectados para determinar as regiões de pele na imagem (Figura 3-5).



Figura 3-5 – Processo de detecção da face pela cor da pele. De esquerda a direita, imagem original, pixels marcados como tendo cor de pele e por último, a imagem resultante após ter sido submetido a uma análise de componentes conectados.

Fonte: (JEBARA et al., 1997)

3.1.4 Outras características

Em (REISFELD et al., 1995), o autor propõe um sistema para detecção de simetrias locais, isto é, áreas relativamente pequenas da imagem que apresentam simetria. Nesse trabalho a técnica é utilizada para encontrar focos de atenção, porém, se faz referência a outro trabalho dos mesmos autores no qual é utilizada a mesma técnica para extrair características de uma face.

O algoritmo percorre cada ponto da imagem e calcula para o mesmo, um vetor de simetria formado pela magnitude, que é proporcional a quão simétrico o ponto é; e um ângulo que define a direção na qual o ponto apresenta mais simetria. Desta forma, o algoritmo

mapeia, no lugar de bordas ou gradientes, áreas com maior ou menor simetria. Como foi mencionado anteriormente, utilizando a mesma técnica, mas calculando a simetria numa determinada direção, podem ser extraídas as características de uma face.

A simetria para cada ponto $p_k = (x_k, y_k)$ pertencente à imagem, é definida da seguinte forma:

Seja $\nabla p_k = \left(\frac{\partial p_k}{\partial x}, \frac{\partial p_k}{\partial y} \right)$ o gradiente de intensidade no ponto p_k , e se assume que o vetor $v_k = (r_k, \theta_k)$ associado com p_k é tal que $r_k = \log(1 + \nabla p_k)$ e $\theta_k = \arctan\left(\frac{\partial p_k}{\partial y} / \frac{\partial p_k}{\partial x}\right)$.

Dados dois pontos da imagem, p_i e p_j , chama-se l à linha que os une, e α_{ij} o ângulo anti-horário entre esta linha e a horizontal.

Define-se:

$\Gamma(p) = \left\{ (i, j) \mid \frac{p_i + p_j}{2} = p \right\}$ como o conjunto de pares de pontos que tem p como ponto médio.

$D_\sigma(i, j) = \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{\|p_i - p_j\|}{2\sigma}}$ como uma função de peso da distância, e
 $P(i, j) = (1 - \cos(\theta_i + \theta_j + 2\alpha_{ij}))(1 - \cos(\theta_i - \theta_j))$ como uma função de peso da fase.

Define-se então a contribuição dos pontos p_i e p_j como:

$$C(i, j) = D_\sigma(i, j) \cdot P(i, j) \cdot r_i \cdot r_j$$

Por último, a *simetria isotrópica* no ponto p é definida como:

$$S_\sigma(p) = (M_\sigma(p), \phi(p))$$

onde $M_\sigma(p) = \sum_{(i,j) \in \Gamma(p)} C(i, j)$ é a magnitude da simetria,

e $\phi(p) = \left\{ \phi(i, j) \mid C(i, j) = \text{MAX}(C(k, l)), (k, l) \in \Gamma(p) \right\}$ é a direção da linha que une os dois pontos com maior contribuição.

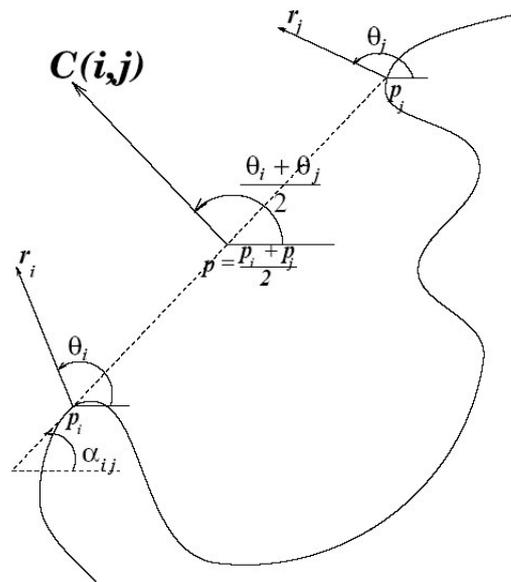


Figura 3-6 - Contribuição à simetria dos gradientes nos pontos p_i e p_j .

Fonte: (REISFELD et al., 1995).

Em alguns casos pode ser necessário detectar pontos com simetria em várias direções, e não somente numa, com este propósito o autor define a *simetria radial* como

$$RS_{\sigma}(p) = \sum_{(i,j) \in \Gamma(p)} C(i,j) \sin^2(\varphi(i,j) - \phi(p)).$$

Os detectores de simetria apresentados são livres de contexto, quer dizer, não importa qual a imagem, eles simplesmente detectam pontos com simetria. Mas também pode se valer do conhecimento prévio do objeto a ser detectado. Por exemplo, o autor define a *projeção de simetria* como $PS_{\sigma}(p, \psi) = M_{\sigma}(p) \cos(\phi(p) - \psi)$, a qual detecta simetrias no ponto p e na direção ψ . Como mencionado anteriormente, os autores utilizaram este operador de simetria, num outro trabalho, para detecção de características faciais.

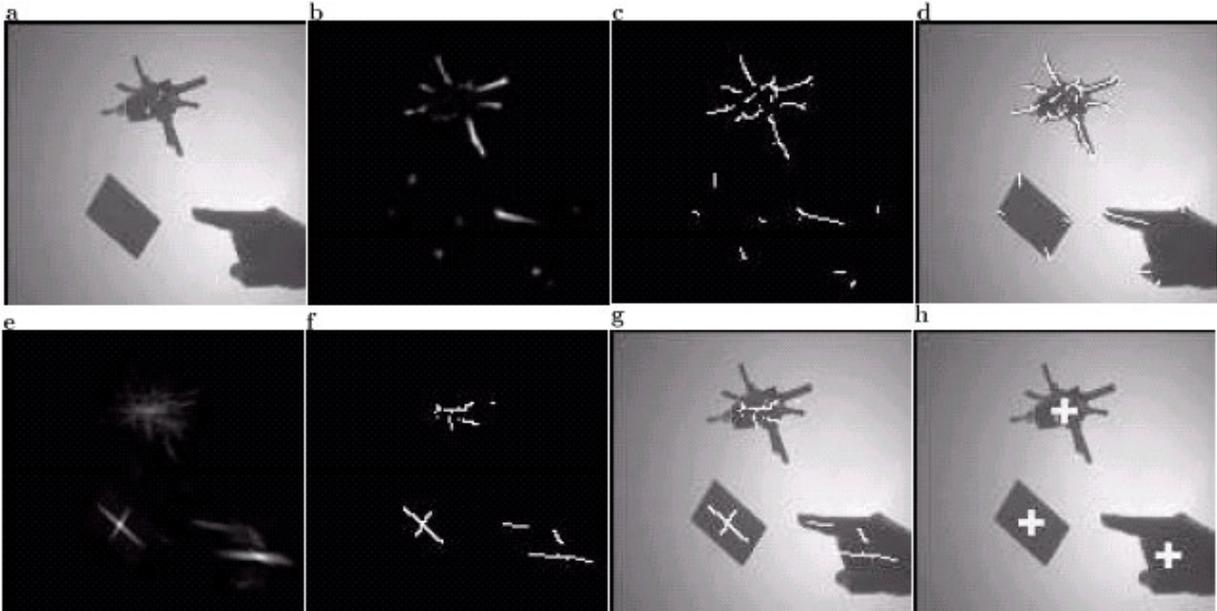


Figura 3-7 - (a) Imagem original; (b) Mapa de simetria isotrópica para um σ pequeno; (c) Binarização da imagem (b); (d) Imagens (a) e (c) sobrepostas; (e) Mapa de simetria isotrópica para um σ grande; (f) Binarização da imagem (e); (g) Imagens (a) e (f) sobrepostas; (h) Valores máximos de simetria radial marcados com uma cruz sobre a imagem original.

Fonte: (REISFELD et al., 1995).

3.2 Análise de características

Na seção anterior foram apresentados os métodos de extração de características de baixo nível. O problema com estas é a sua ambigüidade, quer dizer, não são características exclusivas da face; por exemplo, se aplicado sobre a imagem um detector pela cor da pele, pode acontecer que no fundo da mesma, fora da região da face, sejam detectados alguns pixels (vide Figura 3-5 central). A análise de características visa trabalhar em conjunto com a extração de características de baixo nível para, baseando-se em conhecimento prévio sobre a disposição espacial das características faciais, validar as características detectadas ou determinar onde devem ser buscadas.

3.2.1 Abordagem estrutural

Também chamada de busca de características, esta abordagem faz uma busca hierárquica das características faciais. Começa-se detectando alguma característica proeminente, da qual se tenha um alto grau de certeza, e usando as já detectadas como referência, encontra-se as restantes.

Como já foi mencionado na seção 3.1.2, uma vez segmentada a região da face, Kotropoulos (1997) utiliza o dobro de definição da utilizada para detectar a região candidata e, portanto a mesma fica dividida em 8×8 células. O autor começa procurando a localização

dos olhos, buscando na projeção vertical, o mínimo local dentro da metade superior da região da face. Esse mínimo determina a fila na qual se encontram os olhos. Em seguida procuram-se nessa fila, duas células escuras, e que sejam simétricas em relação ao centro. Analogamente, as narinas são detectadas como mínimos locais, mas na área restringida entre as colunas dos olhos e embaixo da linha dos mesmos. Finalmente a boca é detectada como uma área homogênea de 2 ou 3 células, e no máximo 3 filas embaixo das narinas.

Já Nikolaidis (2000) utiliza um método parecido para a extração dos olhos, nariz e boca (veja seção 3.1.2), mas vai além, extraindo também as bochechas e o queixo. Para estas últimas o autor se aproveita do fato das mesmas poderem ser aproximadas com curvas geométricas. As bochechas podem ser aproximadas por uma linha aproximadamente vertical, expressa pela equação:

$$x=my+c \quad \text{com } m \cong 0$$

sendo m a inclinação ou pendente da linha, e c o ponto onde a linha corta o eixo y . Por sua vez, o queixo pode ser aproximado por uma parábola positiva com eixo vertical, expressa pela equação:

$$x - x_v = g(y - y_v)^2$$

sendo (x_v, y_v) o vértice da parábola e g é proporcional à distância do vértice ao foco. Como ferramenta para detectar estas linhas na figura, o autor utiliza uma variante da Transformada de Hough, proposta por (LI, apud NIKOLAIDIS, 2000), chamada de Transformada de Hough Adaptativa (AHT). Esta técnica procura por uma curva pré-determinada, iterativamente na área de interesse, e redefinindo a área a cada passo. A área de interesse inicial é determinada a partir das características encontradas anteriormente (Figura 3-8).

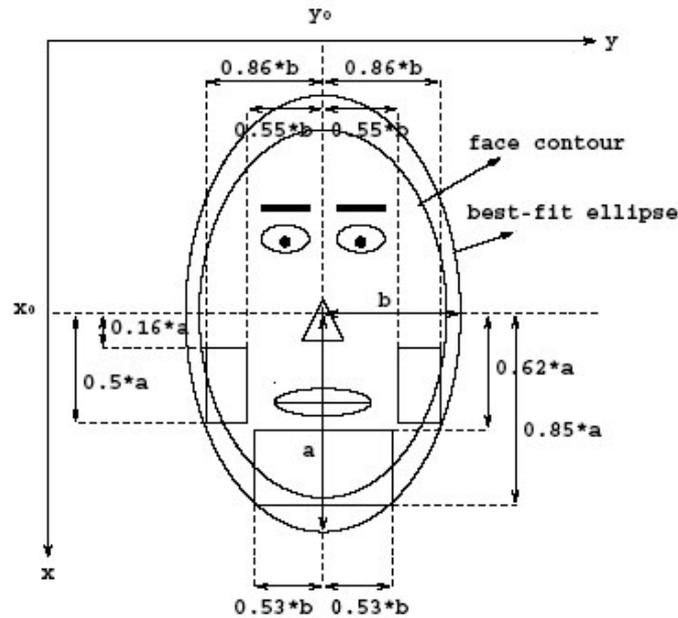


Figura 3-8 - Ilustração das sub-imagens relevantes para a transformada de Hough.
 Fonte: (NIKOLAIDIS et al., 2000)

Já as sobrancelhas não podem ser extraídas pelo método anterior por ter uma variação de forma muito grande de uma pessoa para outra. O autor utiliza então uma abordagem por cassação de padrões para detectar a borda superior da sobrancelha. Com esse objetivo é definido um bloco protótipo, P , que será usado como máscara binária. O bloco P é aplicado como uma janela deslizante sobre o gradiente limiarizado da imagem. A seguir, define-se o bloco ótimo B_{opt} , que melhor representa a característica procurada, como aquele que tem máxima correlação com P :

$$B_{opt} = \left\{ B_j \in S \mid \sum P \cdot B_j = \max_{B_i \in S} \left(\sum P \cdot B_i \right) \right\}$$

onde $B_i \in S$ e S é o conjunto dos blocos dentro da área de busca. O tamanho do bloco pode ser normalizado pela distância entre os olhos.

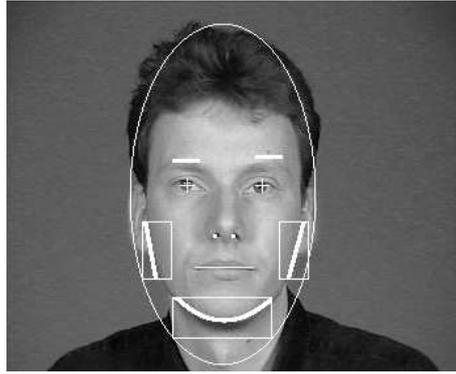


Figura 3-9 - Exemplo de características detectadas pelo método proposto por Nikolaidis.
Fonte: (NIKOLAIDIS, 2000)

3.2.2 Filtros de Gabor

Basicamente um filtro de Gabor⁵ é um plano ondulado, modulado por uma Gaussiana, o que configura um filtro 2D que resalta linhas com uma determinada direção e uma determinada frequência. Os filtros de Gabor estão relacionados com o funcionamento do córtex visual.

O algoritmo descrito em (SMERALDI et al., 2000), faz a localização dos olhos de uma pessoa por meio da decomposição da imagem utilizando filtros de Gabor. Estes trabalham no domínio da frequência, quer dizer, eles têm uma frequência associada e ressaltam as partes da imagem que conferem com a mesma. O problema com os filtros de Gabor é seu alto custo computacional. Para resolver esta dificuldade, os autores utilizam um algoritmo com busca sacádica. A busca sacádica é a forma como o olho humano explora a cena visual, efetuando longos saltos entre os pontos de interesse. Neste trabalho, é utilizada uma grade polar-logarítmica (Figura 3-10) onde os filtros de Gabor são aplicados somente nos pontos da grade, e com uma frequência menor quanto maior a distância ao centro de atenção. A grade é aplicada sucessivamente em diferentes pontos da imagem, escolhidos pelo resultado da grade corrente, até encontrar seu alvo.

Os autores utilizaram como função de Gabor base:

$$\hat{G}(\vec{\omega} | \sigma_{\rho}, \sigma_{\phi}, \rho_0) = \exp\left(-\frac{(\rho - \rho_0)^2}{2\sigma_{\rho}}\right) \cdot \exp\left(-\frac{\omega_{\phi}}{2\sigma_{\phi}^2}\right)$$

⁵ Neste trabalho não será apresentado o embasamento teórico dos filtros de Gabor. Um detalhamento dos mesmos aplicados ao tratamento de imagens pode ser encontrado em Lee (1996).

onde $(\rho, \phi) = \left(\ln\left(\frac{1}{\omega}\right), \tan^{-1}(\omega_y/\omega_x) \right)$ é o mapeamento do plano de frequências para as coordenadas polar-logarítmicas.

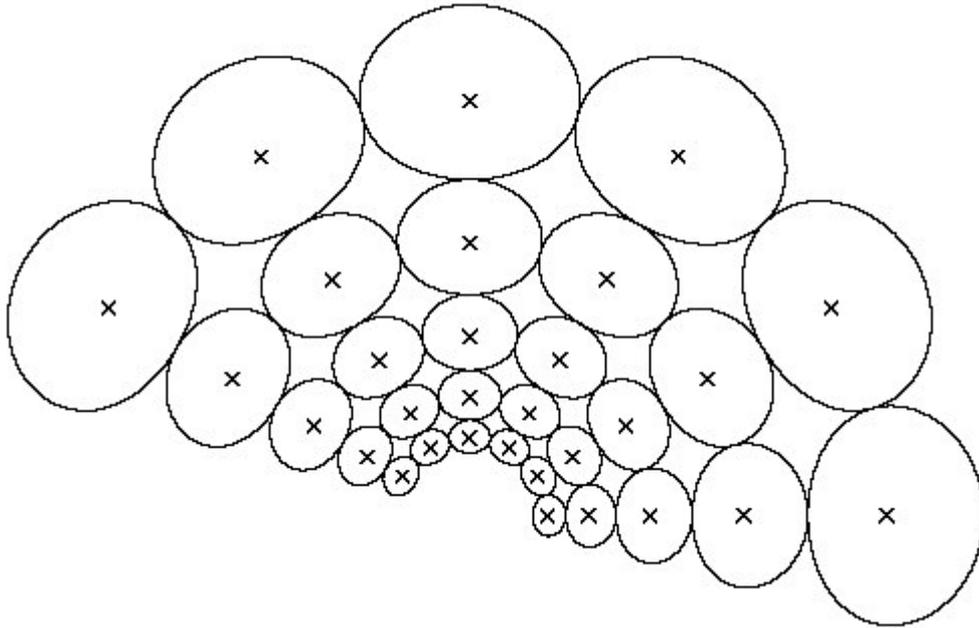


Figura 3-10 – Grade polar-logarítmica usada no algoritmo GAZE.
Fonte: (SMERALDI et al., 2000)

Na fase de treino é computado o vetor e_{av} , descritor do alvo (o olho), por meio do cálculo da média da resposta dos filtros de Gabor, aplicados no centro do olho, para todas as imagens de treino (Figura 3-11). O vetor resultante é composto por seis respostas orientação-seletivas para cada um dos cinco canais de frequência utilizados.

No início da busca, a grade é colocada num lugar aleatório da imagem, e o vetor de Gabor é computado para cada um dos pontos da mesma. Na seqüência é calculada a distância Euclidiana entre cada um destes vetores e o vetor e_{av} . O ponto da grade com a menor distância ao vetor e_{av} é escolhido como o centro da grade para a próxima iteração (*saccade*). A busca é finalizada quando as *saccades* ficam menores que um determinado limiar. Se após algumas iterações o alvo não for encontrado, a busca é reinicializada num novo ponto aleatório (Figura 3-12).



Figura 3-11 - A grade polar-logarítmica aplicada sobre o olho direito de uma face na fase de treinamento.

Fonte: (SMERALDI et al., 2000)

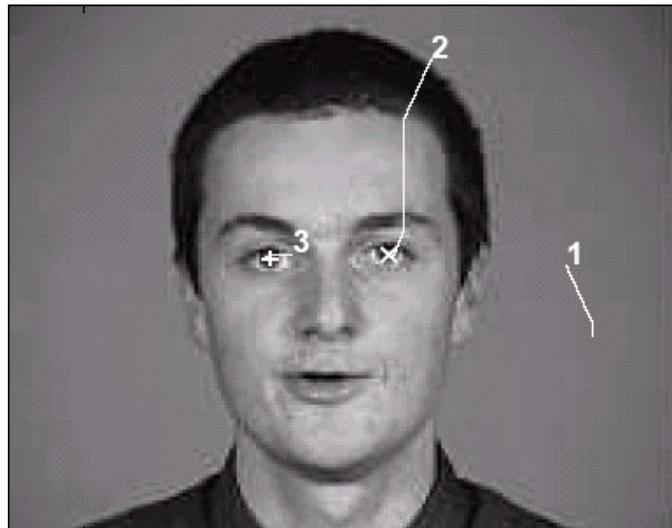


Figura 3-12 - Representação gráfica do processo de busca sacádica. Os números representam os pontos de início, enquanto o X aponta a detecção do olho esquerdo.

Fonte: (SMERALDI et al., 2000)

3.2.3 Constelações

Nas técnicas baseadas em constelações, a face é detectada utilizando um modelo conformado por elementos, ou partes, e as relações espaciais entre eles (distâncias e ângulos). De forma geral, estes algoritmos buscam todas as regiões da imagem que conferem com alguma das partes do modelo, e depois, utilizando as relações entre as mesmas, buscam uma configuração válida.

Em (VOGELHUBER et al., 2000) é apresentado um detector de faces baseado em descritores locais genéricos e relações espaciais entre eles (Figura 3-13). Estes descritores caracterizam os olhos, o nariz e os cantos da boca, e são formados por modelos de misturas Gaussianas, computados a partir de um conjunto de imagens de treinamento. A estimativa destes modelos é feita pelo algoritmo Expectation Maximation (EM), combinado com o algoritmo Minimum Description Length (MDL).

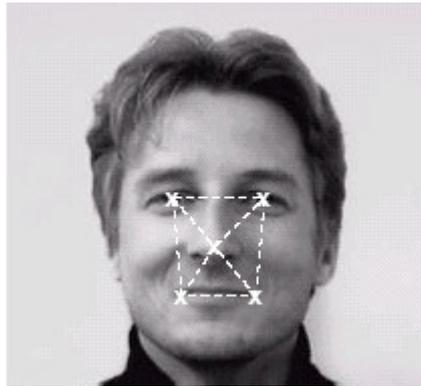


Figura 3-13 - Representação da face com os descritores genéricos e as respectivas relações espaciais.

Fonte: (VOGELHUBER et al., 2000)

Para cada face do conjunto de treinamento são marcadas as mesmas características (p.ex. centro do olho), e para cada uma destas é calculado o descritor local. O descritor local é um vetor que descreve uma pequena área no entorno da característica. Os autores utilizaram derivadas Gaussianas para computar os descritores locais. Terminado o treinamento, para cada característica é calculada a distribuição estatística dos descritores da mesma usando um modelo de misturas Gaussiano.

As relações espaciais correspondem a ângulos e distâncias entre os diferentes pontos característicos, e são representados no modelo como variáveis Gaussianas com média e desvio padrão, o que permite uma certa variabilidade devido a diferenças morfológicas.

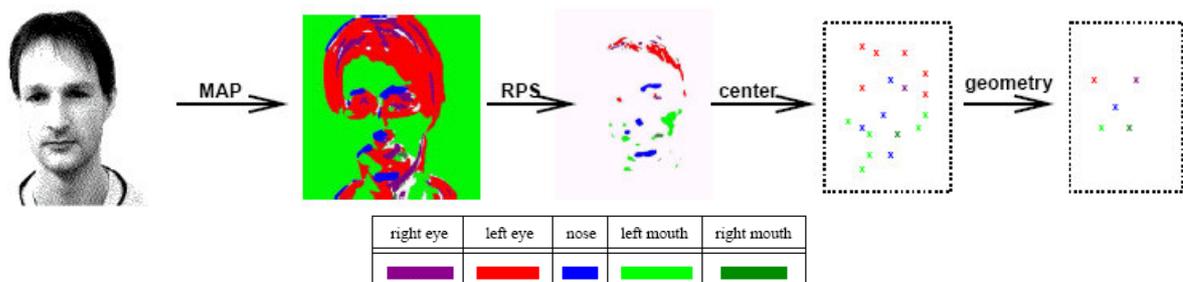


Figura 3-14 - Processo de detecção de uma face.

Fonte: (VOGELHUBER et al., 2000).

O processo de detecção de faces numa imagem desconhecida começa calculando o descritor para cada pixel e escolhendo qual descritor genérico mais se aproxima, utilizando o princípio de Máximo a Priori (MAP).

A probabilidade de um descritor d-dimensional X para um descritor genérico D_i é:

$$p(X|D_i) = \sum_{j=1}^{K_i} \omega_{ij} p(X|\mu_{ij}, \Sigma_{ij}) \text{ onde } K_i \text{ é o número de gaussianas, e}$$

$$p(X|\mu_{ij}, \Sigma_{ij}) = \frac{1}{(2\pi)^{d/2} |\Sigma_{ij}|^{1/2}} \exp\left(-\frac{1}{2}(X - \mu_{ij})^T \Sigma_{ij}^{-1} (X - \mu_{ij})\right).$$

A probabilidade de um descritor genérico D_i , dado X é $p(D_i|X) = \frac{p(X|D_i) \cdot p(D_i)}{p(X)}$.

Finalmente, o descritor genérico com maior semelhança é $\hat{D} = \arg \max_{D_i} p(D_i|X)$.

Em seguida tem-se a etapa RPS (Refused Point Step) que rejeita pontos que estão muito longe da distribuição. Para tais efeitos é calculada a distância de Mahalanobis entre o descritor X e o descritor genérico \hat{D} , $d_M(X, D_i) = \sum_j \sqrt{(X - \mu_{ij})^T \Sigma_{ij}^{-1} (X - \mu_{ij})}$.

Cada descritor genérico tem associado um limiar, computado na etapa de treinamento, pelo qual é determinada a rejeição ou não de cada ponto.

Para os pontos restantes são verificados os componentes conectados e é calculado o ponto médio de cada aglomerado de pontos. Finalmente são verificadas as relações espaciais e somente as configurações que confirmam com estas serão aceitas.

Uma abordagem semelhante é apresentada por Weber (2000) utilizando modelos de classes de objetos. Estas classes são formadas por partes e por uma forma, onde as partes são pedaços da imagem que podem ser detectados e caracterizados por descritores, e a forma descreve a posição geométrica entre as partes. O processo de detecção de faces está constituído por três etapas:

1. Os detectores são aplicados à imagem obtendo um conjunto de candidatos
2. Com as partes detectadas são formados conjuntos que tenham semelhança com uma classe de objeto (ex. constelação de características da face: olhos, nariz, boca), constituindo cada um destes conjuntos uma hipótese de objeto. São consideradas as constelações totais e parciais para tratar os casos com oclusão parcial da face.

3. Cada hipótese é avaliada usando a densidade de probabilidade de conjunto do objeto para calcular a semelhança da hipótese com o objeto.

O processo de construção dos descritores das partes começa com a identificação de pontos de interesse nas imagens de treinamento. Utiliza o operador de interesse proposto por Förstner (HARALICK, apud WEBER, 2000) que é capaz de detectar cantos de objetos, cruzamento de linhas assim como centros de padrões circulares (Figura 3-15).

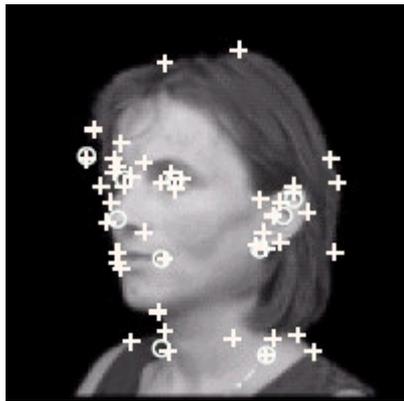


Figura 3-15 - Pontos de interesse detectados usando o método de Förstner. Cruzes indicam padrões do tipo canto, enquanto que os círculos indicam padrões do tipo circular.

Fonte: (WEBER et al., 2000)

Para cada ponto de interesse de cada imagem de treino é calculado um vetor e estes são aglomerados usando o algoritmo k-means (CAMPOS, 2001). Na Figura 3-16 pode se ver o resultado deste processo, onde pode se observar em muitos deles a presença de características faciais.

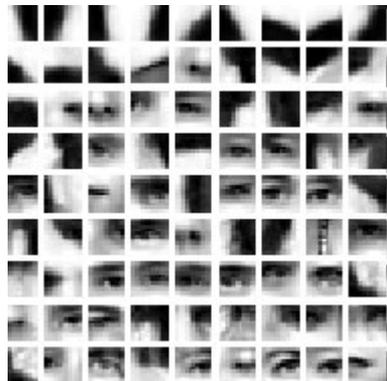


Figura 3-16 - Exemplo de um conjunto de padrões obtidos usando aglomeração por k-means de pequenas áreas da imagem.

Fonte: (WEBER et al., 2000)

A escolha do sub-conjunto das características selecionadas que será utilizado no modelo é feita tentando várias configurações, iterativamente, e avaliando a sua performance utilizando o algoritmo Expectation Maximization (EM).

3.2.4 Modelos probabilísticos

Em (YOW et al., 1997), a detecção da face é feita por um modelo de percepção de duas etapas. Na primeira etapa, chamada *Pre-attentive*, são extraídos os pontos de interesse da imagem original, os quais indicam as possíveis localizações das características faciais. Na segunda etapa, chamada *Attentive*, estes pontos são examinados, agrupados e etiquetados, de acordo com o conhecimento adquirido na fase de treinamento. Para finalizar, os pontos etiquetados são agrupados de acordo com o modelo (Figura 3-17).

O modelo da face está definido por seis características: sobrancelhas(2), olhos(2), nariz e boca. Os sub-grupos de duas e quatro características foram definidos para lidar com a oclusão de uma ou mais características da face (Figura 3-17).

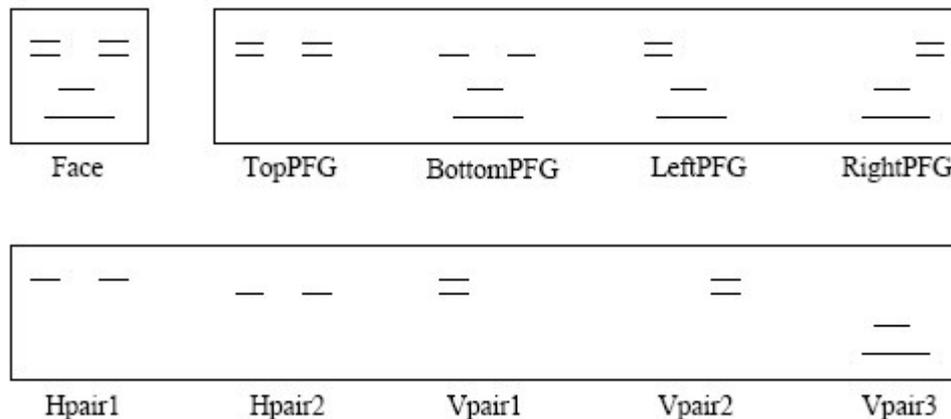


Figura 3-17 - O modelo da face e os grupos de componentes da face.

Fonte: (YOW et al., 1997)

Na primeira etapa, o autor se vale da observação de que numa imagem de baixa resolução, as seis características procuradas se apresentam como uma área escura horizontalmente alongada sobre uma área mais clara. Para detectá-las, então, a imagem é suavizada e passada através de um filtro passa-banda. O filtro utilizado pelos autores é a derivada segunda de uma Gaussiana alongada numa relação 3:1. A saída deste processo é um conjunto de pontos marcando a possível localização de características. A seguir são examinadas as bordas ao redor de cada ponto detectado. São procurados pares de bordas quase horizontais (Figura 3-18). Quando as bordas são encontradas, o ponto de interesse é

marcado como ponto de característica facial. Finalmente é demarcado um retângulo enquadrando as duas bordas ao redor da característica (Figura 3-19), no qual são feitas várias medidas (longitude da borda, largura da borda, variação dos níveis de cinza) e armazenadas num vetor de característica, x . Para cada uma das características faciais, baseado no correspondente conjunto de imagens de treino, é calculada a média e a matriz de covariância. Ex: para o olho tem-se μ_{olhc} e Σ_{olhc} os quais definem a classe de vetores da característica “olho”.

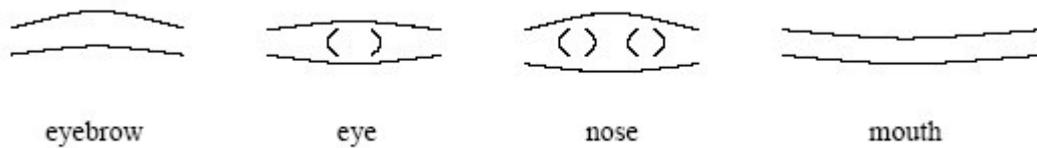


Figura 3-18 - Os modelos de características faciais.

Fonte: (YOW et al., 1997)



Figura 3-19 - Processo de seleção de característica prévio.

Fonte: (YOW et al., 1997)

Dada uma região demarcada, i , candidata à característica j , é calculada a distância de Mahalanobis entre o vetor de característica x_i , e o centro da classe, μ_j . Se a distância for menor que um determinado limiar, a característica é validada.

$$M_{ij} < \tau_j, \text{ onde } M_{ij} = (x_i - \mu_j)^T \Sigma_j^{-1} (x_i - \mu_j)$$

Este procedimento é repetido para cada uma das quatro classes de características.

No final da primeira etapa temos vários pontos candidatos que foram descartados por não possuírem os pares de bordas ou por não se enquadrar em nenhuma das classes de características. Portanto, os pontos restantes estão atribuídos a alguma destas classes.

Na segunda etapa, os pontos são agrupados baseando-se no modelo de conhecimento da face (Figura 3-17). As características são agrupadas em pares verticais e horizontais; os

pares são agrupados em grupos faciais parciais; e os grupos faciais parciais são agrupados em candidatos a face.

Para cada nível k de agrupamento são feitas medidas, e armazenadas num vetor n_k - *dimensional*. Este vetor é projetado no espaço n_k -*dimensional*, determinado na fase de treinamento e, utilizando a distância de Mahalanobis, é determinado se o vetor pertence à classe de agrupamento. As medidas utilizadas são:

- A relação entre a longitude da característica e o tamanho da imagem;
- A relação entre a longitude de uma característica e a longitude de outra característica;
- A relação entre a largura e a altura da região da característica;
- A relação entre distâncias inter-características;
- A diferença de orientação entre características;
- O número de pixels de bordas direcionais numa região (normalizado pelo tamanho da região);
- A relação entre a largura da borda numa região e a largura da borda das características faciais;
- A média dos níveis de cinza de uma região (normalizada pela distribuição de intensidades);
- A variância na distribuição de níveis de cinza de uma região;

Neste ponto, o que se tem são candidatos a face. Para validá-los é utilizada uma rede Bayesiana, de forma que, a existência ou não de cada característica, contribui para o cálculo da probabilidade dos pares de características, as quais são utilizadas no cálculo da probabilidade dos grupos faciais parciais, e estas, por sua vez, utilizadas no cálculo da probabilidade do candidato a face. Se a probabilidade do candidato estiver por cima de um determinado limiar, a face é confirmada.

3.3 Casamento de padrões

Nos métodos por casamento de padrões é definido um modelo da face. A detecção é feita através de uma série de medidas correlacionadas entre o modelo e a imagem. No entanto podem ser enquadradas nesta seção, muitas das técnicas abordadas neste trabalho, trataremos especificamente duas abordagens: padrões pré-definidos e contornos ativos.

3.3.1 Padrões pré-definidos

Em (SCASSELATTI, 1998) é utilizado um molde de face pré-definido para a detecção da face. O artigo descreve um sistema de visão ativa focalizada, composto por duas câmeras, uma de visão periférica e a outra de visão focalizada. A primeira, que tem uma visão mais geral da cena, deve localizar a face, enquanto que a segunda utilizará a informação de posição obtida com a outra câmera, para focalizar a face.

Como foi mencionado anteriormente, para detecção da face o autor utiliza um molde da mesma que define relações de luminosidade entre as diferentes regiões. O molde tem um tamanho de 14 x 16 pixels, e é composto por 16 regiões e 23 relações (Figura 3-20). Para a detecção é utilizado o método de janela deslizante, onde para cada bloco cinza é calculada a média dos valores de tons de cinza dos pixels que estão sob a mesma. Após ter sido calculado o valor para todas as regiões, o sistema confere as relações entre as mesmas. A relação entre duas regiões confere se a divisão entre os valores das mesmas for superior ao valor especificado no modelo para essa relação. Finalmente, a área de 14 x 16 pixels é classificada como uma face se 18 das 23 relações são confirmadas.

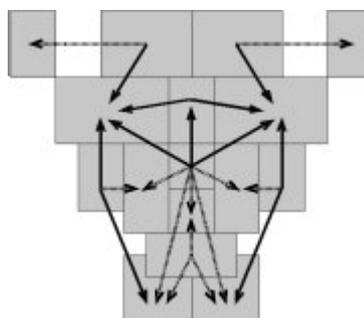


Figura 3-20 - Molde da face de 14 x 16 pixels, composto por 16 regiões (blocos cinza) e 23 relações (setas).

Fonte: (SCASSELATTI, 1998)

3.3.2 Contornos ativos

As técnicas de contornos ativos utilizam um molde que se deforma em função da imagem na qual é aplicado, adquirindo a forma de uma característica específica.

Nikolaidis (2000) utiliza a técnica de deformação dinâmica de contornos ativos (*snakes*) para definir o contorno do rosto com precisão. Como curva inicial utiliza uma elipse

cujos eixos são a metade do tamanho dos eixos da elipse que melhor se ajusta. A energia de cada ponto v_i da curva *snake* é definida por:

$$E_{i-1}(v_{i-1}, v_i, v_{i+1}) = \lambda_i \cdot E_{\text{int}}(v_{i-1}, v_i, v_{i+1}) + (1 - \lambda_i) \cdot E_{\text{ext}}(v_i)$$

onde $\lambda_i \in [0, 1]$ é um parâmetro de regularização, E_{int} indica a energia interna da elipse e afeta a continuidade e curvatura da *snake*, e E_{ext} indica a energia externa da *snake* e representa as forças que a puxam em direção às características significativas da imagem. A energia interna está definida por:

$$E_{\text{int}}(v_i) = \alpha \cdot (|v_i - v_{i-1}|^2 + |v_i - v_{i+1}|^2) + \beta \cdot |v_{i-1} - 2v_i + v_{i+1}|^2$$

onde α e β são parâmetros de normalização que dependem do tamanho da imagem. A energia externa é composta por dois elementos. O primeiro corresponde a saída de um operador gradiente (p.ex. Sobel) e se encarrega de atrair a *snake* em direção às bordas da imagem. O outro componente se refere ao grau de variação da distância euclidiana entre dois pontos adjacentes da *snake*. Este componente se encarrega de manter a suavidade da curva em regiões de baixo contraste entre a face e o fundo.

$$E_{\text{ext}}(v_i) = E_{\text{imagem}} + E_{\text{dist}}$$

onde: $E_{\text{imagem}} = -|\nabla I(x, y)|$ sendo $I(x, y)$ a função tons de cinza da imagem original.

$$E_{\text{dist}} = k \cdot \left| 1 - \frac{d(v_{i-1}, v_i)}{d(v_i, v_{i+1})} \right|$$

onde k é um fator de normalização e $d(v_i, v_j)$ representa a função distância euclidiana entre dois pontos.

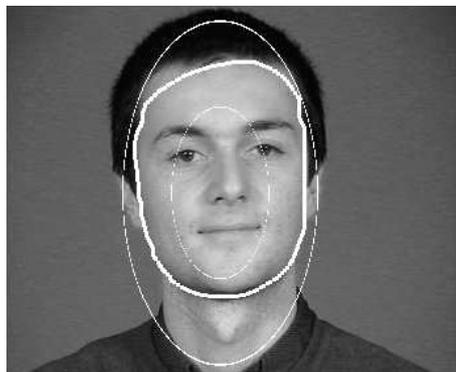


Figura 3-21 - Exemplo de extração do contorno da face.
Fonte: (NIKOLAIDIS et al., 2000)

Apresentou-se neste capítulo os principais métodos baseados em características para a detecção de faces e características faciais em imagens estáticas. No próximo capítulo, serão analisados os métodos baseados na aparência.

4 MÉTODOS BASEADOS NA APARÊNCIA

A principal diferença dos métodos baseados na aparência em relação aos já vistos é que eles não precisam de conhecimento prévio. Em consequência, os sistemas baseados nestes métodos necessitam de treinamento com imagens de exemplo. Na seqüência são explicados os principais métodos baseados na aparência.

4.1 Métodos de sub-espços lineares

Os métodos baseados em sub-espços lineares se apóiam na idéia, bastante intuitiva, de que imagens de faces formam um sub-espço dentro do espço de imagens. O objetivo destes é, então, obter uma base ortonormal para representar este sub-espço, o qual tem uma dimensão muito menor que o espço das imagens. Os autovetores desta base são conhecidos como *eigenfaces*, já que a sua representação gráfica tem grande semelhança com uma face. Por outro lado, é de destacar que os mesmos conceitos podem ser aplicados na representação das diferentes características da face, olhos, nariz, boca, sendo os autovetores chamados de *eigeneyes*, *eigennose* e *eigenmouth* respectivamente. Por sua vez, podem se encontrar na literatura diferentes abordagens, dependendo do método usado para chegar na base canônica, Análise de Componentes Principais (PCA), Análise de Componentes Independentes (ICA) e Análise Fatorial (FA).

Em (ANDRADE, 1997), a obtenção da base de *eigenfaces* usando PCA é descrita da seguinte forma:

Dado o conjunto de imagens de treino $S = \{\Gamma_1, \Gamma_2, \Gamma_3 \dots \Gamma_{M_i}\}$, sendo M a quantidade de imagens, define-se a média do conjunto S como $\Psi = \frac{1}{M} \sum_{n=1}^M \Gamma_n$. Define-se também

$\Phi_i = \Gamma_i - \Psi$, como a diferença entre a i -ésima imagem de treino e a média do conjunto. Na Figura 4-1 pode-se ver a imagem média de um conjunto de 115 imagens e a diferença de uma delas com a imagem média.



Figura 4-1 - Imagem média de um conjunto de 115 imagens (esquerda) e desvio de uma imagem do conjunto com a imagem média.

Fonte: (SIROVICH et al., 1987)

Como mencionado anteriormente, procura-se o conjunto de vetores ortonormais, u_n , que descrevam totalmente o espaço de imagens de faces. Cada vetor u_k é definido de forma que $\lambda_k = \frac{1}{M} \sum_{n=1}^M (u_k^T \cdot \Phi_n)^2$ é máximo, sujeito à condição $u_l^T \cdot u_k = \delta_{lk} = \begin{cases} 1 & \text{se } l = k \\ 0 & \text{c.c.} \end{cases}$

Os vetores u_k e os escalares λ_k são, respectivamente, os autovetores e os autovalores da matriz de covariância $C = \frac{1}{M} \sum_{n=1}^M \Phi_n \cdot \Phi_n^T$ ou, expressado de outra forma $C = AA^T$, onde $A = \{\Phi_1, \Phi_2 \dots \Phi_M\}$.

No entanto, a matriz C é de dimensões $N^2 \times N^2$, sendo N^2 a quantidade de pixels da imagem (considerando imagens quadradas de tamanho $N \times N$), e portanto determinar seus autovetores pode se tornar uma opção inviável.

Considerem-se então os vetores v_i , de dimensão $M \times 1$, autovetores de $A^T A$, tais que $A^T A v_i = \mu_i v_i$

Pré-multiplicando ambos os lados por A , $AA^T A v_i = \mu_i v_i$ se deduz que $A v_i$ são os autovetores de $C = AA^T$. Por outro lado define-se a matriz L , de dimensão $M \times M$, de correlação das imagens de treino $L = A^T A$ onde

$$L_{mn} = \Phi_m^T \Phi_n$$

Para obter os autovetores u_l de C , baseados nos v_i da matriz L , se tem que:

$Av_l = u_l$, $l = 1..M$, são os autovetores de C , então $u_l = \sum_{k=1}^M v_{lk} \Phi_k$, $l = 1..M$, onde v_{lk} é o k -ésimo elemento do vetor v_l da matriz L . A ordem dos cálculos passa então de N^2 para M , que é muito menor.

A representação de uma face, Γ , utilizando *eigenfaces* é determinada pela equação:

$$w_k = u_k^T (\Gamma - \Psi), k = 1..M', \text{ onde:}$$

- w_k é o peso calculado para a k -ésima *eigenface*;
- u_k é a k -ésima *eigenface*;
- Ψ é a média das imagens do conjunto de treino;
- M' é o número de *eigenfaces* mais significativas, que possuem a maior variância da base de dados.

Demirel (1996) utiliza PCA para a detecção de características faciais numa imagem. O diferencial do sistema proposto é a divisão do treinamento em duas etapas. A primeira etapa é um treinamento convencional, feito com imagens que têm a característica já segmentada. Na segunda etapa são utilizadas imagens sem segmentar, mas contendo uma face. Estas imagens são processadas pelo sistema treinado na primeira etapa, e já com as características detectadas, são acrescentadas ao conjunto de treinamento. Finalmente, o sistema é treinado novamente, desta vez utilizando o conjunto de treinamento incrementado.

Como base de classificação, o sistema utiliza a distância ao espaço de características (DFFS), definida como:

$$\varepsilon_e^2 = \|\phi - \phi_f\|^2 = \sum_{j=1}^N (\phi_j - \sum_{i=1}^{M'} \omega_i \cdot \mu_{ij})^2, \text{ sendo } N \text{ o número de pixels nas características}$$

de treino e μ_{ij} o j -ésimo pixel da i -ésima *eigenfeature*.

Na fase de treinamento, para cada Γ^i , sendo a i -ésima imagem segmentada do conjunto de treinamento, aplica PCA para mapear cada característica no espaço de características. Para cada Γ^i calcula a distância ao espaço de características, ε_i . Finalmente, é estimada a distribuição de probabilidade $P_{DFFS}(\varepsilon)$ para o conjunto de distâncias.

Na fase de detecção, o seguinte processo é executado para cada característica a ser detectada. Para cada pixel da imagem de entrada, trata o pixel como o centro de um retângulo

do tamanho do retângulo usado no treinamento da característica. Projeta o retângulo no espaço da característica (*eigenspace*) e calcula a distância do mesmo ao espaço da característica (DFFS). No final do processo se tem um mapa DFFS, o qual será utilizado para determinar o ponto que melhor se aproxima à característica.

O método apresentado por Sênior (1999) utiliza várias técnicas. Começa considerando uma sub-imagem C , com $m \times n$ pixels, para a qual é calculada a proporção de pixels com cor semelhante à cor da pele. Se esta proporção não for maior que certo limiar, a sub-imagem é descartada. Depois é calculado o vetor de C (da mesma forma que é calculado no discriminante de Fisher), excluindo os pixels dos cantos e subtraindo o vetor da média. A seguir é efetuado o produto escalar com o vetor de Fisher e se este valor for suficientemente grande, é calculado o DFFS e subtraído do discriminante. Se o valor resultante estiver por cima de um limiar pode-se dizer que a sub-imagem tem aparência de face.

Na segunda parte é utilizado o chamado discriminante linear de Fisher. Para calcular este discriminante, as imagens de treinamento devem estar com os locais onde se encontram os olhos e o nariz, etiquetados. Basicamente, para cada imagem de treinamento é extraída uma sub-imagem $R(I, m, n, x, \theta, E, k)$ sendo I a imagem de entrada, m e n o tamanho da sub-imagem em pixels, x a localização do nariz e centro da sub-imagem, θ a inclinação da linha dos olhos, E a distância entre os olhos, e k o fator de escala. Neste ponto tem-se, então, sub-imagens padronizadas em tamanho, posição e rotação. Com os pixels destas sub-imagens, mas desconsiderando os pixels dos cantos, é formado um vetor. Com o conjunto de vetores da fase de treinamento, mais outro conjunto de vetores gerados aleatoriamente como vetores fora da classe, é calculado o Vetor Discriminante de Fisher (DUDA, apud SÊNIOR, 1999).



Figura 4-2 - Imagem de uma face com as características a serem localizadas marcadas em branco.

Fonte: (SÊNIOR, 1999)

Para a detecção das características, Senior (1999) utiliza a medida da Semelhança Logarítmica. A partir da imagem já normalizada, no conjunto de treinamento, são medidas as distâncias entre cada par de características (Figura 4-2). No final do treinamento é calculada a média (μ) e a variância (σ) para cada uma destas distâncias. A Semelhança Logarítmica é definida por:

$$L(s) = \sum_{i,j \in S} \left(-\log(\sigma_{ij}) + \frac{(x_i - x_j - \mu_{ij})^2}{\sigma_{ij}^2} + K \right), \text{ sendo } S \text{ o conjunto de características.}$$

É definido também: $\partial L_i = \frac{L(S \setminus i) - L(S)}{\|S \setminus i\|}$ como a contribuição da característica *i-ésima* à semelhança, sendo $S \setminus i$ o conjunto resultante de excluir a característica *i-ésima* do conjunto S . Qualquer característica que provoque um valor alto de ∂L_i deve ser porque sua localização em relação as outras características não é usual e deve ser descartada. Esta operação é repetida até que o valor da semelhança seja alto o suficiente para indicar uma constelação de características consistente.

ICA (Independent Component Analysis) pode ser considerado uma generalização do PCA. O objetivo deste método é encontrar uma transformada linear para a imagem de entrada, usando uma base o mais estatisticamente independente possível. PCA tenta obter uma representação da imagem baseado em variáveis não correlativas, enquanto que o ICA provê uma representação baseada em variáveis estatisticamente independentes. Na Figura 4-3 são apresentados alguns exemplos de *eigenfaces* obtidas pelo método ICA, e comparadas com *eigenfaces* obtidas pelo método PCA sobre o mesmo conjunto de imagens. Déniz (2001) utiliza ICA para o reconhecimento de faces, implementado com o algoritmo FastICA⁶.

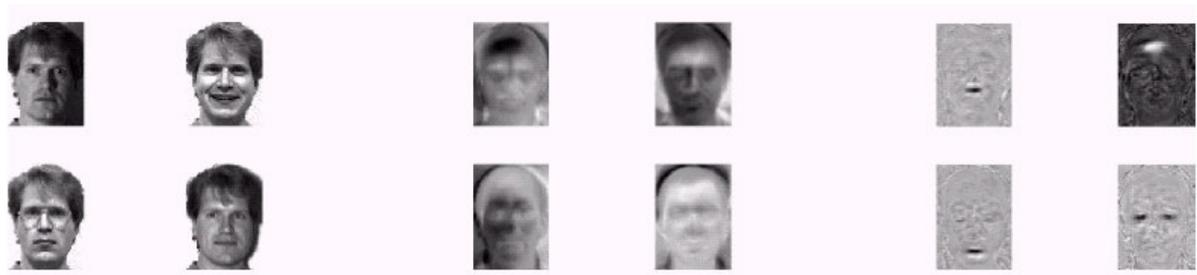


Figura 4-3 - Imagens originais (esquerda), PCA (centro) e ICA (direita).

Fonte: (DÉNIZ et al., 2001)

⁶ FastICA é um algoritmo desenvolvido na Universidade de Helsinque para estimação do ICA. Mais detalhes sobre este algoritmo podem ser encontrados em <http://www.cis.hut.fi/projects/ica/fastica/index.shtml>.

4.2 Métodos baseados em distribuição

Nos modelos baseados em distribuição é feita uma modelagem da distribuição de probabilidade da classe *face*.

O trabalho apresentado por Schneiderman (1998) se baseia na regra de decisão de Bayes, apresentada como uma relação de semelhanças:

$$\text{Se } \frac{P(\text{imagem}|\text{objeto})}{P(\text{imagem}|\overline{\text{objeto}})} > \frac{P(\overline{\text{objeto}})}{P(\text{objeto})} \text{ então um objeto (face) está presente na imagem.}$$

Para obter a função de probabilidade a posteriori, uma série de simplificações foram feitas:

- A resolução da imagem da face foi normalizada em 64 x 64 pixels;
- As imagens de faces foram subdivididas em regiões de 16 x 16 pixels e a dependência estatística entre estas não foi modelada;
- As regiões foram projetadas num espaço 12-dimensional utilizando PCA;
- A região da face foi normalizada para ter média igual a 0 e variância unitária.

Após estas simplificações a regra ficou expressa como:

$$\prod_{j=1}^{n_{magn}} \prod_{i=1}^{n_{subr}} \frac{P(q1^j|\text{objeto})P(pos^j|q2_i, \text{objeto})}{P(q1^j|\overline{\text{objeto}})} > \lambda = \frac{P(\overline{\text{objeto}})}{P(\text{objeto})}$$

n_{subr}

onde n_{magn} é o número de diferentes resoluções testadas e n_{subr} é o número de sub-regiões. O sistema foi treinado com um conjunto de 951 imagens de faces e 1552 de não faces. Para cada imagem do conjunto de faces foram sintetizadas, a partir de alterações na orientação, tamanho, intensidade e fundo, outras 120 imagens de faces.

Para cada escala de resolução j são estimados $P(q1^j|\text{objeto})$ e $P(q1^j|\overline{\text{objeto}})$.

Para tal são contadas as vezes que aparece $q1$ no conjunto de treinamento. Depois é estimado $P(pos^j|q1^j, \text{objeto})$, contando quão frequentemente cada padrão $x = q1^j$ aparece na pos^j . Finalmente, a partir deste último, se deriva $P(pos^j|q2^j, \text{objeto})$.

Já Sung (1997) cria um modelo estatístico com misturas Gaussianas, usando seis aglomerados para os padrões de faces e seis para os padrões de não-faces. Cada aglomerado é representado por uma função Gaussiana elíptica. O algoritmo percorre a imagem diversas vezes, extraindo janelas de vários tamanhos (Figura 4-4-b). Antes de ser classificada, a janela é re-dimensionada para o tamanho padrão de 19 x 19 pixels, é calculada uma medida da

diferença do padrão com o modelo de distribuição, e finalmente, este valor é avaliado por um classificador que determina se a janela contém uma face ou não (Figura 4-4-c).

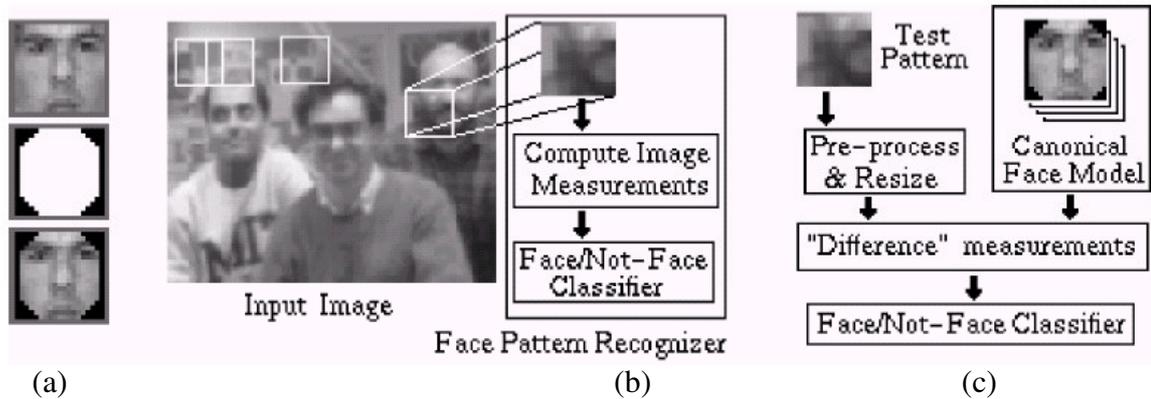


Figura 4-4 - (a) Modelo "canônico" da face, máscara para eliminar pixels próximos das bordas e modelo da face mascarado. (b) Macro-processo do algoritmo de detecção e (c) O algoritmo mais detalhado.

Fonte: (SUNG et al., 1997).

Toda sub-imagem, seja do conjunto de treinamento ou de uma imagem de entrada qualquer, é submetida ao seguinte pré-processamento:

- A sub-imagem é re-dimensionada para o tamanho 19 x 19 pixels.
- É aplicada uma máscara (Figura 4-4-a) para eliminar os pixels próximos das bordas. Assim a dimensão efetiva do padrão fica em 283 pixels.
- Correção do gradiente de iluminação; ajuda a diminuir os contrastes causados por sombras no rosto.
- Equalização de histograma; redistribui os níveis de cinza da imagem tornando-a mais nítida.

Na modelagem da distribuição de padrões da face foram utilizados 4150 padrões de face, 1067 imagens reais e o resto foram sintetizadas a partir das primeiras, por meio de pequenas mudanças na orientação, rotação e escala. Foi utilizada uma versão modificada do algoritmo de aglomeração de *k-means* (CAMPOS, 2001) para gerar aglomerados com forma elíptica no lugar da tradicional forma circular.

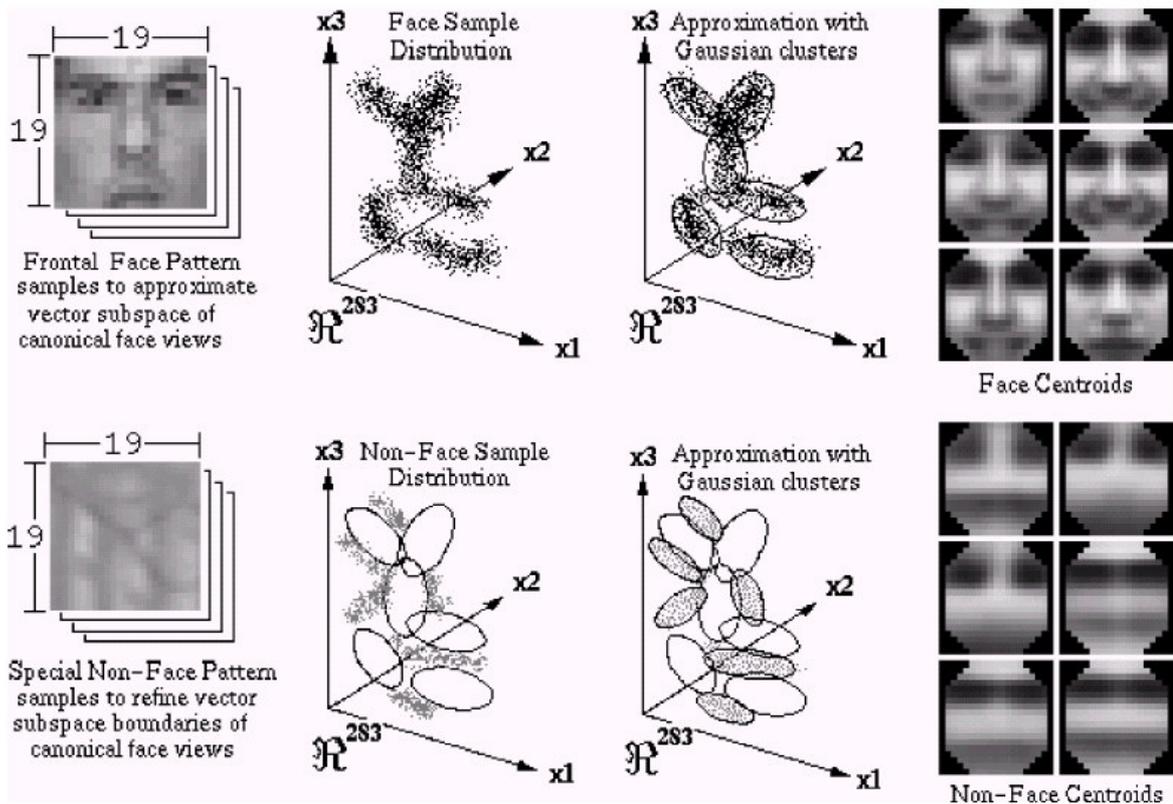


Figura 4-5 - Na fila de cima, de esquerda a direita: Exemplos de padrões de faces; distribuição dos exemplos de faces; aproximação com aglomerados Gaussianos; centroides de cada uma das classes de padrões de faces. Na fila de baixo, idem para as classes de padrões de não faces.

Fonte: (SUNG et al., 1997).

O algoritmo para obtenção do modelo é:

1. São obtidos seis centros de padrões fazendo uma quantização dos vetores das imagens de exemplo, pela distância euclidiana. O conjunto de dados de treino é dividido em seis, atribuindo cada exemplo ao padrão mais próximo, pela distância euclidiana.
2. As matrizes de covariância são inicializadas como matriz identidade.
3. Os centros dos padrões são recalculados para tornarem-se o centróide de cada aglomerado.
4. Usando o conjunto atual de seis centroides e suas respectivas matrizes de covariância, re-computa as partições, atribuindo cada exemplo ao grupo mais próximo. Neste caso utiliza a distância de Mahalanobis normalizada, definida como:

$$\mathcal{M}_n(\bar{x}, \bar{\mu}) = \frac{1}{2} \left(d \cdot \ln 2\pi + \ln |\Sigma| + (\bar{x} - \bar{\mu})^T \Sigma^{-1} (\bar{x} - \bar{\mu}) \right)$$

5. Se as partições permanecem iguais ou o número de iterações internas (passos 3 a 4) for excedido, continua no passo 5, caso contrário volta ao passo 3.
6. Re-calcula as matrizes de covariância.
7. Usando o conjunto atual de seis centroides e suas respectivas matrizes de covariância, re-computa as partições, atribuindo cada exemplo ao grupo mais próximo, utilizando a distância de Mahalanobis normalizada. Se as partições permanecem iguais ou o número de iterações externas (passos 3 a 6) for excedido, continua no passo 7, caso contrário volta ao passo 3.
8. Retorna o conjunto atual de seis centroides e suas respectivas matrizes de covariância.

Para a modelagem das classes de padrões de não-faces é seguido o mesmo algoritmo mas usando um conjunto de 6189 padrões com aparência de face (Figura 4-6).

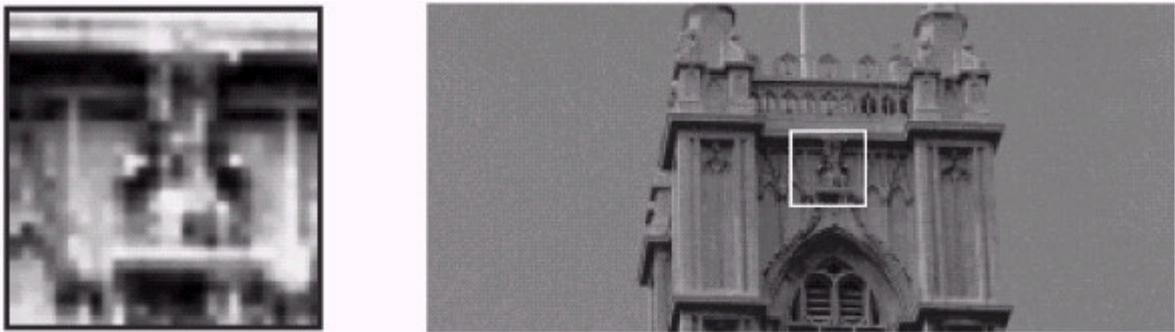


Figura 4-6 - Exemplo de uma imagem com aparência de face.

Fonte: (SUNG et al., 1997).

No processo de detecção de faces, a imagem de entrada é percorrida várias vezes, com janelas deslizantes de diferentes tamanhos. O conteúdo da janela deslizante é submetido ao pré-processamento e a continuação é obtido um conjunto de medidas de diferenças. Este conjunto é um vetor de 12 distâncias entre a sub-imagem e cada uma das classes do modelo. Cada uma destas distâncias é composta por duas medidas. Para o cálculo destas medidas são definidos, para cada classe, um sub-espço de dimensão 75, utilizando os 75 autovetores mais significativos. Define-se então E_{75} como a matriz com 75 colunas onde a i -ésima coluna é o vetor unitário na direção do i -ésimo autovetor mais significativo da classe; W_{75} é a matriz diagonal correspondente aos 75 autovetores, e Σ_{75} é a matriz de covariância do sub-espço 75-dimensional.

A primeira medida é a distância de Mahalanobis normalizada entre a sub-imagem e o centróide da classe:

$$D_1(\bar{x}, \bar{\mu}) = \frac{1}{2} \left(75 \ln 2\pi + \ln |\Sigma_{75}| + (\bar{x} - \bar{\mu})^T \Sigma_{75}^{-1} (\bar{x} - \bar{\mu}) \right)$$

A segunda medida é a distância euclidiana entre a sub-imagem e a sua projeção no sub-espço 75-dimensional:

$$D_2(\bar{x}, \bar{\mu}) = \left\| \bar{x} - \bar{x}_p \right\|^2 = \left\| (I - E_{75} E_{75}^T) (\bar{x} - \bar{\mu}) \right\|^2$$

Finalmente, o vetor de distâncias é processado por um classificador implementado por uma rede neural Multi Layer Perceptron, com 12 pares de entradas e 24 unidades ocultas.

4.3 Métodos baseados em discriminantes

O algoritmo utilizado por Rowley (1998A) (Figura 4-7) tem como principal elemento um filtro implementado com uma rede neural (NN). O filtro tem como entrada uma imagem de 20 x 20 pixels, e gera uma saída no intervalo 1 a -1, indicando a presença ou ausência de face na imagem, respectivamente. O filtro é aplicado para todos os pixels da imagem, ainda para detectar faces de diferentes tamanhos, o processo é repetido, reduzindo a imagem de entrada pelo fator 1,2. Cada região de 20 x 20 pixels, antes de ser submetida ao filtro é pré-processada, corrigindo a iluminação e aplicando uma equalização por histograma. O filtro, como já foi mencionado, está implementado com uma rede neural. Esta está formada por uma camada receptiva e uma segunda camada de unidades ocultas, dividida em três grupos: o primeiro grupo possui 4 unidades, cada uma das quais processa uma sub-região de 10 x 10 pixels; o segundo grupo possui 16 unidades, onde cada uma processa uma sub-região de 5 x 5 pixels; o terceiro grupo se compõe de 6 unidades que processam 6 sub-regiões sobrepostas de 20 x 5 pixels. Para treinar a rede foram utilizados mais de 1000 exemplos de imagens contendo faces de diferentes tamanhos, orientações, posições e intensidades. Os olhos e o centro do lábio superior foram marcados manualmente, e estes pontos foram utilizados para normalizar a imagem. Ainda, para cada imagem do conjunto de treinamento foram geradas outras 15, obtidas a partir da rotação, translação, espelhamento e escalamento da imagem original, de forma aleatória. Para o conjunto de treino de não-faces foi utilizado o seguinte procedimento, adaptado de (SUNG, apud ROWLEY, 1998A):

1. Foi criado um conjunto de 1000 imagens, geradas de forma aleatória.
2. A rede foi treinada para produzir 1 para o conjunto de imagens de faces e -1 para o conjunto criado no passo 1.

3. Foi rodado o sistema com imagens normais que não continham faces e foram coletadas as sub-imagens nas quais foi detectado incorretamente a presença de uma face.
4. Foram selecionadas até 250 das sub-imagens com falso positivo, acrescentadas ao conjunto de exemplos negativos, e repetido o processo a partir do passo 2.

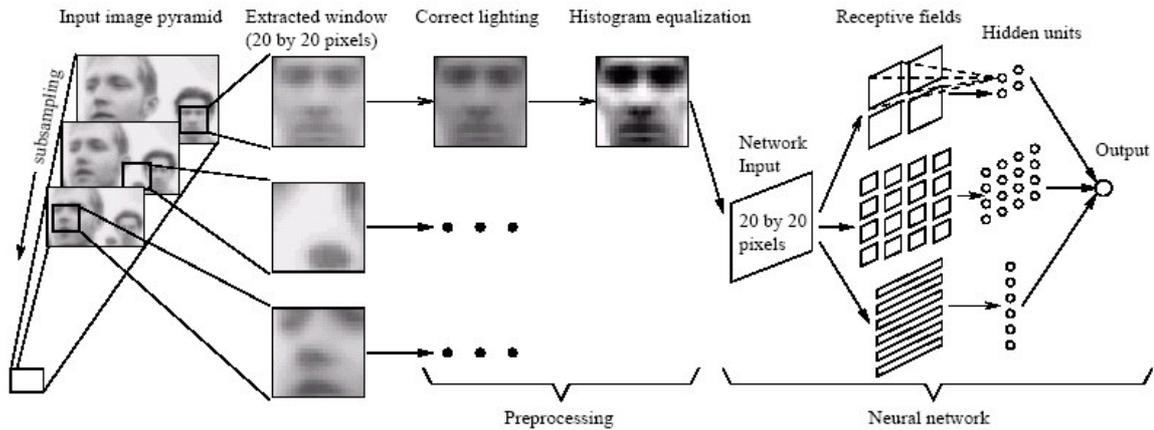


Figura 4-7 - Algoritmo básico para detecção da face.

Fonte: (ROWLEY et al., 1998A)

Com base no levantamento teórico obtido nos capítulos 3 e 4, no próximo capítulo é apresentado o sistema proposto e sua implementação.

5 SISTEMA PROPOSTO

O sistema proposto utiliza o conhecimento adquirido durante o trabalho, e apresentado nos capítulos 3, 4 e 5, para detectar uma face presente numa imagem, detectar os dois olhos, o nariz e a boca, para finalmente, definir os pontos FDPs destas características. O desenvolvimento do sistema se baseia nos seguintes critérios:

- As imagens sempre contêm uma e somente uma face;
- A face corresponde a uma face neutra, como definida no capítulo 2;
- O fundo da imagem é homogêneo;
- A iluminação é frontal e difusa;

O banco de imagens utilizado foi obtido no seguinte endereço: http://www2.imm.dtu.dk/~aam/datasets/face_data.zip e autorizadas por Stegmann(2003). É composto por 240 imagens (640 x 480) do tipo JPeg, correspondentes a rostos de 40 pessoas diferentes, 33 homens e 7 mulheres, todos sem óculos, sendo seis imagens por pessoa. Cada uma destas seis imagens responde a um padrão, conforme segue:

1. Face frontal, expressão neutra, luz difusa.
2. Face frontal, expressão “alegre”, luz difusa.
3. Face rotacionada aprox. 30° à direita da pessoa, expressão neutra, luz difusa.
4. Face rotacionada aprox. 30° à esquerda da pessoa, expressão neutra, luz difusa.
5. Face frontal, expressão neutra, foco de luz à esquerda da pessoa.
6. Face frontal, expressão arbitraria, luz difusa.

O sistema foi desenvolvido na linguagem C/C++, no ambiente de desenvolvimento Dev-Cpp, com o compilador GCC. A interface gráfica foi montada usando as bibliotecas API do Microsoft Windows®. A maior parte da mesma foi desenvolvida no âmbito do projeto Biometria-Reconhecimento da digital, do grupo GPTI desta instituição de ensino. Foram usadas também as bibliotecas: Jpeglib, para abrir e salvar imagens Jpeg; e GSL, para o cálculo dos autovalores e autovetores de uma matriz.

Para os testes finais foram obtidas imagens de 14 pessoas, entre alunos e professores da Feevale. As imagens foram obtidas em um ambiente com fundo homogêneo, da cor verde e iluminação controlada. A iluminação estava composta pela iluminação da sala, com lâmpadas do tipo fluorescente, e por uma lâmpada do tipo incandescente, de 100W, colocada a 1,5m na frente da pessoa. O equipamento utilizado para obtenção das imagens foi uma máquina fotográfica digital, do tipo Sony DSC-T5, com imagens JPeg de tamanho 640 x 480. De cada pessoa foram obtidas três fotos, conforme segue:

1. Face frontal, expressão neutra, iluminação controlada.
2. Face frontal, expressão “alegre”, iluminação controlada.
3. Face rotacionada aproximadamente 30° à direita da pessoa, expressão neutra, iluminação controlada.

5.1 Detecção do rosto

Para detecção do rosto foi utilizado um método baseado na cor da pele. O algoritmo compõe-se de duas fases; a primeira, na qual são detectados os pixels com cor semelhante à cor da pele; e a segunda, na qual são determinados os limites da área detectada. A seguir são apresentadas estas duas fases.

5.1.1 Detecção da cor da pele

Para detecção dos pixels com cor semelhante à cor da pele foi utilizado um método, baseado no apresentado por Yang (1996), que aproxima a distribuição da cor da pele no espaço de cores com um modelo gaussiano, definido pela equação:

$$P(x|\mu, \Sigma) = 2\pi^{-\frac{k}{2}} |\Sigma|^{-\frac{1}{2}} e^{-\frac{1}{2}(x-\mu)^T \Sigma^{-1}(x-\mu)}$$

onde μ é a média do espaço de amostras, Σ é a matriz de covariâncias, e k representa a quantidade de variáveis independentes. Como foi visto na seção 3.1.3, o espaço de cores é representado por duas cores normalizadas:

$$r = \frac{R}{R+G+B} \quad \text{e} \quad b = \frac{B}{R+G+B}$$

Sendo assim, o modelo tem duas variáveis independentes e, portanto, a média μ é um vetor de dimensão 2, e Σ é uma matriz 2 x 2.

O primeiro passo é fazer a coleta de amostras; com tal objetivo foi implementada uma rotina que, quando o usuário clica na imagem, captura a área circundante ao pixel onde

foi clicado. A área capturada é de 21 pixels de largura por 21 pixels de altura, para cada um destes é feita a normalização, e os dois valores resultantes, r e b , são gravados num arquivo de texto. A coleta de amostras foi feita clicando em vários lugares do rosto para várias imagens de treino. Na Figura 5-1 pode-se ver o resultado deste processo.

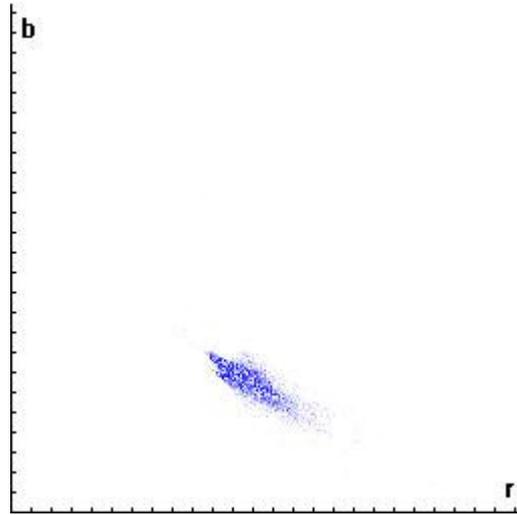


Figura 5-1 - Representação gráfica das amostras utilizadas na modelagem da cor da pele.

A continuação, são calculados os parâmetros necessários para a definição do modelo

Gaussiano $N = (\mu, \Sigma^2)$, onde $\mu = (\bar{r}, \bar{b})$, e $\Sigma = \begin{bmatrix} \sigma_{rr} & \sigma_{rb} \\ \sigma_{br} & \sigma_{bb} \end{bmatrix}$, sendo:

$$\bar{r} = \frac{1}{M} \sum_{i=1}^M r_i \text{ é a média da variável } r,$$

$$\bar{b} = \frac{1}{M} \sum_{i=1}^M b_i \text{ é a média da variável } b,$$

$$\sigma_{rr} = \frac{1}{M} \sum_{i=1}^M r_i^2 - \bar{r}^2 \text{ é a variância da variável } r,$$

$$\sigma_{bb} = \frac{1}{M} \sum_{i=1}^M b_i^2 - \bar{b}^2 \text{ é a variância da variável } b, \text{ e}$$

$$\sigma_{rb} = \sigma_{br} = \frac{1}{M} \sum_{i=1}^M r_i \cdot b_i - \bar{r} \cdot \bar{b} \text{ é a covariância entre as duas variáveis.}$$

M é o número de amostras. Estes parâmetros são armazenados num arquivo de configurações para sua posterior utilização.

Na Figura 5-2 é apresentado o modelo resultante, onde pode-se notar a semelhança com a representação gráfica das amostras, vista na Figura 5-1 - Representação gráfica das amostras utilizadas na modelagem da cor da pele.

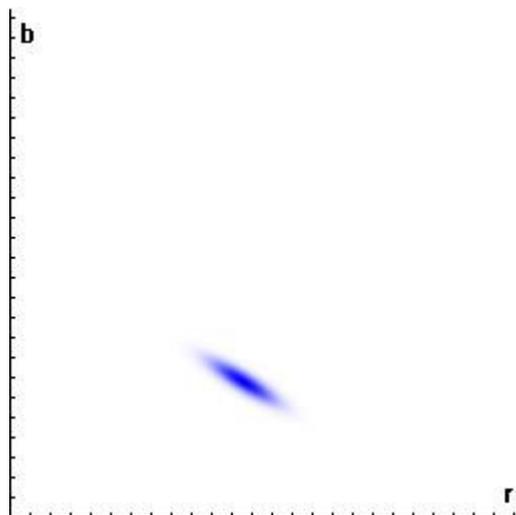


Figura 5-2 - Representação gráfica do modelo Gaussiano da cor da pele.

Uma vez criado o modelo Gaussiano, o sistema está pronto para detectar pixels com cor semelhante à cor da pele. Com esse fim, a imagem é percorrida pixel a pixel, obtendo as cores normalizadas r e b , e submetendo as mesmas ao modelo, sendo que $x = (r, b)$. O resultado deste processo pode ser visto na Figura 5-3-b. Finalmente, a imagem obtida é limiarizada, quer dizer, todo pixel que estiver por cima de um certo limiar é levado ao valor máximo (255), do contrário é zerado (Figura 5-3-c). O limiar usado no sistema é 0,2 e foi determinado empiricamente, de forma que detectasse corretamente o rosto na maior parte das imagens de treino, e ao mesmo tempo minimizasse as áreas detectadas erroneamente.



Figura 5-3 - a) imagem original; b) detecção dos pixels com cor semelhante à cor da pele; c) limiarização da imagem (b).

5.1.2 Determinação dos limites do rosto

Para determinação dos limites do rosto, o sistema se aproveita da restrição de que só haverá uma face na imagem e sobre um fundo diferenciado e homogêneo. Sendo assim, a cabeça pode ser localizada por meio dos perfis horizontal e vertical. Como já foi mencionado

na seção 3.1.2, a projeção horizontal é um vetor do tamanho da largura da imagem, e onde o valor i -ésimo corresponde à média dos valores da coluna i -ésima (Figura 5-4-b); analogamente, o perfil vertical é um vetor do tamanho da altura da imagem, com o valor i -ésimo sendo igual à média da fila i -ésima da imagem (Figura 5-4-c).

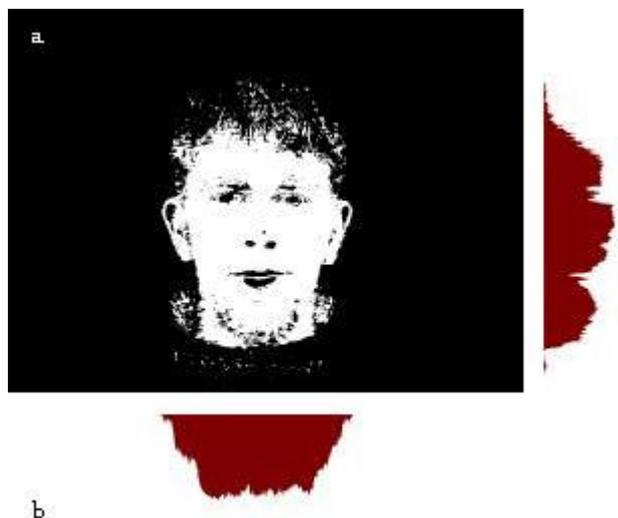


Figura 5-4 - a) Imagem limiarizada; b) Perfil horizontal; c) Perfil vertical.

Para determinar os limites direito e esquerdo do rosto, calcula-se o valor máximo do perfil horizontal. O limite esquerdo do rosto é determinado pela posição do perfil horizontal tal que, partindo da posição 0 e avançando em direção ao final, é a primeira posição para a qual o perfil supera uma certa porcentagem do valor máximo. O limite direito é determinado de forma análoga, mas partindo da posição final do perfil horizontal e avançando em direção ao início.

Analogamente, para os limites superior e inferior do rosto, calcula-se o valor máximo do perfil vertical; avançando do início em direção ao final do perfil vertical encontra-se o limite inferior e, avançando do final para o início, determina-se o limite superior. Na Figura 5-5 pode-se ver o resultado do algoritmo de detecção do rosto.

A porcentagem utilizada para determinar os quatro limites é de 30%, sendo que este valor foi determinado empiricamente, buscando detectar o rosto corretamente em todas as imagens de treinamento.

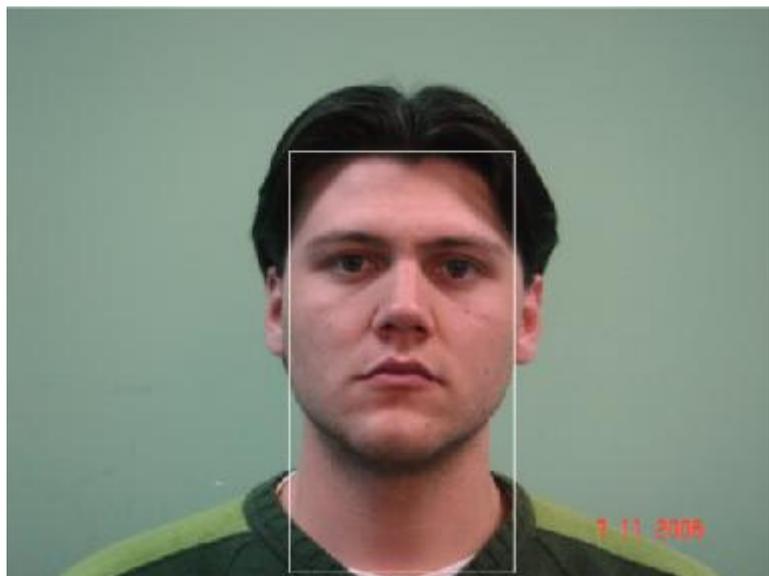


Figura 5-5 - Exemplo de detecção do rosto.

5.2 Detecção dos olhos

A idéia inicial para detecção dos olhos era usar um detector de bordas. Com esse objetivo foram desenvolvidos os detectores de bordas de Sobel, Robert Cross e de Canny. Para auxiliar nesta tarefa também foi desenvolvido um filtro Gaussiano e uma rotina de equalização da imagem pelo histograma. Foram feitos diversos testes; cada um dos diferentes detectores de bordas foi testado, limiarizando ou não a sua saída, filtrando ou não com o filtro de Gauss, equalizando ou não a imagem, considerando a imagem toda ou só a área de interesse. Nenhum destes testes apresentou um resultado satisfatório, às vezes detectando muitas bordas espúrias, às vezes não detectando bordas que eram do interesse do trabalho. Na Figura 5-6 pode-se ver alguns dos resultados obtidos com o detector de bordas de Sobel, enquanto que na Figura 5-7 são apresentados alguns dos resultados com o detector de Bordas de Robert Cross e com o detector de Canny. Mas percebe-se que tendo uma localização aproximada do olho, estas informações poderiam ser de utilidade na determinação dos pontos FDPs do padrão MPEG-4.



Figura 5-6 - Detectores de bordas de Sobel; a) sem limiarizar; b) limiarizado; c) com a imagem previamente equalizada por histograma, sem limiarizar; d) idem (c) mas limiarizado; e) imagem filtrada pelo filtro gaussiano; f) detector aplicado sobre a imagem filtrada (e).

Em vista dos resultados terem sido tão pobres, decidiu-se partir para outra técnica para detecção dos olhos. A técnica escolhida foi o PCA. A idéia é encontrar uma base canônica que represente o espaço dos olhos, e utilizar a distância ao espaço de características (DFFS) para determinar qual a região da imagem que mais se assemelha a um olho.



Figura 5-7 – a) Detector de bordas de Robert Cross; b) Idem (a), com a imagem previamente equalizada pelo histograma; c) Detector de bordas de Canny; d) Idem (c) , com a imagem previamente equalizada pelo histograma.

5.2.1 PCA

Como já foi apresentado na seção 4.1, o PCA (Principal Component Analysis) é uma técnica baseada na aparência, e como tal necessita de treinamento. O treinamento consiste em capturar áreas de imagens contendo um olho na mesma. Todas as áreas capturadas devem ser do mesmo tamanho, o qual fica armazenado no arquivo de configuração. Para capturar estas amostras foi utilizada uma rotina semelhante à usada no treinamento do modelo da cor da pele; o usuário clica no centro do olho, e a área circundante ao ponto que foi clicado é capturada. A diferença, porém, é que no lugar de gravar as informações num arquivo de texto, foi gerada uma imagem para cada amostra e gravada dentro de uma pasta específica para as amostras do olho. A área de captura utilizada tem 56 pixels de largura e 30 de altura. Estas dimensões foram determinadas empiricamente, buscando uma área que compreendesse todas as amostras de olho utilizadas. Na Figura 5-8 podem-se ver algumas das imagens capturadas.



Figura 5-8 - Exemplos de amostras utilizadas na determinação do espaço vetorial do olho direito.

Feita a captura de amostras, já pode ser calculado o espaço vetorial. O primeiro passo é o cálculo da imagem média das amostras (Figura 5-9) definida como:

$$\Psi = \frac{1}{M} \sum_{n=1}^M \Gamma_n \text{ sendo } M \text{ a quantidade de amostras e } \Gamma_i \text{ a } i\text{-ésima amostra.}$$



Figura 5-9 - Imagem média das amostras capturadas para o olho direito.

O objetivo é encontrar os vetores ortonormais, u_n , que descrevam totalmente o espaço de imagens dos olhos. Como foi visto na seção 4.1, $u_l = A.v_l$, com $l = 1..M$, sendo M a quantidade de amostras, $A = \{\Phi_1, \Phi_2 \dots \Phi_M\}$ e v_l são os autovetores de A . Já $\Phi_i = \Gamma_i - \Psi$, é a diferença entre a amostra i -ésima e a média das amostras.

Então, o primeiro a ser feito é criar a matriz A , cujas colunas são os vetores Φ_i . Depois é calculado o produto entre a matriz transposta de A e A ; e como A é uma matriz $M \times N$, A^T será uma matriz $N \times M$, e o produto resultante será a matriz $L: M \times M$. A seguir são calculados os autovetores de L^7 , obtendo-se M vetores de dimensão M . Os autovetores obtidos são pré-multiplicados pela matriz A , obtendo-se M vetores de dimensão N . Para finalizar, os vetores são normalizados, dividindo-os pelo módulo da matriz A . Na Figura 5-10 é apresentado o conjunto de *eigeneyes* obtido, após ter selecionado só os mais significativos. O critério adotado foi manter os autovetores com autovalor maior que o 5% do maior autovalor, e excluir o resto. A matriz de autovetores é gravada num arquivo binário, onde primeiro são gravadas as duas dimensões da matriz e na seqüência são gravados os vetores, um a continuação do outro.

⁷ Para esta tarefa foi utilizada a biblioteca GSL (GNU Scientific Library)

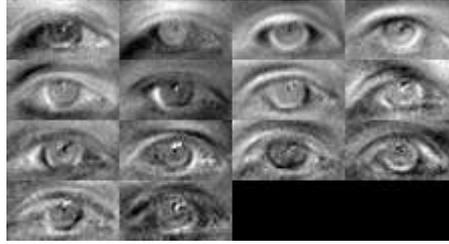


Figura 5-10 – Representação dos *eigeneyes* do olho direito.

Uma vez criado o espaço da característica, o sistema está pronto para fazer a detecção da mesma. O método de detecção utilizado está baseado na distância ao espaço da característica (DFFS), usada por Demirel (1996) como método de classificação. A DFFS é a distância entre o vetor correspondente a uma área da imagem, do mesmo tamanho usado na criação do espaço, e a projeção da mesma no espaço.

Para a detecção da característica, a imagem é percorrida pixel a pixel, com uma janela do mesmo tamanho da área usada na criação do espaço. Esta área é projetada no espaço, quer dizer, primeiro é representada no espaço e logo é reconstruída pelo espaço. Representar uma imagem no espaço significa que são calculados os pesos w_k para cada vetor da base. Como foi visto na seção 4.1:

$$w_k = u_k^T (\Gamma - \Psi), k = 1..M', \text{ onde:}$$

- w_k é o peso calculado para o k-ésimo *eigeneye*;
- u_k é o k-ésimo *eigeneye*;
- Ψ é a média das imagens do conjunto de treino;
- M' é o número de *eigeneyes* mais significativos, que possuem a maior variância da base de dados.

Já reconstruir a imagem significa somar os vetores da base, ponderados pelos pesos calculados anteriormente, ou seja:

$$\Phi' = \sum_{i=1}^{M'} w_i u_i, \text{ onde } \Phi' \text{ é a imagem } \Phi \text{ projetada no espaço dos olhos, } w_i \text{ é o peso}$$

para o i-ésimo *eigeneye*, e u_i é o i-ésimo *eigeneye*.

Uma vez projetada a imagem, é calculada a distância euclidiana:

$$\varepsilon_e^2 = \|\Phi - \Phi'\|^2 = \sum_{i=1}^N (\Phi_i - \Phi'_i)^2$$

O ponto central da área com a menor distância ao espaço é considerado o ponto onde foi detectado o olho. Na Figura 5-11 é apresentado o resultado da rotina de detecção dos olhos.



Figura 5-11 - Detalhe do resultado da detecção dos olhos.

5.3 Detecção do nariz

A técnica usada na detecção do nariz foi PCA, a mesma utilizada na detecção dos olhos. As características de imagens de treino são as mesmas, tendo como diferença a região selecionada no rosto para o treino. Na Figura 5-12 podem-se ver algumas das amostras utilizadas na criação do espaço vetorial do nariz, e na Figura 5-13 apresenta-se a média destas amostras. Já na Figura 5-14 estão representadas os autovetores (*eigennoses*) encontrados para o espaço vetorial do nariz.



Figura 5-12 - Exemplos de amostras utilizadas na determinação do espaço vetorial do nariz.



Figura 5-13 - Imagem média das amostras capturadas para o nariz.

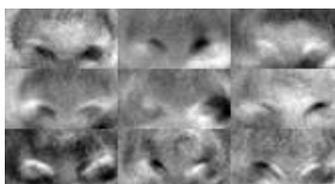


Figura 5-14 - Representação do *eigennoses* do nariz.

Na Figura 5-15 é apresentado o resultado da rotina de detecção do nariz.



Figura 5-15 - Detalhe do resultado da detecção do nariz.

5.4 Detecção da boca

Para detecção da boca a área de interesse é delimitada por:

- Por cima: 40% da distância entre a linha dos olhos e o nariz, por baixo do nariz;
- Por baixo: a distância entre a linha dos olhos e o nariz, por baixo do limite superior;
- Pela esquerda: 20% da distância entre os olhos à esquerda do olho direito;
- Pela direita: 20% da distância entre os olhos à direita do olho esquerdo;

Na Figura 5-16-a pode-se ver a área de interesse assim delimitada. A seguir é aplicado nesta área, a componente horizontal do detector de bordas de Sobel, limiarizado (Figura 5-16-b). Depois é calculado o perfil vertical na imagem das bordas (Figura 5-16-d); note-se o máximo do perfil coincidindo com a boca. Desta forma, então, a coordenada y do centro da boca é determinada pelo máximo do perfil vertical. Finalmente, para determinar a coordenada x do centro da boca, os limites superior e inferior da área de interesse são redefinidos, passando a ser 10 pixels por cima e 10 pixels por baixo da coordenada y encontrada. Na nova área é calculado o perfil horizontal (Figura 5-16-c). Procedendo de forma semelhante a como foi determinada a área do rosto (seção 5.1.2), são determinados os limites direito e esquerdo da boca. A coordenada x do centro da boca fica determinada, então, pelo ponto meio entre o limite esquerdo e o direito.

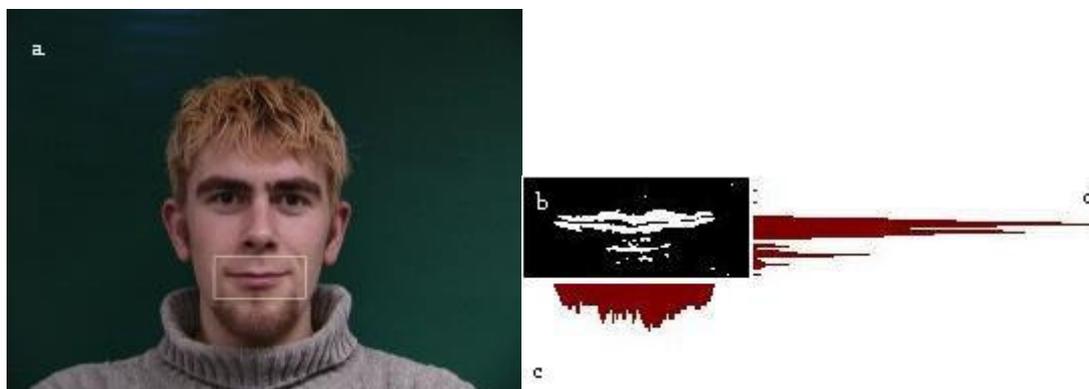


Figura 5-16 – a) Região de interesse da para detecção da boca; b) Detecção de bordas horizontais na região da boca; c) Perfil horizontal da imagem das bordas; d) Perfil vertical da imagem das bordas.

Na Figura 5-17 pode-se comprovar o resultado final desta rotina.

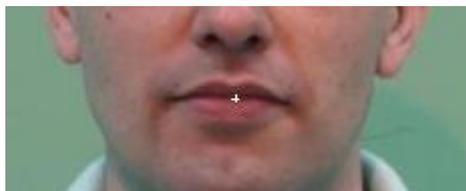


Figura 5-17 - Detalhe do resultado da detecção da boca.

5.5 Detecção dos pontos FDPs

Os pontos FDPs são determinados a partir das características já encontradas, utilizando as componentes horizontal e vertical do detector de bordas de Sobel. As coordenadas dos pontos FDPs detectados são gravadas num arquivo de texto com o seguinte formato:

- para cada ponto uma linha;
- cada linha estará formada pelo número do ponto FDP, um sinal de igual (=) e as duas coordenadas do ponto separadas por vírgula;

5.5.1 Pontos FDPs do nariz

Para detecção dos pontos 9.1 a 9.5 e o 9.15, é considerada a área de interesse determinada por: por cima, pela coordenada y do ponto do nariz detectado; por baixo, metade da distância entre o nariz e a boca, por baixo do limite superior; pela esquerda, pela coordenada x do olho direito; pela direita, pela coordenada x do olho esquerdo. Dentro desta área é calculado o perfil horizontal; a coordenada x do ponto 9.15 é determinada como o máximo entre os dois grandes mínimos do perfil; já as coordenadas x dos pontos 9.4 e 9.5 pelos flancos externos dos dois mínimos. A seguir, na mesma área, é calculado o perfil vertical. Nele aparece uma grande depressão correspondente às narinas; as coordenadas y dos pontos 9.4, 9.5 e 9.15 ficam determinadas pelo flanco inferior da depressão no perfil vertical.

Os pontos 9.1 e 9.2 são encontrados procurando por duas bordas verticais perto dos pontos 9.5 e 9.4 respectivamente. O ponto 9.3 fica com a mesma coordenada x do ponto 9.15 e a mesma coordenada y do ponto central do nariz.

Para determinar os pontos 9.6 e 9.7, primeiro é calculado o ponto meio entre os dois olhos, para depois buscar duas bordas verticais, próximas, em direção aos olhos.

Finalmente, o ponto 9.12 é o ponto meio entre o ponto meio dos olhos e o ponto 9.3; o ponto 9.14 é o ponto meio entre os pontos 9.6 e 9.2; e o ponto 9.13 é o ponto meio entre os pontos 9.7 e 9.1. Na Figura 5-18 pode-se observar o resultado desta rotina.

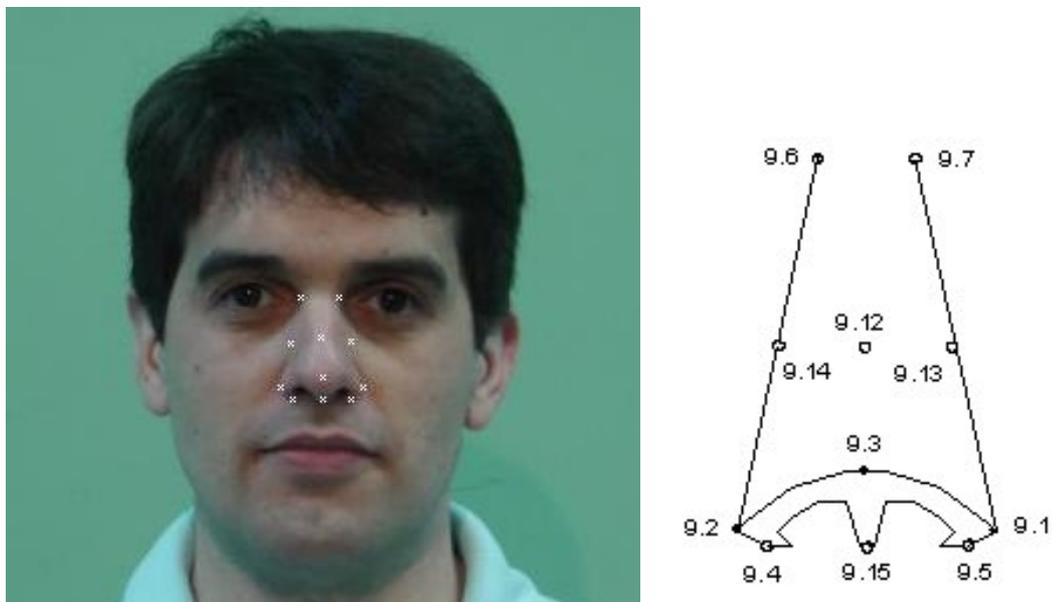


Figura 5-18 - Detalhe da detecção dos pontos FDPs do nariz.

5.5.2 Pontos FDPs do olho⁸

Para detectar os pontos FDPs do olho é considerada uma área de interesse centrada no ponto detectado do olho, e 20% maior que a área de detecção.

Primeiro é calculado o perfil horizontal. O mesmo apresenta um mínimo local correspondente à íris. Encontrando o ponto central desta região, encontra-se a coordenada x para os pontos 3.2, 3.4, 3.6, 3.10 e 3.14. Procedendo da mesma forma, mas com o perfil vertical, é encontrada a coordenada y do ponto 3.6. A seguir é aplicada a componente horizontal do detector de bordas de Sobel. Partindo do ponto 3.6 para cima, a primeira borda encontrada corresponde ao ponto 3.2, e a seguinte ao 3.14. Da mesma forma, do ponto 3.6 para baixo, encontra-se o ponto 3.4, na primeira borda, e o 3.10 na segunda.

Ainda sobre a imagem das bordas horizontais, são encontrados os pontos 3.8 e 3.12, procurando os extremos de uma borda forte, usando os perfis vertical e horizontal. Na Figura 5-19 é apresentado um exemplo do resultado desta rotina.

⁸ Somente são apresentados os pontos FDPs do olho direito, já que os do olho esquerdo são obtidos de forma análoga.

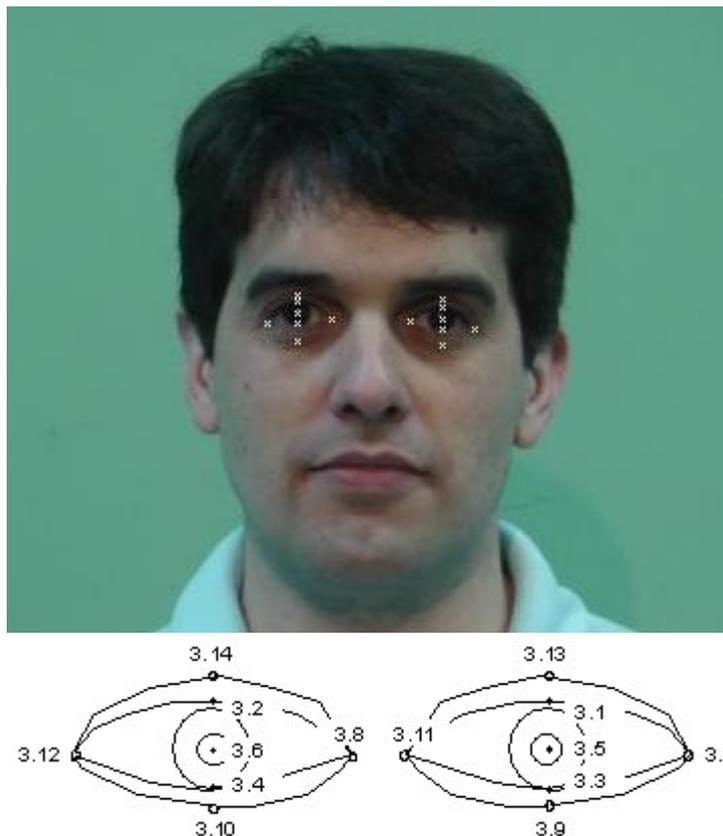


Figura 5-19 - Detalhe da detecção dos pontos FDPs dos olhos.

5.5.3 Pontos FDPs da boca

Para detecção dos pontos FDPs da boca a imagem é convertida a tons de cinza. A seguir são apresentados os passos para detecção dos mesmos:

1. Inicia-se com o ponto detectado da boca como o pixel atual.
2. Considera-se a coluna de pixels formada pelo pixel atual, os cinco pixels superiores e os 5 inferiores.
3. Procura-se o pixel com menor valor entre os onze.
4. O pixel atual passa a ser o pixel à direita do mínimo.
5. Repete-se os passos 2 a 4 até que a diferença entre dois mínimos contíguos seja menor que 50. Quando isso acontecer, terá se encontrado o ponto 2.4.
6. O ponto detectado do nariz volta a ser o pixel atual.
7. Considera-se a coluna de pixels formada pelo pixel atual, os cinco pixels superiores e os 5 inferiores.
8. Procura-se o pixel com menor valor entre os onze.
9. O pixel atual passa a ser o pixel à esquerda do mínimo.
10. Repete-se os passos 7 a 9 até que a diferença entre dois mínimos contíguos seja menor que 50. Quando isso acontecer, terá se encontrado o ponto 2.5.

O ponto 2.2, que para uma face neutra é o mesmo que o 2.3, corresponde ao pixel com coordenada x igual à média entre as coordenadas x dos pontos 2.4 e 2.5, e coordenada y igual à coordenada y do pixel mínimo para esse y . De forma análoga, os pontos 2.7 e 2.9 são encontrados na metade do caminho entre os pontos 2.5 e 2.2, e os pontos 2.6 e 2.8, na metade do caminho do 2.4 e o 2.2.

Finalmente, os FDPs externos são encontrados por detecção de bordas. Os pontos 8.4 e 8.5, numa face neutra, são iguais aos pontos 2.4 e 2.5 respectivamente. O ponto 8.1 é determinado pela primeira borda horizontal encontrada logo acima do ponto 2.2, e o ponto 8.2, pela primeira borda encontrada logo abaixo do 2.3. O ponto 8.8 fica na metade do caminho entre os pontos 8.4 e o 8.2, seguindo a borda inferior. Da mesma forma, o ponto 8.7 fica na metade do caminho entre os pontos 8.2 e o 8.3. Os pontos 8.9 e 8.10 são as terminações das duas bordas horizontais encontradas logo acima do ponto 8.1. Para terminar, de forma análoga aos pontos 8.8 e 8.7, o ponto 8.6 é encontrado na metade do caminho entre os pontos 8.4 e 8.9, e o 8.5 na metade do caminho entre os pontos 8.3 e 8.10. Na Figura 5-20 pode-se ver o resultados da detecção dos pontos FDPs da boca.

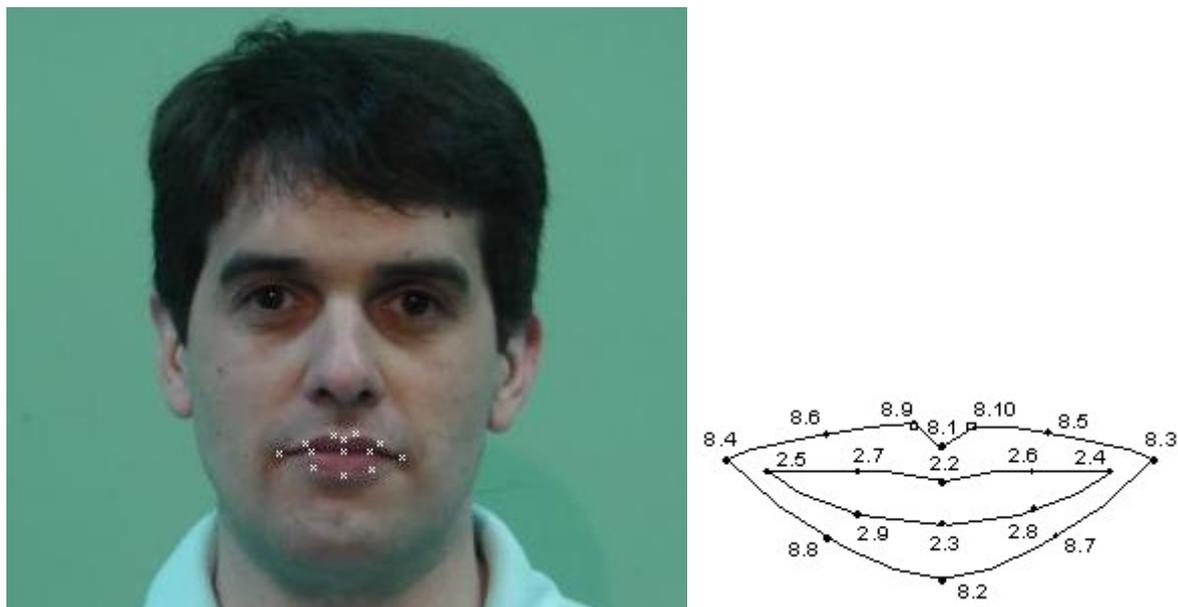


Figura 5-20 - Detalhe da detecção dos pontos FDPs da boca.

5.6 O arquivo de configuração

O sistema utiliza o arquivo *ProgramaFaces.ini* para armazenar as configurações e os parâmetros obtidos no treinamento. No Anexo C podemos ver o conteúdo do mesmo. A seguir é feita a descrição do conteúdo de cada campo:

[Cores], seção que contém informações da rotina de detecção pela cor da pele;

ArquivoAmostras, nome do arquivo onde são gravadas as amostras de cor da pele;

MediaX, média da primeira variável;

MediaY, média da segunda variável;

VarianciaX, variância da primeira variável;

VarianciaY, variância da segunda variável;

CovarianciaXY, covariância entre as duas variáveis;

[Olho_Direito], seção com informações para o treinamento do olho direito;

Pasta, nome da pasta onde são gravadas as amostras;

Numero_amostras, quantidade de amostras;

Largura, largura da amostra;

Altura, altura da amostra;

[Olho_Esquerdo], seção com informações para treinamento do olho esquerdo;

Pasta, nome da pasta onde são gravadas as amostras;

Numero_amostras, quantidade de amostras;

Largura, largura da amostra;

Altura, altura da amostra;

[Nariz], seção com informações para treinamento do nariz;

Pasta, nome da pasta onde são gravadas as amostras;

Numero_amostras, quantidade de amostras;

Largura, largura da amostra;

Altura, altura da amostra;

5.7 Implementação

A primeira implementação do sistema começava com a detecção do rosto. Na seqüência era aplicado o detector do olho direito, tendo como área de busca a região definida pelo detector de rosto. Neste ponto foram encontradas as duas primeiras dificuldades: o processo de detecção do olho era muito lento (de 20 a 30 seg.), e a taxa de acerto era muito baixa (11 de 37 imagens usadas no treinamento do olho direito).

O primeiro problema foi resolvido reduzindo a imagem, porém, isto é feito internamente pelo sistema, sendo que os pontos detectados na imagem reduzida são projetados novamente na imagem original. Esta alteração acarretou também a redução das amostras, tendo dois caminhos: reduzir as amostras da mesma forma que eram reduzidas as imagens, ou re-capturar as amostras nas imagens reduzidas. Foi optado pelo primeiro caminho por ser mais rápido, já que foi implementada uma rotina para fazer isso pelo sistema, e pelo fato de que capturar as amostras em imagens menores poderia induzir mais erros. O valor

utilizado como fator de redução foi encontrado empiricamente, visando deixar a detecção o suficientemente rápida, mas sem comprometer de mais a qualidade das imagens; sobre tudo das amostras, que ficaram, com o valor empregado (0.43), com dimensões de 24 x 12 pixels.

Para a redução da imagem foi criada uma rotina que, conceitualmente falando, coloca uma grade cobrindo a imagem original, com tantas células como pixels terá a imagem re-dimensionada. Ou seja, se a imagem re-dimensionada terá M pixels de largura por N de altura, a grade sobreposta terá M células na horizontal e N na vertical. O valor de cada pixel da nova imagem será então, a média dos pixels cobertos pela correspondente célula da grade; sendo que os pixels fracionados são considerados como tais. Na prática, é criada uma janela de tamanho igual à relação de tamanhos entre a imagem original e a re-dimensionada, e com esta é percorrida toda a imagem, de forma não sobreposta, calculando a média dos pixels cobertos por ela.

O segundo problema foi em parte resolvido limitando a região de busca pela metade superior da região do rosto (Figura 5-21-a). Desta forma o sistema passou a detectar o olho direito em 32 das 37 imagens utilizadas no treinamento.

Com esses resultados já considerados aceitáveis, continuou-se com a detecção das outras características. Para detecção do olho esquerdo, limitou-se a região de busca anterior, pela coordenada x do olho direito. Quer dizer, a região de busca do olho esquerdo ficava delimitada: por cima, pelo limite superior da região do rosto; por baixo, pela linha que divide a região do rosto ao meio na vertical; pela esquerda, pela coordenada x do olho direito; e pela direita pelo limite direito da região do rosto (Figura 5-21-b).

Detectados o rosto e os dois olhos, o passo seguinte era a detecção do nariz. Para tal, a região de busca ficava definida da seguinte forma: por cima, pela coordenada y do olho esquerdo; por baixo, pelo limite inferior da região do rosto; pela esquerda, pela coordenada x do olho direito; e pela direita pela coordenada x do olho esquerdo (Figura 5-21-c).

Finalmente, a procura da boca era feita na região determinada por: por cima, 40% da distância entre os olhos e o nariz considerados logo abaixo do nariz; por baixo, a distância entre os olhos e o nariz considerados logo abaixo do limite superior; pela esquerda, 20% da distância entre os olhos considerados à esquerda do olho direito; pela direita, 20% da distância entre os olhos considerados à direita do olho esquerdo (Figura 5-21-d).



Figura 5-21 – Regiões de busca da primeira implementação; a) Região de busca do olho direito; b) Região de busca do olho esquerdo; c) Região de busca do nariz; d) Região de busca da boca.

O resultado para o conjunto do processo ficou aproximadamente igual ao obtido só com o olho direito, 31 das 37 imagens usadas no treinamento. Das seis imagens que deram problema, três foi porque a linha da metade da região do rosto ficava por cima dos olhos, ou tocando os mesmos, e já que esta linha é o limite inferior da região de busca dos olhos, os mesmos terminaram por não serem detectados. Por sua vez, a causa dos olhos não ficarem completamente na parte superior do rosto foi, ou porque o cabelo foi detectado com a cor da pele, ou porque a pessoa possuía barba, a qual não foi detectada como pele, qualquer um dos dois motivos provocou a elevação relativa da linha do meio do rosto. Este problema já evidencia uma grande limitação na utilização do detector pela cor da pele.

A seguir foi feito um teste para comparar o desempenho entre o detector do olho direito e o detector do nariz, aplicados na região detectada do rosto, sem nenhuma restrição adicional. Neste teste verificou-se que o desempenho do detector do nariz era melhor que o do olho direito. O primeiro falhou na detecção em 5 das 37 imagens usadas para o treinamento, enquanto que o detector do olho direito falhou em 10, 5 nas quais não detectou nada, e 5 nas que detectou o olho esquerdo no lugar do direito. Esta comprovação levou a investir no detector do nariz. Visando detectar o 100% das imagens de treinamento, foram aplicadas

algumas restrições na área de busca. Depois de algumas tentativas, chegou-se a duas restrições com as quais foi atingido o objetivo de 37 das 37 imagens de treino detectadas. As restrições à área de busca, que foram determinadas empiricamente, são: por cima, metade da largura da área do rosto considerado logo abaixo do limite superior da área do rosto; por baixo, 130% da largura da região do rosto considerados logo abaixo do limite superior da região do rosto; pela direita e pela esquerda ficou igual aos respectivos limites da região do rosto (Figura 5-22-a).

Em vista do sucesso dos testes feitos com o detector do nariz, foi alterada a ordem de execução dos detectores por PCA e, por conseguinte, também as regiões de busca da cada um.

A segunda e definitiva implementação começa igual à primeira; detectando a área do rosto pela cor da pele. A seguir é aplicado o detector do nariz na região de busca antes mencionada. Uma vez detectado o nariz, o olho direito é procurado na região determinada por: por cima, pelo limite superior da região do rosto; por baixo, pela coordenada y do nariz; pela esquerda pelo limite esquerdo da região do rosto; e pela direita, pela coordenada x do nariz (Figura 5-22-b). Por sua vez a área de busca do olho esquerdo fica determinada por: por cima e por baixo, igual aos correspondentes da região de busca do olho direito; pela esquerda, pela coordenada x do nariz; e pela direita, pelo limite direito da região do rosto (Figura 5-22-c). Finalmente, a boca é detectada usando como área de busca a mencionada na seção 5.4 (Figura 5-22-d).



Figura 5-22 – Regiões de busca da segunda implementação; a) Região de busca do nariz; b) Região de busca do olho direito; c) Região de busca do olho esquerdo; d) Região de busca da boca.

Assim implementado, obteve-se sucesso em 35 das 37 imagens usadas no treinamento, sendo que nas duas imagens que apresentaram problemas só não foi detectado corretamente o olho esquerdo. Na Figura 5-23 pode-se ver o fluxograma do processo de detecção da face e das características faciais.

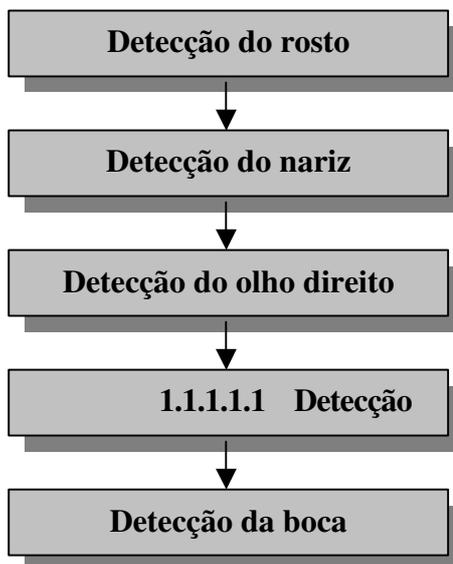


Figura 5-23 - Fluxograma do processo de detecção da face e suas características.

5.8 Testes

Para os testes foi utilizado o banco de imagens criado pelo autor, descrito na seção 5. Nos primeiros testes, efetuados sobre as 34 imagens compostas pelas duas primeiras séries do banco, somente duas foram completamente detectadas. Na verdade, o principal problema com estas imagens é que o sistema estava detectando de forma pobre a área do rosto.

O seguinte passo então, foi re-treinar o modelo do espaço da cor da pele. Mas precisamente, novas amostras foram acrescentadas ao modelo já existente. Foram capturadas três amostras para cada uma das oito primeiras imagens do banco. Na Figura 5-24 são comparadas as representações gráficas das amostras, antes e após o re-treinamento. Na mesma pode-se observar principalmente a expansão do aglomerado em direção ao eixo b , quer dizer que são pontos que tem mais b e menos r . Isto pode ser explicado pela diferença nas condições de luz deste novo conjunto de imagens, em comparação com o conjunto original. Mas uma conclusão pode ser tirada antes de prosseguir com os testes; pixels que não eram detectados como da cor da pele no primeiro banco de imagens usado, agora podem vir a ser detectados, o que poderia provocar algum problema.

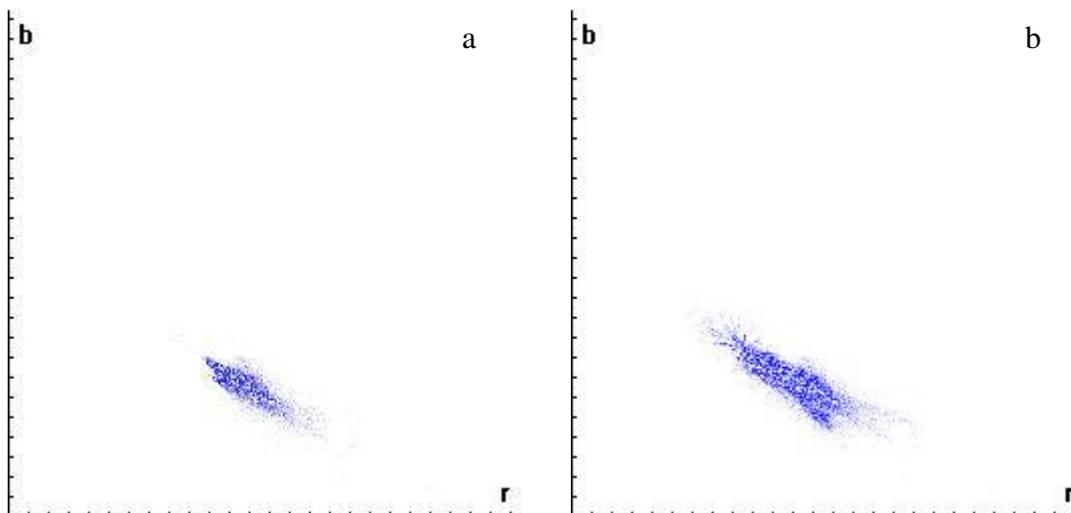


Figura 5-24 - a) Representação gráfica do conjunto de amostras original; b) Representação gráfica do novo conjunto de amostras.

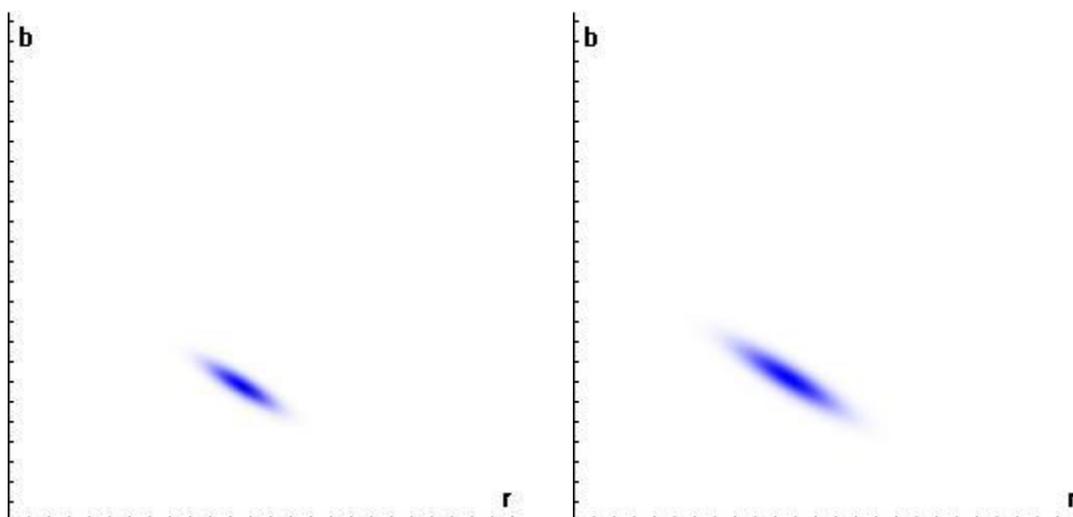


Figura 5-25 - a) Representação gráfica do modelo da cor da pele original; b) Representação gráfica do novo modelo da cor da pele.

O resultado com o sistema re-treinado foi de 17 imagens detectadas sobre 34 testadas, 8 da primeira série e 7 da segunda. O que se pode notar na maioria das imagens que não foram detectadas é que tem uma tonalidade mais avermelhada que as imagens detectadas. Estas podem corresponder, observando o gráfico na Figura 5-24-b, à língua que sobre-sai do aglomerado principal, na parte inferior direita do gráfico, e que claramente fica excluída do modelo gerado (Figura 5-25-b). Esta situação justifica o método usado por Jebara (1997), o qual cria um modelo de misturas gaussianas para modelar a cor da pele, precisamente para lidar com as diferentes condições de iluminação.

Para confirmar a hipótese de que o problema nas imagens não detectadas estava na cor avermelhada, foram criadas duas novas rotinas; uma para diminuir em 10% a cor vermelha da imagem, e a outra para aumentá-la. Das 17 imagens que não tinham sido detectadas, 13 conseguiram ser detectadas depois de aplicar duas vezes a rotina de diminuição da cor vermelha.

As rotinas de redução e de aumento da cor vermelha foram desenvolvidas utilizando o que em processamento de imagens é chamado de Função de Transformação Pontual. Uma transformação pontual é aquela onde o novo valor do pixel depende somente do valor atual do mesmo, podendo ser expressa como $p' = F(p)$, onde p' é o novo valor do pixel, p é o valor atual do pixel e F é a função de transformação. Como o valor do pixel é um valor discreto entre 0 e 255^9 , a função pode ser representada com um vetor de 256 posições, onde a posição i corresponde ao valor atual do pixel (p) e o valor contido na posição i , ao novo valor ($F(p)$).

⁹ Considerando uma imagem com pixel de 8 bits ou, como é o caso, considerando só uma cor numa imagem com pixel de 24 bits.

Desta forma, a transformação é aplicada percorrendo a imagem e substituindo o valor de cada pixel, pelo valor correspondente no vetor função.

Voltando às imagens do banco original, comprovou-se, já na primeira imagem, a mudança de comportamento do detector de cor. A roupa da pessoa nesta imagem foi levemente detectada, não chegando a atrapalhar a detecção, mas marcando uma possível fonte de problemas.

Foi feito, então, um último teste. Voltou-se ao conjunto de amostras inicial, e foi recriado o modelo da cor da pele. A continuação, foram repetidos os testes sobre as 34 imagens do segundo banco, aplicando, antes da detecção, a rotina de diminuição da cor vermelha se a imagem tinha um tom avermelhado, ou a rotina de aumento da cor vermelha se a mesma tinha um tom mais azulado. O resultado obtido foi o mesmo que o anterior, 30 das 34 imagens.

Uma possível solução a este problema poderia ser um sistema de balanceamento de branco automático. Para isto, as imagens de treino teriam que ter uma imagem de um quadro branco, uma imagem para cada condição de iluminação diferente, o que permitiria a normalização da cor nas imagens.

Nas figuras a seguir podem-se ver alguns exemplos dos resultados obtidos. Na Figura 5-26 são apresentados exemplos obtidos sobre as duas primeiras séries do banco de imagens criado para este trabalho, onde podem se observar resultados positivos em rostos com barba e óculos (b), rosto de uma pessoa de sexo feminino (c), e numa cabeça levemente inclinada. Já na Figura 5-27 são apresentados alguns casos especiais; na imagem (a) pode-se observar o resultado positivo numa imagem com parte da face oculta; as imagens (b) e (c) são dois casos de rostos com olhos fechados, no primeiro caso falhou a detecção dos olhos, e no segundo o resultado foi positivo.



Figura 5-26 - Exemplos de resultados obtidos.



Figura 5-27 - exemplos de resultados obtidos com casos especiais.

6 CONCLUSÕES

Atualmente tem-se observado um crescente número de investigações na área de análise de imagens faciais, podendo ser destacadas diversas aplicações no uso desta tecnologia, como: sistemas de reconhecimento facial, recuperação de imagens baseada em conteúdo, codificação de vídeo, vídeo conferência, vigilância de multidões, interfaces humano-computador inteligentes, etc. O primeiro passo para a detecção de faces e suas características em imagens é a identificação da presença de uma ou mais faces (se houver), e determinar a localização e tamanho das mesmas.

No problema específico da modelagem facial, deve-se também detectar as características faciais e seus pontos característicos, para posteriormente serem utilizados para calibração do modelo, colagem da textura e animação facial. Neste caso, uma das abordagens utilizadas frequentemente baseia-se no padrão MPEG-4, estudado em profundidade no capítulo 2.

Para facilitar o entendimento do trabalho, as abordagens utilizadas pelos principais investigadores foram divididas em dois grupos: métodos baseados em características e métodos baseados na aparência.

Conforme apresentado no capítulo 3, os métodos baseados em características dependem fortemente de conhecimentos prévio. Neste capítulo os métodos foram divididos nos baseados em características de baixo nível e em análise de características. A informação obtida pelos primeiros, por si só não tem muita significância, mas estes são utilizados pelos métodos do segundo grupo como ferramenta para a detecção das características. Os dois grupos utilizam informação prévia, porém a informação utilizada por cada um deles é de natureza diferente. Os métodos baseados em características de baixo nível utilizam características comuns à maioria dos objetos, como forma e cor, assim, a íris do olho pode ser detectada como uma borda com forma circular. Já os métodos de análise de características utilizam informações espaciais e dimensionais das características faciais; assim o nariz é procurado embaixo e no meio dos dois olhos.

No capítulo 4 foram apresentados os métodos baseados na aparência. Nestas abordagens não se necessita o conhecimento prévio. Na sua maioria estão baseados em algoritmos estatísticos para implementação de um classificador face/não-face, e por tanto necessitam de uma fase de treinamento, na qual são apresentados ao sistema imagens de exemplo para ambas as classes.

Conforme pode ser observado no decorrer do trabalho, os algoritmos são bastantes complexos. Entendê-los e escrever sobre os mesmo requer o conhecimento de várias áreas da ciência, exigindo um tempo considerável de pesquisa.

No capítulo 5 foi apresentado o sistema proposto para detecção da face e de suas características (olhos, nariz e boca).

Para a detecção da face foi usado um método baseado na cor da pele. Os resultados obtidos foram bons, porém, o método utilizado mostrou algumas limitações, como a sensibilidade às condições de iluminação da imagem e a possibilidade de falsos positivos em regiões do fundo da imagem. Estas limitações indicam a necessidade de, em trabalhos futuros, utilizar outro método para a detecção da face ou, acompanhar o mesmo com outros métodos que possibilitem superar estas limitações.

Na detecção dos olhos e do nariz foi utilizado um método baseado no PCA, usando o DFFS como forma de determinar qual a região da imagem que mais se aproxima com a característica procurada. Os resultados obtidos também foram bons, no entanto, a alta taxa de falsos positivos obrigou a acompanhar esta técnica com informação da estrutura da face, delimitando as regiões de busca de cada característica. Uma possibilidade para melhorar o desempenho dos detectores individuais pode ser a utilização de um conjunto de imagens de treino maior.

A última característica, a boca, foi detectada usando o método de detecção de bordas aliado às técnicas de projeção na horizontal e na vertical. Os resultados obtidos foram bons, porém, uma iluminação diferente poderia chegar a atrapalhar os resultados.

Finalmente, seguindo o padrão MPEG-4 foram detectados os pontos FDPs das características faciais, utilizando para isto os pontos detectados nas fases anteriores conjuntamente com o detector de bordas de Sobel, com algumas aproximações.

No final deste trabalho, confirma-se a afirmação feita na introdução do mesmo, sobre a dificuldade desta tarefa, aparentemente tão simples para o ser humano. Reafirma-se, então, a necessidade de continuar pesquisando novas técnicas para aprimorar ainda mais o sistema desenvolvido.

REFERÊNCIAS BIBLIOGRÁFICAS

- ABÁSULO, Maria José (1998). **Creación de mundos virtuales con VRML 97**. Disponível em <http://dmi.uib.es/~abasolo/cursosvrml/indice.htm>. Acessado em 24 mar. 2006.
- ANDRADE, Ernesto L. (1997). **Sistema de Identificação Pessoal Utilizando Técnicas de Reconhecimento e Verificação Facial Automática**. Tese de Mestrado. Universidade estadual de Campinas.
- CAMPOS, Teófilo E. de (2001). **Técnicas de Seleção de Características com Aplicações em Reconhecimento de Faces**. Dissertação de Mestrado. Universidade de São Paulo.
- CANNY, John (1986). **A computational approach to edge detection**. In IEEE: Transactions on Pattern Analysis and Machine Intelligence, 8(6):679-698.
- CRAW, Ian; TOCK, David; BENNETT, Alan (1992). **Finding face features**. In Proceedings of the Second European Conference on Computer Vision, pp. 92-96.
- DEMIREL, Hasan; CLARKE, Thomas; CHEUNG, Peter (1996). **Adaptive automatic facial feature segmentation**. In IEEE Proceedings of 2nd International Conference on Automatic Face and Gesture Recognition, pp. 277-282.
- DEMPSTER, Arthur; LAIRD, Nan; RUBIN, Donald (1997). **Maximum Likelihood from Incomplete Data via the EM Algorithm**. In Proceedings of Journal of the Royal Statistical Society, Series B, 39(1), pp. 1-38.
- DÉNIZ, O.; CASTRILLÓN, M.; HERNÁNDEZ, M. (2001). **Face Recognition Using Independent Component Analysis and Support Vector Machines**. In Proceedings of the IX Spanish Symposium on Pattern Recognition and Image Analysis Volume II, pp. 1-6.
- EKMAN, Paul; OSTER, Harriet (1979). **Facial expressions of emotion**. Annual Review of Psychology, vol. 30, pp 527-554.
- FERIS, Rogério S.; CAMPOS, Teófilo E. de; CESAR, Roberto M. (2000). **Detection and Tracking of Facial Features in Video Sequences**. In Proceedings of Lecture Notes of Artificial Intelligence, vol. 1793, pp 197-206.
- HJELMÅS, Erik; LOW, Boon Kee (2001). **Face Detection: A Survey**. In Proceedings of Computer Vision and Image Understanding, vol. 83, pp 236-274.
- ISO/IEC JTC1/SC29/WG11 (1998). **SNHC, Information Technology – Generics Coding of Audio – Visual Objects, Part. 2: Visual**. ISO/IEC 14496-2, Final Draft of International Standard. Version of 13. Nov. 1998, ISO/IEC JTC1/SC29/WG11 N2502a. Atlantic City.
- JEBARA, Tony; PENTLAND, Alex (1997). **Parametrized structure from motion for 3D adaptive feedback tracking of faces**. In Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, pp. 144-150.

- KOTROPOULOS, Constantine; PITAS, Ioannis (1997). Rule-based face detection in frontal views. In Proceedings of International Conference on Acoustic, Speech and Signal Processing.
- LEE, T. S. (1996). **Image Representation Using 2D Gabor Wavelets**. In Proceedings of IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 18, n° 10.
- NIKOLAIDIS, Athanasios; PITAS, Ioannis (2000). **Facial feature extraction and pose determination**. Pattern Recognition 33:1783-1791.
- POPOVICI, Vlad (2004). **Kernel-based classifiers with applications to face detection**. Thèse pour l'obtention du grade de docteur ès Sciences. École Polytechnique Fédérale de Loussane.
- REISFELD, Daniel; WOLFSON, Haim; YESHURUN, Yehezquel (1995). **Context-free attentional operators: the generalized symmetry operator**. International Journal of Computer Vision (14):57-74.
- ROWLEY, Henry; BALUJA, Shumeet; KANADE, Takeo (1998A). **Neural network-based face detection**. IEEE Transactions on Pattern Analysis and Machine Intelligence 20(1):23-38.
- SCASSELLATI, Brian (1998). **Eye finding via face detection for a foveated, active vision system**. In Proceedings of the 15th National Conference on Artificial Intelligence.
- SCHNEIDERMAN, Henry; KANADE, Takeo (1998). **Probabilistic modeling of local appearance and spatial relationships for object recognition**. In Proceedings of IEEE International Conference on Computer Vision and Pattern Recognition, pp. 45-51.
- SENIOR, Andrew (1999). **Face and feature finding for a face recognition system**. In Proceedings of the Second International Conference on Audio- and Video-Based Person Authentication.
- SIROVICH, L.; KIRBY, M. (1987). **Low-dimensional procedure for the characterization of human faces**. Journal of the Optical Society of America A 4(3):519-524.
- SMERALDI, F.; CARMONA, O.; BIGÜN, J (2000). **Saccadic search with Gabor features applied to eye detection and real-time head tracking**. Image and Vision Computing 18(4):323-329.
- SUN Microsystems (1999). **Java 3D API Specification, version 1.1.2**. Disponível em <http://java.sun.com/products/java-media/3D/forDevelopers/j3dguide/j3dTOC.doc.html>. Acesso em 20/03/2006.
- SUNG, Kha-Kay; POGGIO, Tomaso (1997). **Example-based learning for view-based human face detection**. IEEE Transactions on Pattern Analysis and Machine Intelligence 20(1):39-51.
- VOGELHUBER, Veronika; SCHMID, Cordelia (2000). **Face detection based on generic local descriptors and spatial constraints**. In International Conference on Pattern Recognition, vol. I, pp. 1084-1087.
- WEBER, M. et al. (2000). **Viewpoint-invariant learning and detection of human heads**. In Proceedings of the Fourth IEEE International Conference on Automatic Face and Gesture Recognition, pp. 20-27.
- YANG, Guangzheng; HUANG T. S. (1994). **Human Face Detection in a Complex Background**. In Proceedings of Pattern Recognition, vol. 27, pp. 53-63.
- YANG, Jie; WAIBEL, Alex (1996). **A real-time face tracker**. In Proceedings of the Third Workshop on Applications of Computer Vision, pp. 142-147.

YOW, Kin; CIPOLLA, Roberto (1997). **Feature-based human face detection**. *Image and Vision Computation* 15(9):713-735.

ANEXOS

Anexo A Descrição dos pontos FDP

	FDP	COR.	POS.	DESCRIÇÃO	X	Y	Z
C A R A	E X T E R N O S	2.1		FL	Ponta do queixo	7.1	
		2.11	2.12	FLE	Lado esquerdo queixo		
		2.12	2.11	FLD	Lado direito do queixo		
		2.13	2.14	FLE	Lado esquerdo da mandíbula		
		2.14	2.13	FLD	Lado direito da mandíbula		
		10.7	10.8	FLE	Contato inferior entre a orelha e a cara, lado esquerdo		
		10.8	10.7	FLD	Contato inferior entre a orelha e a cara, lado direito		
		10.9	10.10	FLE	Contato superior entre a orelha e a cara, lado esquerdo		
		10.10	10.9	FLD	Contato superior entre a orelha e a cara, lado direito		
		11.1		FL	Borde superior central da cara	7.1	
		11.2	11.3	FLD	Borde superior direito da cara		
		11.3	11.2	FLE	Borde superior esquerdo da cara		
		I N T E R	7.1		L	Centro de rotação da cabeça	
2.10				FL	Centro do queixo	7.1	
5.1	5.2			FLE	Centro da bochecha esquerda		
5.2	5.1			FLD	Centro da bochecha direita		
5.3	5.4			FLE	Osso da bochecha esquerda		
5.4	5.3	FLD	Osso da bochecha direita				
B O C A	I N T E R N A	2.2		FL	Ponto médio superior dos lábios internos	7.1	
		2.3		FL	Ponto médio inferior dos lábios internos	7.1	
		2.4	2.5	FLE	Extremo esquerdo dos lábios internos		
		2.5	2.4	FLD	Extremo direito dos lábios internos		
		2.6	2.7	FLE	Ponto médio esquerdo dos lábios internos superior	$(2.2 + 2.4) / 2$	
		2.7	2.6	FLD	Ponto médio direito dos lábios internos superior	$(2.2 + 2.5) / 2$	
		2.8	2.9	FLE	Ponto médio esquerdo dos lábios internos superior	$(2.3 + 2.4) / 2$	
		2.9	2.8	FLD	Ponto médio direito dos lábios internos superior	$(2.3 + 2.5) / 2$	
E X T E R N A	8.1		F	Meio externo contorno lábio superior	7.1		
		8.2		FL	Meio externo contorno lábio inferior	7.1	
		8.3	8.4	FLE	Extremo esquerdo lábio superior externo		
		8.4	8.3	FLD	Extremo direito lábio superior externo		
		8.5	8.6	FLE	Médio lábio superior lado esquerdo	$(8.3 + 8.1) / 2$	
		8.6	8.5	FLD	Médio lábio superior lado direito	$(8.4 + 8.1) / 2$	

	FDP	COR.	POS.	DESCRIÇÃO	X	Y	Z	
	8.7	8.8	FLE	Médio lábio inferior lado esquerdo	$(8.3 + 8.2) / 2$			
	8.8	8.7	FLD	Médio lábio inferior lado direito	$(8.4 + 8.2) / 2$			
	8.9	8.10	FLD	Ponta mais alta lábio superior lado direito				
	8.10	8.9	FLE	Ponta mais alta lábio superior lado esquerdo				
LÍNGUA	6.1		FL	Ponta da língua	7.1			
	6.2		F	Centro da língua	7.1			
	6.3	6.4	FLE	Lado esquerdo da língua			6.2	
	6.4	6.3	FLD	Lado direito da língua			6.2	
DENTES	9.8	9.9	F	Parte dos dentes mais perto dos lábios superiores	7.1			
	9.9	9.8	F	Parte dos dentes mais perto dos lábios inferiores	7.1			
	9.10	9.11	F	Ponta dos dentes superiores	7.1			
	9.11	9.10	F	Ponta dos dentes inferiores	7.1			
OLHOS	DETERECHE	3.2	3.1	FLD	Centro superior interno do olho direito	$(3.8 + 3.12) / 2$		
		3.4	3.3	FLE	Centro inferior interno do olho esquerdo	$(3.8 + 3.12) / 2$		
		3.6	3.5	F	Centro da pupila olho direito			
		3.10	3.9	FLD	Centro inferior olho direito	$(3.8 + 3.12) / 2$		
		3.8	3.8	FLE	Extremo direito do olho esquerdo			
		3.12	3.7	FLD	Extremo direito do olho direito			
		3.14	3.13	FLD	Centro superior do olho direito	$(3.8 + 3.12) / 2$		
	IZQUIE	3.1	3.2	FLE	Centro superior interno do olho esquerdo	$(3.7 + 3.11) / 2$		
		3.3	3.4	FLE	Centro inferior interno do olho esquerdo	$(3.7 + 3.11) / 2$		
		3.5	3.6	F	Centro da pupila olho esquerdo			
		3.7	3.12	FLE	Extremo esquerdo do olho esquerdo			
		3.11	3.11	FLD	Extremo esquerdo do olho direito			
		3.9	3.10	FLE	Centro inferior do olho esquerdo	$(3.7 + 3.11) / 2$		
		3.13	3.14	FLE	Centro superior do olho esquerdo	$(3.8 + 3.12) / 2$		
NARIZ	9.1	9.2	FLE	Extremo esquerdo borde do nariz				
	9.2	9.1	FLD	Extremo direito borde do nariz				
	9.3		FL	Ponta do nariz	7.1			
	9.4	9.5	F	Parte extrema inferior narina direita				
	9.5	9.4	F	Parte extrema inferior narina esquerda				
	9.6	9.7	FLD	Parte do nariz mais perto do olho direito				
	9.7	9.6	FLE	Parte do nariz mais perto do olho esquerdo				
	9.12		FL	Metade superior do nariz	7.1	$(9.6 + 9.3) / 2$		
	9.13	9.14	FLE	Metade superior do nariz, borde esquerda		$(9.6 + 9.3) / 2$		

	FDP	COR.	POS.	DESCRIÇÃO	X	Y	Z	
	9.14	9.13	FLD	Metade superior do nariz borde direito		$(9.6 + 9.3) / 2$		
	9.15		FL	Borde inferior metade do nariz	7.1			
O R E	D	10.2	10.1	FLD	Borde superior orelha direita			
	I	10.4	10.3	FLD	Metade orelha direita		$(10.2 + 10.6) / 2$	
	R	10.6	10.5	FLD	Borde inferior orelha direita			
	E		10.1	10.2	FLE	Borde superior orelha esquerda		
			10.3	10.4	FLE	Metade orelha esquerda		$(10.1 + 10.5) / 2$
		S	10.5	10.6	FLE	Borde inferior orelha esquerda		
L S O A B R A N C.	Q	4.2	4.1	FLD	Ponta esquerda sobrelha olho direito			
	R		4.4	4.3	FLD	Centro sobrelha olho direito	$(4.12 + 4.56) / 2$	
			4.6	4.5	FLD	Ponta direita sobrelha olho direito		
	E S Q		4.1	4.2	FLE	Ponta direita sobrelha olho esquerdo		
			4.3	4.4	FLE	Centro sobrelha olho esquerdo	$(4.1 + 4.5) / 2$	
			4.5	4.6	FLE	Ponta esquerda sobrelha olho esquerdo		
	CA	11.4		FL	Topo central do crânio	7.1		
BE	11.5		FL	Topo central do cabelo	11.4			
LO	11.6		F	Parte de trás do crânio	7.1			

Fonte : ISO/IEC JTC1/SC29/WG11 (1998).

Anexo B Descrição dos pontos FAP

Nome FAP	Descrição FAP	Unidades	Unid- orBi- dir	Pos- moção	Gr- p	FDP subgrp num	Quant- idade Paso taman- ho	Min/Max I-Frame valores quantitati- vos	Min/Max P-Frame valores quantitati- vos
viseme	Conjunto de valores que determinam a mistura de dois visemas para este frame (p.e. pbm, fv, th)	na	na	na	1	na	1	Viseme_ble nd: +63	Viseme_bl end: +-63
expressão	Conjunto de valores que determinam a mistura de duas expressões faciais	na	na	na	1	na	1	expression_ intensity1, expression_ intensity2: +63	expression_ _intensity1 , expression_ _intensity2 : +-63
open_jaw	Deslocamento vertical da mandíbula (no afeta à abertura da boca)	MNS	U	down	2	1	4	+1080	+360
lower_t_midlip	Deslocamento vertical da parte central interna do lábio superior	MNS	B	down	2	2	2	+600	+180
raise_b_midlip	Deslocamento vertical da parte central interna do lábio inferior	MNS	B	up	2	3	2	+1860	+600
stretch_l_cornerlip	Deslocamento horizontal da parte interna do canto esquerdo do lábio	MW	B	left	2	4	2	+600	+180
stretch_r_cornerlip	Deslocamento horizontal da parte interna do canto direito do lábio	MW	B	right	2	5	2	+600	+180
lower_t_lip_lm	Deslocamento vertical do ponto meio entre o canto esquerdo e o centro interno do lábio superior	MNS	B	down	2	6	2	+600	+180
lower_t_lip_rm	Deslocamento vertical do ponto meio entre o canto direito e o centro interno do lábio superior	MNS	B	down	2	7	2	+600	+180
raise_b_lip_lm	Deslocamento vertical do ponto meio entre o canto esquerdo e o centro interno do lábio inferior	MNS	B	up	2	8	2	+1860	+600
raise_b_lip_rm	Deslocamento vertical do ponto meio entre o canto direito e o centro interno do lábio inferior	MNS	B	up	2	9	2	+1860	+600
raise_l_cornerlip	Deslocamento vertical do canto esquerdo interno do lábio	MNS	B	up	2	4	2	+600	+180
raise_r_cornerlip	Deslocamento vertical do canto direito interno do lábio	MNS	B	up	2	5	2	+600	+180

Nome FAP	Descrição FAP	Unidades	Uni- Bi- dir	Pos- moção	Gr p	FDP subgrp num	Quant idade Paso taman ho	Min/Max I-Frame valores quantitati vos	Min/Max P-Frame valores quantitati vos
thrust_jaw	Deslocamento em profundidade da mandíbula	MNS	U	forward	2	1	1	+600	+180
shift_jaw	Deslocamento lado a lado da mandíbula	MW	B	right	2	1	1	+1080	+360
push_b_lip	Deslocamento em profundidade da parte central do lábio inferior	MNS	B	forward	2	3	1	+1080	+360
push_t_lip	Deslocamento em profundidade da parte central do lábio superior	MNS	B	forward	2	2	1	+1080	+360
depress_chin	Movimento para cima e com compressão do queixo (como na expressão de tristeza)	MNS	B	up	2	10	1	+420	+180
close_t_l_eyelid	Deslocamento vertical da parte superior esquerda da pálpebra	IRISD	B	down	3	1	1	+1080	+600
close_t_r_eyelid	Deslocamento vertical da parte superior direita da pálpebra	IRISD	B	down	3	2	1	+1080	+600
close_b_l_eyelid	Deslocamento vertical da parte inferior esquerda da pálpebra	IRISD	B	up	3	3	1	+600	+240
close_b_r_eyelid	Deslocamento vertical da parte inferior direita da pálpebra	IRISD	B	up	3	4	1	+600	+240
yaw_l_eyeball	Orientação horizontal do globo ocular esquerdo	AU	B	left	3	na	128	+1200	+420
yaw_r_eyeball	Orientação horizontal do globo ocular direito	AU	B	left	3	na	128	+1200	+420
pitch_l_eyeball	Orientação vertical do globo ocular esquerdo	AU	B	down	3	na	128	+900	+300
pitch_r_eyeball	Orientação vertical do globo ocular direito	AU	B	down	3	na	128	+900	+300
thrust_l_eyeball	Deslocamento em profundidade do globo ocular esquerdo	ES	B	forward	3	na	1	+600	+180
thrust_r_eyeball	Deslocamento em profundidade do globo ocular direito	ES	B	forward	3	na	1	+600	+180
dilate_l_pupil	Dilatação da pupila esquerda	IRISD	B	growing	3	5	1	+420	+120
dilate_r_pupil	Dilatação da pupila direita	IRISD	B	growing	3	6	1	+420	+120
raise_l_i_eyebrow	Deslocamento vertical da parte interna da sobrancelha esquerda	ENS	B	up	4	1	2	+900	+360
raise_r_i_eyebrow	Deslocamento vertical da parte interna da sobrancelha direita	ENS	B	up	4	2	2	+900	+360

Nome FAP	Descrição FAP	Unidades	Uni-orBi dir	Pos-moção	Grp	FDP subgrp num	Quantidade Paso tamanho	Min/Max I-Frame valores quantitativos	Min/Max P-Frame valores quantitativos
raise_l_m_eyebrow	Deslocamento vertical da parte central da sobrancelha esquerda	ENS	B	up	4	3	2	+900	+360
raise_r_m_eyebrow	Deslocamento vertical da parte central da sobrancelha direita	ENS	B	up	4	4	2	+900	+360
raise_l_o_eyebrow	Deslocamento vertical da parte externa da sobrancelha esquerda	ENS	B	up	4	5	2	+900	+360
raise_r_o_eyebrow	Deslocamento vertical da parte externa da sobrancelha direita	ENS	B	up	4	6	2	+900	+360
squeeze_l_eyebrow	Deslocamento horizontal da sobrancelha esquerda	ES	B	right	4	1	1	+900	+300
squeeze_r_eyebrow	Deslocamento horizontal da sobrancelha direita	ES	B	left	4	2	1	+900	+300
puff_l_cheek	Deslocamento horizontal da bochecha esquerda	ES	B	left	5	1	2	+900	+300
puff_r_cheek	Deslocamento horizontal da bochecha direita	ES	B	right	5	2	2	+900	+300
lift_l_cheek	Deslocamento vertical da bochecha esquerda	ENS	U	up	5	3	2	+600	+180
lift_r_cheek	Deslocamento vertical da bochecha direita	ENS	U	up	5	4	2	+600	+180
shift_tongue_tip	Deslocamento horizontal da ponta da língua	MW	B	right	6	1	1	+1080	+420
raise_tongue_tip	Deslocamento vertical da ponta da língua	MNS	B	up	6	1	1	+1080	+420
thrust_tongue_tip	Deslocamento em profundidade da ponta da língua	MW	B	forward	6	1	1	+1080	+420
raise_tongue	Deslocamento vertical da língua	MNS	B	up	6	2	1	+1080	+420
tongue_roll	Dobramento da língua em forma de U	AU	U	concave upward	6	3, 4	512	+300	+60
head_pitch	Angulo de pitch da cabeça desde o inicio da coluna	AU	B	down	7	na	170	+1860	+600
head_yaw	Angulo de yaw da cabeça desde o inicio da coluna	AU	B	left	7	na	170	+1860	+600
head_roll	Angulo de roll da cabeça desde o inicio da coluna	AU	B	right	7	na	170	+1860	+600
lower_t_midlip_o	Deslocamento vertical da parte central externa do lábio superior	MNS	B	down	8	1	2	+600	+180
raise_b_midlip_o	Deslocamento vertical da parte central externa do lábio inferior	MNS	B	up	8	2	2	+1860	+600

Nome FAP	Descrição FAP	Unidades	UniorBidir	Pos-moção	Grp	FDP subgr num	Quantidade Paso tamanho	Min/Max I-Frame valores quantitativos	Min/Max P-Frame valores quantitativos
stretch_l_cornerlip_o	Deslocamento horizontal da parte externa do canto esquerdo	MW	B	left	8	3	2	+600	+180
stretch_r_cornerlip_o	Deslocamento horizontal da parte externa do canto direito	MW	B	right	8	4	2	+600	+180
lower_t_lip_lm_o	Deslocamento vertical do ponto meio entre o canto esquerdo e o centro externo do lábio superior	MNS	B	down	8	5	2	+600	+180
lower_t_lip_rm_o	Deslocamento vertical do ponto meio entre o canto direito e o centro externo do lábio superior	MNS	B	down	8	6	2	+600	+180
raise_b_lip_lm_o	Deslocamento vertical do ponto meio entre o canto esquerdo e o centro externo do lábio inferior	MNS	B	up	8	7	2	+1860	+600
raise_b_lip_rm_o	Deslocamento vertical do ponto meio entre o canto direito e o centro externo do lábio inferior	MNS	B	up	8	8	2	+1860	+600
raise_l_cornerlip_o	Deslocamento vertical da parte externa do canto esquerdo do lábio	MNS	B	up	8	3	2	+600	+180
raise_r_cornerlip_o	Deslocamento vertical da parte externa do canto direito do lábio	MNS	B	up	8	4	2	+600	+180
stretch_l_nose	Deslocamento horizontal da parte esquerda do nariz	ENS	B	left	9	1	1	+540	+120
stretch_r_nose	Deslocamento horizontal da parte direita do nariz	ENS	B	right	9	2	1	+540	+120
raise_nose	Deslocamento vertical da ponta do nariz	ENS	B	up	9	3	1	+680	+180
bend_nose	Deslocamento horizontal da ponta do nariz	ENS	B	right	9	3	1	+900	+180
raise_l_ear	Deslocamento vertical da orelha esquerda	ENS	B	Up	10	1	1	+900	+240
raise_r_ear	Deslocamento vertical da orelha direita	ENS	B	Up	10	2	1	+900	+240
pull_l_ear	Deslocamento horizontal da orelha esquerda	ENS	B	left	10	3	1	+900	+300
pull_r_ear	Deslocamento horizontal da orelha direita	ENS	B	right	10	4	1	+900	+300

Fonte: ISO/IEC JTC1/SC29/WG11 (1998)

Anexo C Arquivo de configurações

[Cores]
ArquivoAmostras=AmostrasCor.txt
MediaX=0.454205
MediaY=0.264877
VariânciaX=0.001808
VariânciaY=0.000799
CovariânciaXY=-0.001043

[Olho_Direito]
Pasta=Olho_Direito
Numero_amostras=39
Largura=24
Altura=12

[Olho_Esquerdo]
Pasta=Olho_Esquerdo
Numero_amostras=40
Largura=24
Altura=12

[Nariz]
Pasta=Nariz
Numero_amostras=40
Largura=24
Altura=12