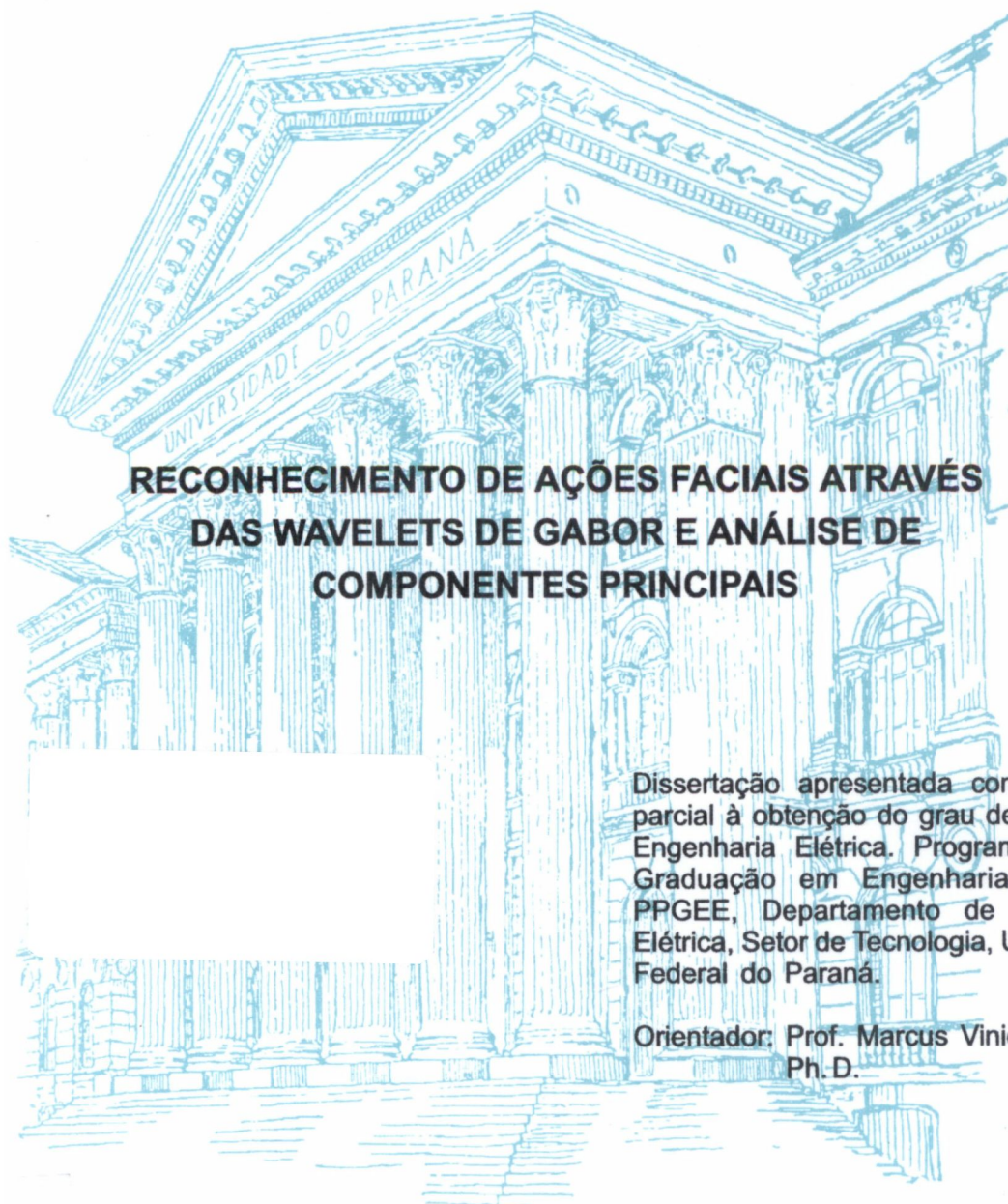


JULIANO JOÃO BAZZO



**RECONHECIMENTO DE AÇÕES FACIAIS ATRAVÉS  
DAS WAVELETS DE GABOR E ANÁLISE DE  
COMPONENTES PRINCIPAIS**

Dissertação apresentada como requisito parcial à obtenção do grau de Mestre em Engenharia Elétrica. Programa de Pós-Graduação em Engenharia Elétrica - PPGEE, Departamento de Engenharia Elétrica, Setor de Tecnologia, Universidade Federal do Paraná.

Orientador: Prof. Marcus Vinicius Lamar,  
Ph.D.

CURITIBA

2003

JULIANO JOÃO BAZZO

**RECONHECIMENTO DE AÇÕES FACIAIS ATRAVÉS  
DAS WAVELETS DE GABOR E ANÁLISE DE  
COMPONENTES PRINCIPAIS**

Dissertação apresentada como requisito parcial à obtenção do grau de Mestre em Engenharia Elétrica. Programa de Pós-Graduação em Engenharia Elétrica - PPGEE, Departamento de Engenharia Elétrica, Setor de Tecnologia, Universidade Federal do Paraná.

Orientador: Prof. Marcus Vinicius Lamar,  
Ph.D.

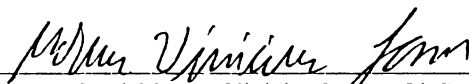
CURITIBA

2003

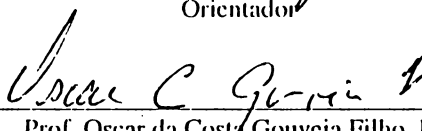
# RECONHECIMENTO DE AÇÕES FACIAIS ATRAVÉS DAS WAVELETS DE GABOR E ANÁLISE DE COMPONENTES PRINCIPAIS

JULIANO JOÃO BAZZO

Dissertação aprovada como requisito parcial para obtenção do grau de Mestre no  
Programa de Pós-Graduação em Engenharia Elétrica da Universidade Federal do  
Paraná

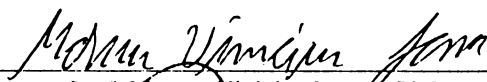


Prof. Marcus Vinicius L. Amar, Ph.D.  
Orientador

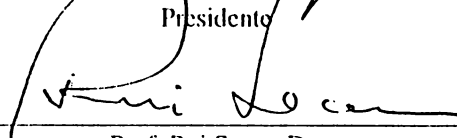


Prof. Oscar da Costa Gouveia Filho, Dr.  
Coordenador do Programa em Pós-Graduação em Engenharia Elétrica

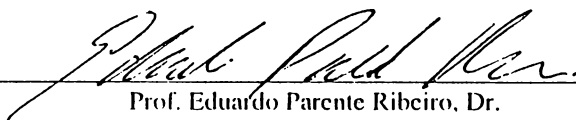
## Banca Examinadora



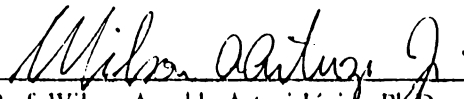
Prof. Marcus Vinicius L. Amar, Ph.D.  
Presidente



Prof. Rui Scara, Dr.



Prof. Eduardo Parente Ribeiro, Dr.



Prof. Wilson Arnaldo Artuzi Júnior, Ph.D.

Curitiba, 12 de dezembro de 2003

O Mestre na arte da vida faz pouca distinção entre o seu trabalho e o seu lazer, entre a sua mente e o seu corpo, entre a sua educação e a sua recreação, entre o seu amor e a sua religião. Ele dificilmente sabe distinguir um corpo do outro. Ele simplesmente persegue sua visão de excelência em tudo que faz, deixando para os outros a decisão de saber se está trabalhando ou se divertindo. Ele acha que está sempre fazendo as duas coisas simultaneamente.

Texto Zen-Budista

*Ao meu pai Osmir*

## Agradecimentos

À coordenação do Programa de Pós-Graduação em Engenharia Elétrica da UFPR pela realização satisfatória de seu trabalho.

Ao Professor Marcus V. Lamar pela orientação, pelos questionamentos realizados na hora certa, pela paciência e por sempre estar disposto a sugerir, revisar e tirar dúvidas referentes a este trabalho.

Agradeço ao Instituto de Robótica da Universidade de Carnegie Mellon dos Estados Unidos pela concessão do banco de dados de expressões faciais.

Ao Flávio Farias, bolsista de iniciação científica, pela dedicação na segmentação das imagens e compreensão da sua importância.

Ao LACTEC pelo uso dos computadores e ao Ivan Chueiri, pela liberação das minhas atividades diárias nos momentos decisivos dessa dissertação.

Agradeço aos professores Eduardo P. Ribeiro e Horácio Tertuliano por me incentivarem desde a graduação.

Agradeço em especial à Kelly, minha futura esposa, pela paciência, apoio e compreensão nos momentos de incertezas e preocupações que surgiram nesses 2 anos.

À minha mãe Inete e ao meu irmão Delcir pela atenção e palavras de incentivo que recebi, mesmo estando tantos anos longe de casa.

A todos os meus amigos que de alguma forma contribuíram para o sucesso desse trabalho.

## SUMÁRIO

|  |      |
|--|------|
| LISTA DE ABREVIACÕES .....   | V    |
| LISTA DE FIGURAS.....  | VI   |
| LISTA DE TABELAS.....  | VIII |
| RESUMO .....   | IX   |
| ABSTRACT .....   | X    |
| 1. INTRODUÇÃO .....  | 1    |
| 2. SISTEMA DE CODIFICAÇÃO DE AÇÃO FACIAL .....                     | 7    |
| 2.1 LIMITAÇÃO DO SISTEMA DE CODIFICAÇÃO DE AÇÃO FACIAL .....       | 8    |
| 2.2 DESENVOLVIMENTO DO SISTEMA DE CODIFICAÇÃO DE AÇÃO FACIAL ..... | 9    |
| 3. WAVELETS DE GABOR.....  | 13   |
| 4. ANÁLISE DE COMPONENTES PRINCIPAIS.....                          | 21   |
| 4.1 GERAÇÃO DO ESPAÇO DE CARACTERÍSTICAS .....                     | 22   |
| 4.2 ESCOLHA DAS CARACTERÍSTICAS .....                              | 26   |
| 4.3 MÉTODO M.....  | 27   |
| 5. SISTEMA DE RECONHECIMENTO DE EXPRESSÕES.....                    | 29   |
| 5.1 BANCO DE DADOS .....   | 31   |
| 5.2 DIFERENÇA DA AU 0 MÉDIA .....                                  | 34   |
| 5.3 APLICAÇÃO DAS WAVELETS DE GABOR .....                          | 36   |
| 5.4 APLICAÇÃO DO PCA.....  | 37   |
| 5.5 REDE NEURAL .....  | 39   |
| 6. RESULTADOS EXPERIMENTAIS.....                                   | 45   |
| 6.1 FACE SUPERIOR .....  | 45   |
| 6.2 FACE INFERIOR.....   | 50   |
| 6.3 DESEMPENHO EM TEMPO REAL .....                                 | 54   |
| 7. CONCLUSÕES.....   | 56   |
| REFERÊNCIAS .....  | 59   |
| APÊNDICE A .....   | 64   |
| APÊNDICE B .....   | 68   |
| APÊNDICE C .....   | 71   |

## Lista de Abreviações

|      |   |
|------|---|
| PCA  | <i>Principal Component Analysis</i>       |
| SNNS | <i>Stuttgart Neural Network Simulator</i> |
| FACS | <i>Facial Action Coding System</i>        |
| AU   | <i>Action Unit</i>                        |
| ICA  | <i>Independent Component Analysis</i>     |
| HCI  | <i>Human-Computer Interaction</i>         |
| RAM  | <i>Random Access Memory</i>               |



# Lista de Figuras

|   |    |
|---|----|
| Figura 1.1: (a) Localização de 20 pontos utilizados para o reconhecimento de Unidades de Ação; (b) Localização de 34 pontos utilizados para o reconhecimento de expressões de emoção.....   | 4  |
| Figura 2.1: (a) Principais músculos faciais; (b) Diversos prolongamentos musculares da cabeça humana .....  | 10 |
| Figura 3.1: Córtex Visual Primário .....  | 13 |
| Figura 3.2: Agrupamento de wavelets de Gabor dispostas de acordo com sua frequência ( $k_v$ ) e orientação ( $\theta_\mu$ ) .....   | 16 |
| Figura 3.3: Projeções do (a) módulo, (b) parte real e (c) imaginária do filtro de Gabor.16  |    |
| Figura 3.4: Distribuição da frequência espacial e orientação no domínio da frequência. ....   | 17 |
| Figura 3.5: Distribuição da parte real dos filtros de Gabor no domínio espacial para frequências espaciais (a) $k_v = \pi/16$ , (b) $k_v = \pi/8$ , (c) $k_v = \pi/4$ e (d) $k_v = \pi/2$ com orientação $\theta_\mu = \pi/2$ .....   | 18 |
| Figura 3.6: Imagem Original.....  | 19 |
| Figura 3.7: Imagens filtradas através de wavelets de Gabor com frequência espacial ( $k_v$ ) e orientação ( $\theta_\mu$ ): (a) $(k_v, \theta_\mu) = \pi/16, \pi/4$ , (b) $(k_v, \theta_\mu) = \pi/4, \pi/4$ , (c) $(k_v, \theta_\mu) = \pi/2, \pi/4$ , (d) $(k_v, \theta_\mu) = \pi/16, \pi/2$ , (e) $(k_v, \theta_\mu) = \pi/4, \pi/2$ e (f) $(k_v, \theta_\mu) = \pi/2, \pi/2$ ..... | 20 |
| Figura 4.1: Representação de uma imagem na forma: (a) matricial e (b) vetorial.....   | 21 |
| Figura 4.2: Média das imagens de treinamento.....   | 23 |
| Figura 4.3: Autovalores <i>versus</i> número de autovetores.....  | 28 |
| Figura 5.1: Estrutura do sistema proposto .....   | 29 |
| Figura 5.2: Face superior segmentada da (a) imagem neutra e (b) com uma expressão de espanto. ....  | 30 |
| Figura 5.3: Face inferior segmentada da (a) imagem neutra e (b) com uma expressão de espanto. ....  | 30 |
| Figura 5.4: Imagens segmentadas e re-escaladas da (a) face superior e (b) face inferior. ....   | 31 |
| Figura 5.5: (a) Primeiro quadro da seqüência e (b) o último quadro de um indivíduo apresentando uma expressão de espanto. ....  | 32 |
| Figura 5.6: Ação facial da face superior: (a) imagem original (b) imagem neutra média e (c) imagem diferença .....  | 35 |
| Figura 5.7: Ação facial da face inferior: (a) imagem original (b) imagem neutra média e (c) imagem diferença.....   | 36 |
| Figura 5.8: Distribuição do percentual da quantidade de informação mantida para os 30 autovetores da (a) face superior (b) face inferior. ....  | 37 |
| Figura 5.9: Percentual da quantidade de informação em relação ao número de componentes principais para a face superior .....  | 38 |
| Figura 5.10: Estrutura da rede neural .....   | 40 |
| Figura 5.11: Representação de 3 classes de dados .....  | 41 |
| Figura 5.12: Exemplos do (a) padrão 5, AU4+6+7+9; (b) padrão 6, AU 4+7+9 .....  | 43 |
| Figura 5.13: Exemplos do (a) padrão 1, AU 0; (b) padrão 2, AU 25; (c) padrão 3, AU 26 .....   | 43 |

|   |    |
|---|----|
| Figura 6.1: Distribuição no domínio de Fourier do filtro de Sobel para orientação espacial: (a) $\frac{\pi}{2}$ e (b) $\pi$ ..... | 47 |
| Figura B.1: Aplicativo utilizado para a segmentação de imagens .....  | 68 |

## Lista de Tabelas

|   |    |
|---|----|
| Tabela 2.1: Exemplo de informação dada pelo FACS .....  | 11 |
| Tabela 5.1: Padrões de ação facial da parte superior da face.....                                 | 33 |
| Tabela 5.2: Padrões de ação facial da parte inferior da face.....                                 | 34 |
| Tabela 5.3: Filtros de Gabor aplicados à imagem original e a imagem diferença. ....               | 37 |
| Tabela 5.4: Dois componentes principais para a imagem original e a imagem diferença .....         | 39 |
| Tabela 5.5: Dimensão do conjunto de treinamento e teste para a parte superior da face             | 42 |
| Tabela 5.6: Dimensão do conjunto de treinamento e teste para a parte inferior da face             | 42 |
| Tabela 6.1: Estrutura e resultados dos experimentos .....   | 46 |
| Tabela 6.2: Matriz de confusão do experimento 3 .....   | 49 |
| Tabela 6.3: Matriz de confusão para o experimento 3 após a aplicação da segunda rede neural ..... | 49 |
| Tabela 6.4: Resultado final do experimento com a face superior .....                              | 50 |
| Tabela 6.5: Matriz de confusão para o experimento com a face inferior .....                       | 50 |
| Tabela 6.6: Matriz de confusão após a aplicação da segunda rede neural.....                       | 51 |
| Tabela 6.7: Resultado final do experimento com a face inferior .....                              | 51 |
| Tabela 6.8: Expressões de emoção resultantes da combinação de AUs.....                            | 53 |
| Tabela 6.9: Desempenho do Sistema.....  | 54 |
| Tabela A.1: Unidades de ação facial e os músculos que são estimulados.....                        | 64 |
| Tabela A.2: Descrição de AUs pouco comuns.....  | 67 |
| Tabela B.1: Localização dos pontos na imagem a ser cortada.....                                   | 69 |

# Resumo

Diversos artigos têm sido publicados utilizando independentemente as Wavelets de Gabor e a Análise de Componentes Principais (PCA) no reconhecimento de ações faciais e expressões de emoção. Este trabalho propõe um sistema automático para reconhecer ações faciais em imagens de elevada complexidade baseado em Wavelets de Gabor seguido pelo PCA, aplicado na parte superior e inferior de uma face subtraída de sua face neutra média. As faces superior e inferior são segmentadas da imagem original e re-escalada para uma janela de dimensão  $30 \times 45$ . Uma técnica de pré-processamento que gera um volume reduzido de dados de saída é proposta. Uma taxa de reconhecimento de 87 % é obtida para classificar 7 ações faciais superior e uma taxa de 82% para reconhecer 7 ações faciais inferior, usando-se um classificador baseado em Redes Neurais. Seqüências de imagens de indivíduos heterogêneos com movimentos de cabeça e com variação na luminosidade foram consideradas.

## **Palavras Chave**

Reconhecimento de expressões faciais, Wavelets de Gabor, PCA e Redes Neurais.

# Abstract

Several research works have been published independently using the Gabor Wavelets and Principal Component Analysis (PCA) to recognize facial actions and emotion expressions. This work proposes an automatic system to recognize facial actions in images of increasing complexity using Gabor Wavelets followed by PCA applied in an upper and lower face subtracted from its neutral face average. The upper and lower faces are extracted from the original image and resized to a  $30 \times 45$  dimensional window. A pre-processing technique that generates a reduced output data volume is proposed. A recognition rate of 87 % is obtained to recognize 7 upper facial actions and 82 % to 7 lower facial actions by using a Neural Network-based classifier. Image sequences from heterogeneous subjects with head motion and with luminosity variation have been considered.

## Keywords

Facial expression recognition, Gabor Wavelets, PCA and Neural Network

# Capítulo 1

## Introdução

A comunicação de pessoas através da expressão facial é uma das formas mais poderosas, naturais e de compreensão mais rápida para se comunicar emoção e intenção. A face pode expressar uma emoção mais rapidamente do que a fala e até mesmo antes que o indivíduo perceba seu próprio sentimento.

De expressões faciais provêm informações não só sobre os estados afetivos, mas também sobre atividades cognitivas, temperamentais e de personalidade, honestidade e psicopatologias. Com o objetivo de se obter uma forma mais adequada de comunicação entre seres humanos e máquinas, o interesse no desenvolvimento de pesquisas em sistemas de Interação Homem-Computador (HCI) têm aumentado consideravelmente nos últimos anos. As formas atuais de interação com o computador, através do uso de um teclado ou *mouse*, estão longe de serem ideais. É importante que as pessoas possam se comunicar com computadores de forma natural, em um estilo similar ao que fazemos dia-a-dia, falando, escutando e executando expressões faciais e corporais. O reconhecimento automático de expressões faciais está sendo um grande desafio para os pesquisadores da área de visão computacional, pois a expressão é gerada por deformações em um objeto não-rígido e essas deformações variam de pessoa para pessoa. Em imagens estáticas, o reconhecimento torna-se ainda mais difícil quando comparado com o reconhecimento de seqüências de imagens, isso se deve ao fato de não se dispor da informação durante uma ação facial.

Para a comunidade científica da área, o Sistema de Codificação de Ação Facial (FACS) é atualmente o principal método para avaliar movimentos faciais [1]. Tal sistema é implementado manualmente por especialistas humanos altamente treinados. Avanços recentes na análise de imagens abriram a possibilidade de se avaliar automaticamente os sinais faciais. Um sistema automático tornaria a avaliação de expressões faciais mais acessível como ferramenta de pesquisa para profissionais médicos e pesquisadores da área. Esse sistema também poderia ser aplicado na tarefa de Interação Homem-Máquina (HCI) [2].

Em 1978, Ekman e Friesen desenvolveram um FACS [3] que é capaz de objetivamente avaliar atividades faciais para uso em pesquisas científicas. Tal sistema fornece uma descrição objetiva dos sinais faciais em termos dos componentes de movimentos, ou “Ações Faciais”. O sistema foi desenvolvido pela palpação, conhecimento da anatomia e gravações de vídeo. Eckman e Friesen definiram 46 Unidades de Ação (AU), que correspondem a cada movimento independente da face. Mais de 7000 combinações de ações faciais foram observadas em comportamentos faciais espontâneos [4]. Um humano treinado no FACS decompõe uma expressão observada em AUs específicas que reproduzem a expressão. O FACS é criado através de um vídeo e as imagens provêm uma informação precisa da dinâmica (duração, início e fim) dos movimentos faciais e da morfologia (específica à ação facial que ocorre).

O FACS é usado, por exemplo, para demonstrar diferenças entre dor verdadeira e simulada [5], diferenças entre quando o indivíduo está falando a verdade ou mentindo, como também diferenças entre sinais faciais de pacientes deprimidos suicidas e não suicidas [6]. Embora o FACS atualmente ser o principal método para se avaliar expressões faciais, o maior impedimento para seu uso em larga escala é o tempo

requerido para treinar humanos especialistas e para classificar manualmente o vídeo. Tal tarefa pode levar mais de 100 horas de treinamento para se alcançar um desempenho mínimo no FACS e cada minuto de vídeo leva aproximadamente uma hora para ser classificado [3].

Um sistema automático de classificação de ação facial tornará as pesquisas sobre a face humana mais acessível. A criação de uma ferramenta automática de análise não só aumentará a velocidade de classificação, mas também aumentará a confiabilidade, precisão e resolução temporal das medidas faciais.

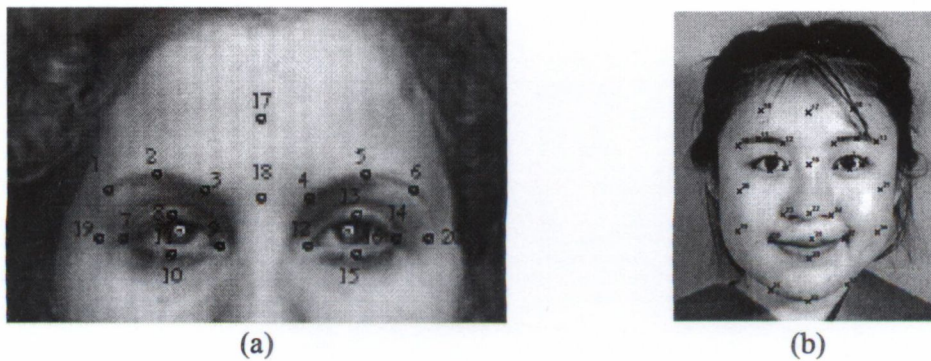
Avanços têm sido feitos em visão computacional visando o reconhecimento automático de expressões faciais em imagens. As técnicas até então exploradas incluem as de movimentos faciais [7-10]; medidas da forma, características faciais e seu arranjo espacial [11,12]; análise de padrões espaciais holísticos usando técnicas baseadas na Análise de Componentes Principais (PCA) [11,13,14]; filtros espaciais locais, utilizando análise de padrão do nível de cinza [12,14] e métodos que relacionam imagens faciais com modelos físicos da pele e musculatura facial [7,10,15,16].

As técnicas de análise de expressões faciais citados anteriormente são relevantes para o objetivo presente, porém possuem uso limitado para investigação científica da face. Muitas dessas técnicas são concebidas com o objetivo de classificar expressões faciais em poucas categorias básicas de emoções, tal como felicidade, tristeza ou surpresa. Para uma investigação científica elementar do comportamento facial, tal como a diferença entre a dor verdadeira e a simulada [5], uma medida objetiva e detalhada da atividade facial tal como a proposta pelo FACS faz-se necessária.

Ying-li Tian *et. al.* [17] definiram manualmente 20 pontos na parte superior da face e criaram um sistema capaz de reconhecer 9 AUs utilizando o método de características geométricas e as wavelets de Gabor combinadas através de uma rede



neural. A técnica de características geométricas consiste em analisar a disposição e relação entre os pontos definidos anteriormente para determinar a expressão facial correspondente. A determinação da expressão com base nas wavelets de Gabor se dá através da aplicação de filtros em distintas orientações e frequências. A Figura 1.1(a) mostra a disposição dos pontos para esse experimento.



**Figura 1.1:** (a) Localização de 20 pontos utilizados para o reconhecimento de Unidades de Ação; (b) Localização de 34 pontos utilizados para o reconhecimento de expressões de emoção.

Em tal experimento foram utilizados 40 filtros de Gabor e as imagens de teste foram extraídas do banco de dados de expressões faciais desenvolvido por Cohn-Kanade [18]. Indivíduos heterogêneos, variação na luminosidade e pequenos movimentos involuntários com a cabeça estão presentes nos testes.

Zang *et. al.* [19] aplicou 18 filtros de Gabor e extraiu a característica geométrica de 34 pontos selecionados manualmente para determinar 6 categorias de expressões de emoção (alegria, tristeza, surpresa, irritação, desgosto e medo), além da imagem neutra. O experimento foi realizado apenas com expressões de indivíduos do sexo feminino. A localização dos pontos pode ser observada na Figura 1.1(b).

Donato *et. al.* [2] aplicou independentemente diversas técnicas (fluxo ótico, PCA, Análise de Componentes Independentes (ICA), Análise de Características Locais

(LFA) e wavelets de Gabor) para reconhecer 6 AUs da face superior e 6 AUs da face inferior de 24 indivíduos. Um cuidadoso alinhamento da face, rotação dos olhos, escalamento e segmentação da face, aumento de contraste e equalização de histograma foi realizado para todas as seqüências de imagens. Também, a expressão que se deseja classificar foi subtraída de todos os quadros anteriores àquela seqüência. Os passos de pré-processamento, presentes no trabalho de Donato, aumentam a complexidade computacional, tornando inviável a aplicação do sistema em aplicações que operam em tempo real.

Fase e Luëttin [20] propuseram um sistema para reconhecer atividades faciais e sua intensidade de 16 distintas ações faciais utilizando a técnica PCA e ICA independentemente. As imagens de todos os indivíduos mantêm uma posição constante ao longo das seqüências, sem mudança na escala, orientação e luminosidade. As faces foram separadas da imagem do fundo através de uma máscara que foi ajustada manualmente. Foi realizada também a diferença entre o quadro de um indivíduo realizando a expressão e o quadro com a face neutra daquele indivíduo.

As abordagens anteriores sugerem técnicas que requerem um elevado esforço no pré-processamento para alinhar a face e escolher os pontos de interesse, ou para segmentar, rotacionar e escalar a face para adaptá-la a um determinado padrão. Alguns trabalhos sugerem que as wavelets de Gabor podem obter uma alta sensibilidade e especificidade para determinadas expressões de emoção [19,20] e AUs individuais ou em combinação [2]. Geralmente a taxa de reconhecimento pode aumentar se algumas restrições e alguns passos de pré-processamento forem seguidos como: (1) Utilizar uma base de dados com indivíduos homogêneos, tais como Japoneses ou Ibero-Americanos; (2) Excluir movimentos da cabeça; (3) efetuar uma minuciosa segmentação e

alinhamento dos quadros; (4) excluir do conjunto de testes combinações de AUs que podem mudar a estrutura individual de cada AU e (5) testar expressões muito distintas.

Neste trabalho, será utilizado o banco de dados de Cohn-Kanade [18], composto por expressões de indivíduos heterogêneos com ancestrais Europeus, Africanos, Asiáticos e outros. Serão consideradas seqüências de vídeo com variação na luminosidade, distintos tamanhos das faces, indivíduos efetuando pequenos movimentos com a cabeça e ações faciais similares. Um alinhamento (aproximado) dos olhos, nariz e do queixo é efetuado no momento da segmentação. A etapa de pré-processamento é composta pela subtração da imagem do indivíduo efetuando a expressão da imagem facial neutra média de todo o banco de dados, em seguida são aplicadas as wavelets de Gabor com baixo custo computacional. O método de extração de características é baseado em wavelets de Gabor seguido pela PCA para redução dimensional.

No Capítulo 2 é descrito de forma sintética o FACS. O Capítulo 3 apresenta os princípios das wavelets de Gabor em imagens. No Capítulo 4, a técnica de redução da dimensão PCA, é apresentada. No Capítulo 5 descreveremos o sistema automático de reconhecimento de ações faciais proposto e, no Capítulo 6, são apresentados alguns resultados experimentais. Finalizando o trabalho, no Capítulo 7, são apresentadas as conclusões.

## *Capítulo 2*

# **Sistema de Codificação de Ação Facial**

Pesquisas anteriores à criação do FACS não utilizavam nenhum método de medida do comportamento facial, eles mediam a informação que os observadores eram capazes de distinguir. No entanto, normalmente os observadores não estão aptos a fazerem uma interferência apurada sobre a emoção, detectar mudanças clínicas ou fazerem diagnósticos e levam em conta mais a voz do que a face. Também, observadores de diferentes culturas podem interpretar expressões faciais diferentemente ou serem influenciados pelo contexto no julgamento da face.

Antes do surgimento do FACS, poucos estudos foram realizados com o objetivo de criar sistemas que realmente fossem capazes de mediar ações faciais. Uma forma mais eficiente de avaliação facial, possível através do FACS, é analisar quais movimentos sinalizam uma emoção, que ocorrem em diferentes contextos e culturas, que diferenciam-se entre tipos de psicopatologias, que são inibidos em certos ambientes sociais e são enfatizados em uma conversação.

As condições de análise da face através de um observador e através de um sistema de codificação são revisadas por Eckman, Friesen e Ellsworth [22].

Diversos impedimentos surgiram nas pesquisas direcionadas à face pelo não desenvolvimento de uma técnica de codificação adequada para medição do comportamento facial. Até o surgimento do FACS, proposto por Eckman e Friesen [3] no final dos anos 70, diversos procedimentos de medidas faciais foram desenvolvidos

[23][24] e mesmo assim não houve consenso entre os pesquisadores sobre como medir o comportamento facial, não sendo desenvolvida nenhuma ferramenta que tivesse se tornado um padrão.

O FACS é capaz de relacionar todos os movimentos faciais visualmente distinguíveis e pode ser utilizado em pesquisas sobre estados emocionais e deficiências faciais que indicam uma lesão cerebral.

O desenvolvimento do FACS foi realizado considerando a anatomia dos movimentos faciais. Todo movimento facial é resultado de uma ação muscular e portanto a criação de um sistema compreensivo pode ser obtido pela descoberta de como cada músculo da face atua para efetuar uma expressão. Através deste conhecimento será possível analisar qualquer movimento facial em unidades de ação mínimas baseada na anatomia da face.

## **2.1 Limitação do Sistema de Codificação de Ação Facial**

Uma das principais limitações do FACS está no fato dele tratar apenas características claramente visíveis na face, ignorando mudanças invisíveis, como certas mudanças nos estímulos musculares, como também desconsiderando mudanças faciais muito súbitas como, por exemplo, as necessárias para a identificação de honestidade.

Este sistema exclui mudanças visíveis nos estímulos que não envolvam movimentos, como por exemplo, mudanças na coloração da pele que não são percebidas em imagens preto e branco, suor facial, inflamações, espinhas e características faciais permanentes. O indivíduo que utiliza o FACS deve aprender a mecânica e a base

muscular dos movimentos faciais, não apenas a consequência dos movimentos ou uma descrição de marcas estáticas.

O FACS enfatiza padrões de movimento, ou seja, a mudança natural da aparência facial. Suas ações são descritas como movimento da pele, mudança temporária na forma e localização de características, saliências e enrugamento da pele.

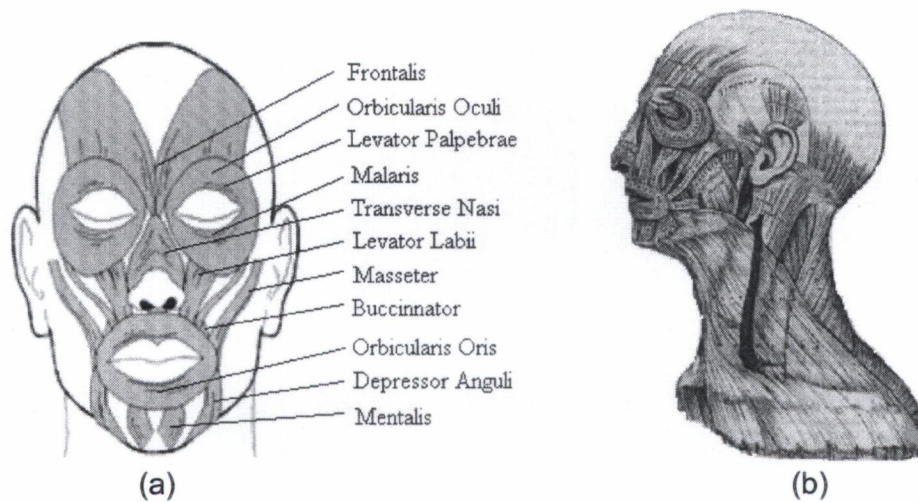
## 2.2 Desenvolvimento do Sistema de Codificação de Ação Facial

O primeiro passo no desenvolvimento do FACS foi o estudo de vários textos anatômicos para se descobrir uma unidade mínima. Os autores do FACS, Eckman e Friesen, procuravam uma lista dos músculos faciais que pudessem atuar separadamente e mudar a aparência facial.

Em seguida, eles examinaram fotografias de várias expressões faciais, misturaram-nas de forma que não soubessem qual músculo foi estimulado por cada expressão facial. Eckman e Friesen esperavam determinar quais ações musculares poderiam ser distinguíveis somente através da aparência facial. Eles chamaram as medidas de unidades de ação (AU) e não de unidade de músculo, isso porque eles combinaram mais de um músculo em suas unidades. Outra razão para usar o termo unidade de ação é devido à mais de uma ação ser designada para um mesmo músculo.

Por exemplo, o músculo *frontalis*, mostrado na Figura 2.1(a), que levanta a sobrancelha, está presente em duas unidades de ação, uma quando é levantada a porção interna da sobrancelha e outra quando é erguida a porção externa da mesma.

A relação das unidades de ação, exemplos de fotografias das ações faciais e a descrição dos músculos estimulados durante a expressão são mostrados no Apêndice A.



**Figura 2.1:** (a) Principais músculos faciais; (b) Diversos prolongamentos musculares da cabeça humana

A Figura 2.1(a) ilustra uma visão frontal dos principais músculos presentes na face. A Figura 2.1(b) mostra os inúmeros prolongamentos musculares existentes na cabeça humana, o que dificulta a classificação de expressões faciais baseando-se apenas em um músculo.

A Tabela 2.1 mostra um exemplo de como uma Unidade de Ação é descrita no manual do FACS. Esta descrição é realizada através de quatro tipos de informação.

- (1) Descrição da base muscular de cada AU através de diagramas;
- (2) Descrição detalhada da mudança de aparência, ilustrada com exemplos de fotografias e filmes;
- (3) Descrição de instruções de como executar um movimento facial na própria face.
- (4) Regras para especificar as mudanças mínimas que devem ser observadas para que uma AU possa ser realizada.

O exemplo a seguir apresenta uma descrição da AU 15 de acordo com o manual do FACS.

**Tabela 2.1:** Exemplo de informação dada pelo FACS

| <b>AU 15 – Borda dos lábios abaixada</b>  |
|---|
| <p>O músculo fundamental da AU 15 surge ao lado do queixo e segue para cima, juntando-se ao ponto próximo à borda dos lábios. Na AU 15 as bordas dos lábios são puxadas para baixo.</p> <p>Considerações sobre a forma, causa e deformações provocadas na execução desta AU.</p> <p>(1) Bordas dos lábios puxadas para baixo.</p> <p>(2) Mudanças na forma dos lábios, de maneira que o ângulo com a horizontal, das bordas dos mesmos sejam negativos.</p> <p>(3) Surgimento de uma espécie de enrugamento da pele abaixo das bordas dos lábios, que normalmente aparece quando a ação facial é mais intensa.</p> <p>(4) Pode causar alisamento ou enrugamento quando surgir em um queixo mais saliente. Pode produzir um enrugamento no meio e logo abaixo do lábio inferior.</p> <p>(5) Se os sulcos nasolabiais estão permanentemente presentes, a AU aprofunda-se e pode aparecer puxada para baixo ou estendida.</p> <p>Fotografias no manual do FACS mostram a versão desta Unidade de Ação tanto levemente como fortemente presente. Note que o item (3), citado anteriormente, é mais aparente quando uma ação mais intensa é executada.</p> <p>Como executar a AU 15:</p> <p>Ponha as bordas dos lábios para baixo. Tome cuidado para não levantar o lábio inferior ao mesmo tempo (não use a AU 17). Se você é incapaz de fazer isso, ponha seus dedos acima da borda dos lábios e os puxe para baixo, notando a mudança na aparência. Agora, tente fazer tal aparência sem o uso dos dedos.</p> <p>Requisitos mínimos para executar a AU 15:</p> <p>(1) Se a linha dos lábios é sutilmente levantada na face neutra, então a borda dos lábios</p> |



deve ser, no mínimo, puxada para baixo para poder ser classificada como AU 15.

Ou (2), se a linha dos lábios é um pouco abaixada na face neutra, então a borda dos lábios deve ser puxada para baixo pouco mais do que na face neutra, com o cuidado de não executar a AU 17 ou AU 20.

O prolongamento da boca é irrelevante, isso está presente na AU 20, AU 15, ou AU 15+20.

A determinação das AUs individuais, como mostrado no Apêndice A, e suas descrições, como o exemplo mostrado na Tabela 2.1, são os primeiros passos no desenvolvimento do FACS. O procedimento para a movimentação dos músculos, fotografia dos movimentos faciais e inspeção das fotos são retomadas com todas as possíveis combinações de 2 AUs. Não há necessidade de descrever combinações de AU que não se interagem.

Combinações de AUs podem ser do tipo aditivas, quando uma AU muda o comportamento da outra e não-aditivas quando as AUs são executadas como se estivessem sozinhas na face. Estudos de diversas expressões faciais mostraram que a maioria das combinações de AUs são aditivas, que são mais difíceis de serem detectadas por um sistema automático ou mesmo através de humanos especialistas em FACS.

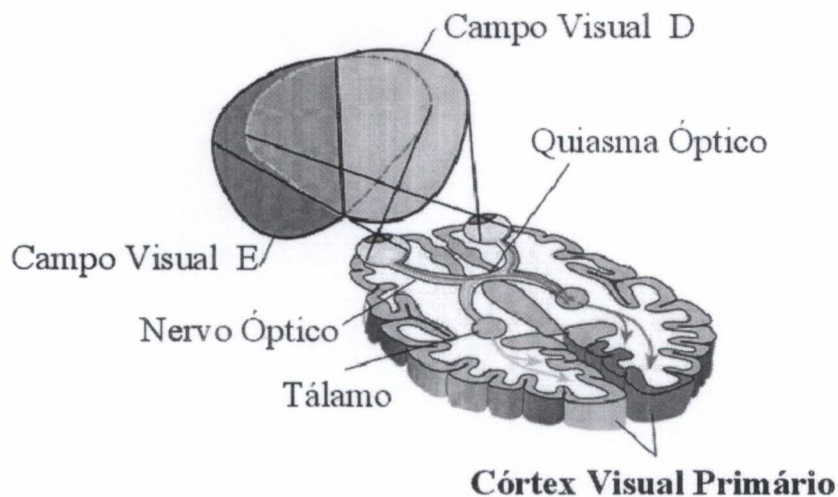
Além da classificação das expressões através de AUs, o manual do FACS define níveis de intensidade para cada AU de acordo com 5 letras do alfabeto romano, A, B, C, D e E, sendo que a última representa a máxima intensidade para a expressão.

A análise do comportamento facial é uma tarefa extremamente difícil, mesmo para humanos treinados em FACS. Inúmeras variáveis estão envolvidas no processo de classificação como a variação na intensidade, a presença de combinações de AUs e também a heterogeneidade de ações faciais entre o lado direito e esquerdo da face.

## Capítulo 3

### Wavelets de Gabor

Desde o descobrimento da organização cristalina do córtex visual primário no cérebro de mamíferos há trinta anos por Hubel e Wiesel [25], um crescente esforço em pesquisas tem permitido fortalecer nosso entendimento nesta área. No âmbito teórico, uma importante descoberta foi realizada por Marcelja [26] e Daugman [27]. Eles descobriram que os sinais elétricos dos neurônios do córtex visual podem ser modelados por funções de Gabor. A Figura 3.1 mostra um esboço do córtex visual primário que é responsável pelos estímulos visuais, no qual são introduzidos eletrodos e, o sinal resultante, é modelado por funções de Gabor 2D.



**Figura 3.1:** Córtex Visual Primário

As funções de Gabor 2D propostas por Daugman [27] são filtros passa-faixas espaciais locais que possuem uma resolução no domínio espacial e de Fourier 2D.

As funções de Gabor 1D foram primeiramente propostas por Dennis Gabor em 1946 [28]. Gabor criou a transformada de Gabor, análoga a transformada de Fourier, que permitia separar um sinal em segmentos considerando uma certa resolução no domínio do tempo e na frequência. A função elementar de Gabor é gerada através de uma Gaussiana fixa, enquanto as frequências das ondas moduladas variam.

Os filtros de Gabor, redescobertos e generalizados para 2D, estão sendo extensivamente usados em várias aplicações de visão computacional. Daugman [27] generalizou as funções de Gabor para a seguinte forma 2D.

$$\Psi_i(\vec{x}) = \frac{|\vec{k}|^2}{\sigma^2} e^{\frac{|\vec{k}|^2 |\vec{x}|^2}{2\sigma^2}} \left( e^{i\vec{k}_i \cdot \vec{x}} - e^{-\frac{\sigma^2}{2}} \right) \quad (3.1)$$

A função  $\Psi_i$  é a resposta ao impulso do filtro de Gabor que é caracterizado pelo vetor  $\vec{k}_i$ , onde  $\sigma$  é o desvio padrão da Gaussiana. O vetor  $\vec{x} = (x, y)$  refere-se a posição do filtro. A frequência central do  $i$ -ésimo filtro é dada pelo vetor característico de onda  $\vec{k}_i$  da seguinte forma:

$$\vec{k}_i = \begin{pmatrix} k_{ix} \\ k_{iy} \end{pmatrix} = \begin{pmatrix} k_v \cos \theta_\mu \\ k_v \sin \theta_\mu \end{pmatrix} \quad (3.2)$$

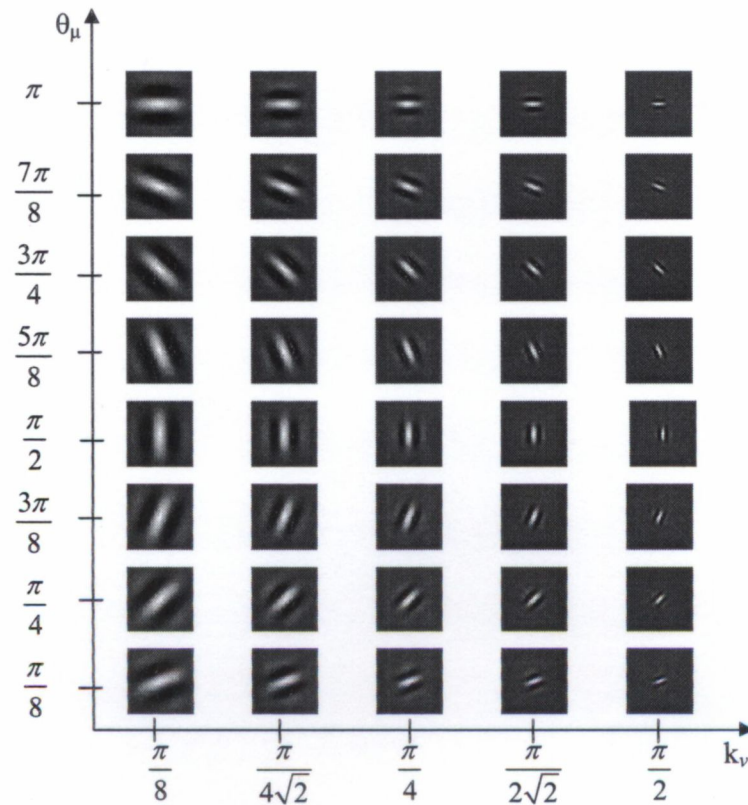
onde  $\vec{k}_i$  é representado por um determinado fator de escala ou frequência espacial  $k_v$  e uma orientação  $\theta_\mu$ . O primeiro termo entre parênteses na Equação (3.1) determina a parte oscilatória do núcleo da wavelet, e o segundo termo faz a compensação do valor DC deste núcleo [29]. Subtraindo a resposta DC, o filtro de Gabor torna-se menos sensível às variações dos níveis de luminosidade.

Evidências recentes em estudos neuropsicológicos sugerem que a estrutura espacial dos campos receptivos dos neurônios com diferentes tamanhos são praticamente invariantes. Daugman [27] e Marcelja [26] sugeriram que os neurônios do córtex visual primário de mamíferos poderiam ser bem modelados através de uma família de wavelets de Gabor 2D. As wavelets 2D são diversos filtros que possuem uma determinada orientação e frequência no espaço 2D. A decomposição de uma imagem  $I$  através de um conjunto de filtros é chamada de transformação wavelet de uma imagem, calculada por:

$$R_i(\vec{x}) = \int I(\vec{x}') \Psi_i(\vec{x} - \vec{x}') d\vec{x}' \quad (3.3)$$

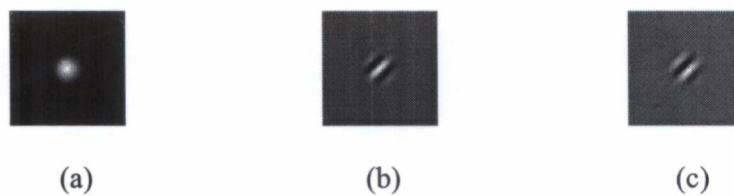
onde  $I(\vec{x})$  é a intensidade do *pixel* da imagem na posição  $\vec{x}$ .

A Figura 3.2 mostra um agrupamento de wavelets de Gabor, no domínio espacial com diferentes frequências espaciais ( $k_\nu$ ) e orientações ( $\theta_\nu$ ). Somente a parte real da resposta ao impulso do filtro de Gabor está representada. As regiões mais claras das imagens da Figura 3.2 representam os coeficientes de maior intensidade do filtro no domínio espacial.



**Figura 3.2:** Agrupamento de wavelets de Gabor dispostas de acordo com sua frequência ( $k_v$ ) e orientação ( $\theta_\mu$ )

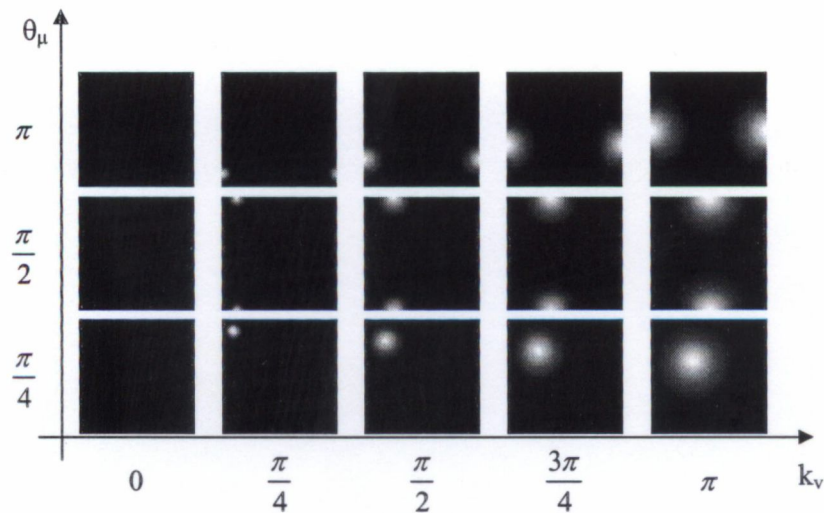
A Figura 3.7 mostra a projeção do módulo, parte real e parte imaginária, no domínio espacial para o filtro de Gabor com frequência espacial  $k_v = \pi/4$  e orientação  $\theta_\mu = \pi/4$ .



**Figura 3.3:** Projeções do (a) módulo, (b) parte real e (c) imaginária do filtro de Gabor.

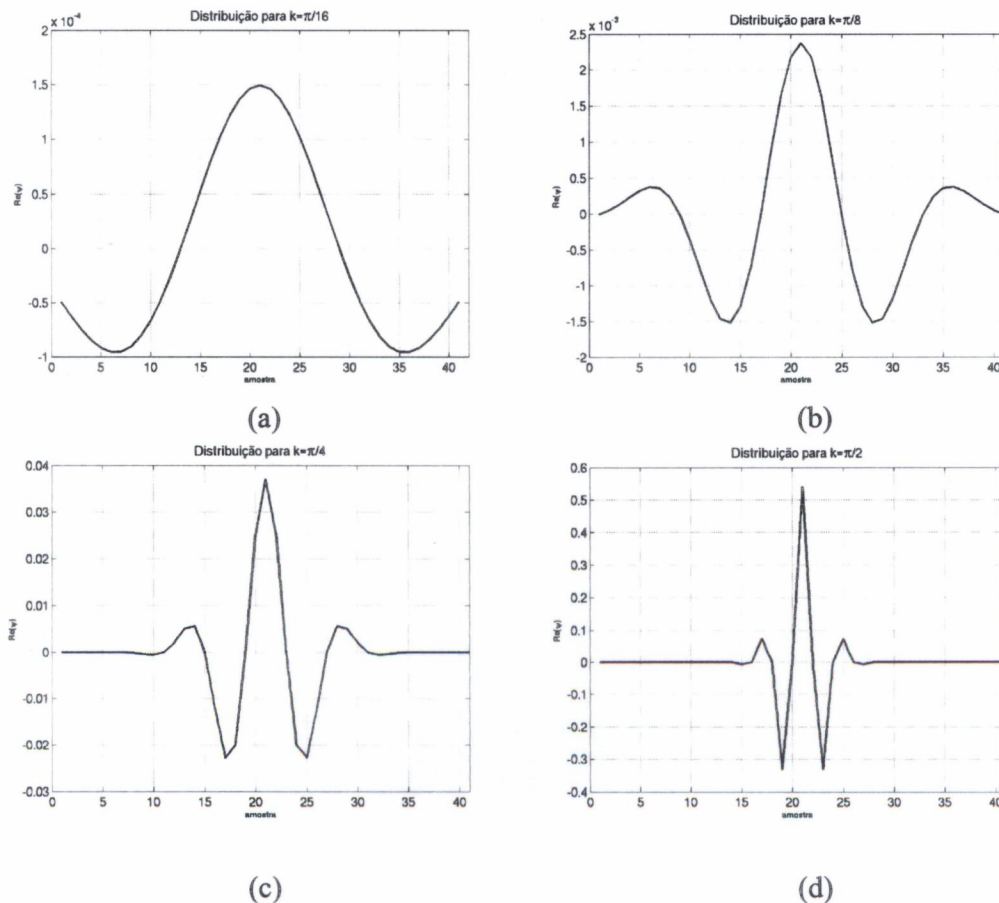
Podemos perceber que tanto a parte real como a imaginária obedecem a orientação determinada por  $\theta_\mu$  e que o módulo do filtro de Gabor possui a forma de uma distribuição Gaussiana.

A Figura 3.4 mostra a distribuição do módulo dos filtros de Gabor no domínio da frequência. As regiões mais claras das imagens da Figura 3.4 representam a maior intensidade do sinal no filtro no domínio da frequência.



**Figura 3.4:** Distribuição da frequência espacial e orientação no domínio da frequência.

É possível notar na Figura 3.4 que quanto maior a frequência espacial do filtro, maior é a sua banda de passagem, isto é, a largura de banda relativa do filtro é mantida. A Figura 3.5 mostra o corte da parte real dos filtros de Gabor com orientação  $\theta_\mu = \pi/2$  e quatro frequências espaciais  $k_v = \pi/16, \pi/8, \pi/4$  e  $\pi/2$ .

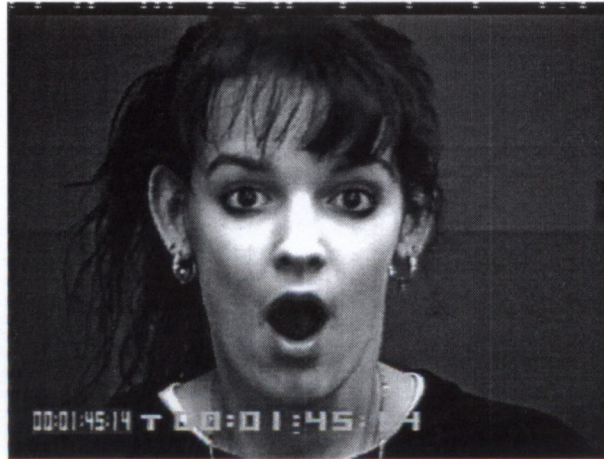


**Figura 3.5:** Distribuição da parte real dos filtros de Gabor no domínio espacial para frequências espaciais (a)  $k_v = \pi/16$ , (b)  $k_v = \pi/8$ , (c)  $k_v = \pi/4$  e (d)  $k_v = \pi/2$  com orientação  $\theta_\mu = \pi/2$

Cada membro desta família de wavelets de Gabor modela a estrutura do campo receptivo espacial de um neurônio do córtex visual primário. A decomposição de Gabor pode ser considerada como uma microscopia direcional sensível à orientação e à escala. Os neurônios do córtex visual são inibidos quando estão presentes retas e linhas com pequenas curvaturas na imagem.

Como tais curvas geralmente correspondem a alguma característica saliente na imagem, estes neurônios podem então mapear certas características da imagem através de uma variação de intensidade.

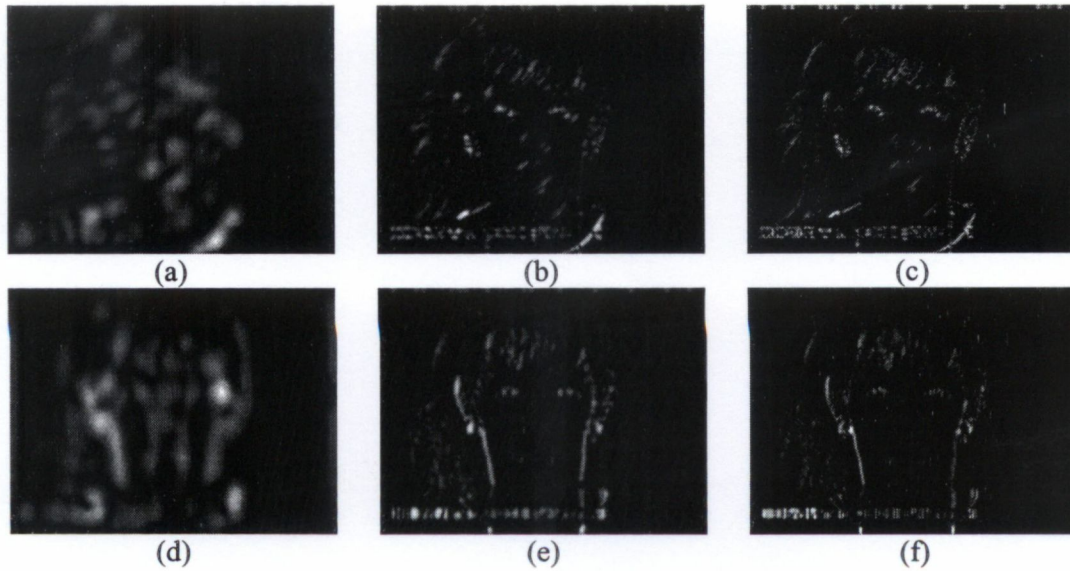
A aplicação das wavelets de Gabor consiste em processar uma imagem através de um conjunto de filtros com uma determinada orientação e frequência espacial 2D. A Figura 3.6 mostra a imagem que será submetida a aplicação dos filtros de Gabor.



**Figura 3.6:** Imagem Original

As Figura 3.7(a), (b) e (c) ilustra a aplicação dos filtros de Gabor para frequências espaciais  $k_v = \frac{\pi}{16}, \frac{\pi}{4}, \frac{\pi}{2}$ , respectivamente, com orientação  $\theta_\mu = \frac{\pi}{4}$ . As Figura 3.7(d), (e) e (f) também mostram a aplicação dos filtros de Gabor para frequências espaciais  $k_v = \frac{\pi}{16}, \frac{\pi}{4}, \frac{\pi}{2}$  porém agora com orientação  $\theta_\mu = \frac{\pi}{2}$ .





**Figura 3.7:** Imagens filtradas através de wavelets de Gabor com frequência espacial ( $k_v$ ) e orientação ( $\theta_\mu$ ): (a)  $(k_v, \theta_\mu) = \pi/16, \pi/4$ , (b)  $(k_v, \theta_\mu) = \pi/4, \pi/4$ , (c)  $(k_v, \theta_\mu) = \pi/2, \pi/4$ , (d)  $(k_v, \theta_\mu) = \pi/16, \pi/2$ , (e)  $(k_v, \theta_\mu) = \pi/4, \pi/2$  e (f)  $(k_v, \theta_\mu) = \pi/2, \pi/2$ .

A Figura 3.7 mostra que as bordas com orientação  $\theta_\mu = \pi/2$  são mais salientes nesta direção, o mesmo acontece para a orientação  $\theta_\mu = \pi/4$ . Também podemos notar que para a frequência espacial  $k_v = \pi/16$  as linhas são mais espessas do que para  $k_v = \pi/2$ . Isto se deve ao fato do filtro de Gabor ser do tipo passa faixas, sendo  $k_v$  a frequência central. Como uma borda representa uma transição e, portanto, composta por componentes de altas frequências, quanto maior o valor de  $k_v$ , maior será a capacidade de distinguir as transições em uma imagem.

A utilização da representação das wavelets de Gabor 2D na área de visão computacional foi introduzida por Daugman nos anos 80. Mais recentemente, B. S. Manjunath [30] desenvolveu um sistema de reconhecimento de face baseado nesta representação. Trabalhos anteriores sugerem que a utilização de filtros motivados biologicamente, como os filtros de Gabor, pode aumentar significativamente a taxa de reconhecimento, quando comparados com outros filtros encontrados na literatura.

# Capítulo 4

## Análise de Componentes Principais

A Análise de Componentes Principais (PCA) [31] envolve um procedimento matemático que transforma um número de variáveis possivelmente correlacionadas em um número menor de variáveis não-correlacionadas, chamado *componentes principais*. A proposta é reduzir a dimensão da representação dos dados tornando o processo de classificação ou tratamento destes mais fácil [32]. O primeiro componente aponta para o eixo de maior variância do conjunto de dados. O segundo componente aponta para o eixo de maior variância do conjunto de dados restantes, e assim sucessivamente. Assim, a idéia do PCA é encontrar os componentes principais ou bases ortogonais  $\{\mathbf{w}_1, \mathbf{w}_2, \dots, \mathbf{w}_n\}$  tal que estas expressem a máxima variância possível do conjunto de dados.

Neste trabalho aplicamos o PCA em imagens, portanto é necessário que as imagens sejam representadas na forma vetorial.

|          |          |          |          |
|----------|----------|----------|----------|
| $P_{11}$ | $P_{12}$ | $P_{13}$ | $P_{14}$ |
| $P_{21}$ | $P_{22}$ | $P_{23}$ | $P_{24}$ |
| $P_{31}$ | $P_{32}$ | $P_{33}$ | $P_{34}$ |

(a)

|          |          |          |          |          |          |          |          |          |          |          |          |
|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|
| $P_{11}$ | $P_{12}$ | $P_{13}$ | $P_{14}$ | $P_{21}$ | $P_{22}$ | $P_{23}$ | $P_{24}$ | $P_{31}$ | $P_{32}$ | $P_{33}$ | $P_{34}$ |
|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|

(b)

**Figura 4.1:** Representação de uma imagem na forma: (a) matricial e (b) vetorial

A Figura 4.1(a) mostra um exemplo de disposição dos *pixels* em uma imagem de dimensão  $3 \times 4$  e a Figura 4.1(b) sua representação na forma vetorial, com dimensão 12.

A transformação PCA é obtida seguindo alguns passos. Primeiro, o conjunto de vetores de treinamento deve ser definido. A matriz de covariância deve ser gerada e todos os autovalores e autovetores desta matriz devem ser determinados. Então, devem ser decididos quais autovetores ou componentes principais serão utilizados.

## 4.1 Geração do Espaço de Características

A projeção de uma imagem, representada na forma vetorial, no subespaço de características é obtida como segue

$$\mathbf{y} = \mathbf{W}^T (\Omega - \mu) \quad (4.1)$$

onde:

$\mathbf{W}$  : autovetores da matriz de covariância arranjados como vetores colunas;

$\Omega$  : vetor imagem a ser projetado;

$\mu$  : vetor imagem média do conjunto de treinamento;

$\mathbf{y}$  : vetor resultante das projeções no espaço de características.

As operações de soma e subtração em uma imagem ou em um vetor imagem é realizado *pixel a pixel*.

Para o cálculo da imagem média do conjunto de dados, vamos considerar um conjunto de imagens arranjadas de forma vetorial como  $\{\Omega_1, \Omega_2, \dots, \Omega_M\}$ , onde cada

imagem que se deseja caracterizar deve ser representada várias vezes. Desta forma, a média do conjunto de imagens de treinamento ( $\mu$ ) é dada por

$$\mu = \frac{1}{M} \sum_{i=1}^M \Omega_i \quad (4.2)$$

onde  $M$  denota o número de imagens de treinamento e  $\Omega_i$ , a  $i$ -ésima imagem de treinamento.

A Figura 4.2 mostra a imagem média considerando todas as faces extraídas de 21 indivíduos, escolhidos aleatoriamente apresentando expressões distintas.



**Figura 4.2:** Média das imagens de treinamento

Os autovetores da matriz  $\mathbf{W}$  são calculados a partir da matriz de covariância  $\mathbf{C}$ . Sejam  $M$  imagens de dimensão  $p \times q$  arranjadas em uma matriz  $\mathbf{X}_{M,p,q}$  onde cada imagem é uma linha de  $\mathbf{X}$ . A matriz de covariância é obtida por

$$\mathbf{C}_{i,j} = E \left\{ (x_i - \mu_i)(x_j - \mu_j)^T \right\} \quad (4.3)$$

onde:

$E$ : Esperança matemática;

$C_{i,j}$ : elemento  $i,j$  da matriz de covariância;

$x_i$ : vetor da  $i$ -ésima coluna da matriz  $\mathbf{X}$ ;

$\mu_i$ : vetor médio da  $i$ -ésima coluna da matriz  $\mathbf{X}$ .

Os elementos da matriz de covariância devem ser calculados para todas as combinações possíveis dentre os vetores colunas  $i$  e  $j$  da matriz  $\mathbf{X}$ , dessa forma  $C_{i,j} = C_{j,i}$ . A matriz de covariância é simétrica em relação a diagonal principal.

A esperança  $E$  de um grande conjunto de dados pode ser aproximada para a média aritmética desse conjunto. Assim

$$C_{i,j} \approx \frac{1}{M} \sum \left[ (x_i - \mu_i)(x_j - \mu_j)^T \right] \quad (4.4)$$

Notamos que a aproximação da Equação (4.4) mede a variância entre os vetores de dados  $x_i$  e  $x_j$ .

Os autovetores da matriz de covariância são determinados através da seguinte expressão

$$\det(\mathbf{C} - \lambda_i \mathbf{I}) = 0 \quad (4.5)$$

onde:

$\lambda_i$ :  $i$ -ésimo autovalor

$\mathbf{I}$ : matriz identidade

O cálculo dos autovetores da matriz de covariância é uma tarefa com um alto custo computacional, devido a sua elevada dimensão. Por exemplo, uma imagem de

dimensão  $640 \times 480 = 307200$  resultaria em uma matriz de covariância de dimensão  $307200 \times 307200$ , o que seria inviável computacionalmente. Em consequência disso, outra forma de se determinar os autovetores é mostrada a seguir.

Seja  $\Pi_i = \Omega_i - \mu$ , onde  $\mu$  é o vetor imagem média e  $\mathbf{V} = [\Pi_1, \Pi_2, \dots, \Pi_M]_{N,M}$ ,  $N$  é o número de *pixels* da imagem e  $M$  o número de imagens a serem projetadas. Vamos então representar a matriz de covariância da seguinte forma.

$$\mathbf{C} = \mathbf{V}\mathbf{V}^T \quad (4.6)$$

Note que na Equação (4.6) não foi levado em conta o fator  $1/M$  devido ao fato deste somente escalar os autovalores, não implicando nenhuma mudança na determinação final dos autovetores.

A Equação (4.5) pode ser reescrita da seguinte forma.

$$\mathbf{V}\mathbf{V}^T \mathbf{e}_i = \lambda_i \mathbf{e}_i \quad (4.7)$$

onde  $\mathbf{e}_i$  denota os autovetores da matriz de covariância e  $\lambda_i$  representa os autovalores.

Para resolver a Equação (4.7), pode-se escrever a seguinte expressão.

$$\mathbf{V}^T \mathbf{V} \mathbf{b}_i = \lambda_i \mathbf{b}_i \quad (4.8)$$

como a disposição das matrizes  $\mathbf{V}$  e  $\mathbf{V}^T$  foram comutadas, obtém-se novos autovetores  $\mathbf{b}_i$ .

Multiplicando a Equação (4.8) por  $\mathbf{V}$ , obtém-se

$$\mathbf{V}\mathbf{V}^T\mathbf{V}\mathbf{b}_i = \mathbf{V}\lambda_i\mathbf{b}_i \quad (4.9)$$

$$\mathbf{V}\mathbf{V}^T(\mathbf{V}\mathbf{b}_i) = \lambda_i(\mathbf{V}\mathbf{b}_i) \quad (4.10)$$

Comparando a Equação (4.7) e (4.10) notamos que  $\mathbf{e}_i = \mathbf{V}\mathbf{b}_i$ . Assim os autovetores do problema original podem ser obtidos simplesmente multiplicando os autovetores  $\mathbf{b}_i$  por  $\mathbf{V}$ . Através de uma nova representação da matriz de covariância, Equação (4.6), o problema do autovalor foi agora reduzido de uma dimensão  $N \times N$  para  $M \times M$ , onde  $N$  é a dimensão original da matriz de covariância e  $M$  é o número de imagens de treinamento consideradas.

## 4.2 Escolha das Características

Uma vez definidos os autovetores, o conjunto de imagens de treinamento deve ser mapeado em um espaço de características. O problema agora é reduzir e escolher um número reduzido de autovetores de forma que estes representem o espaço original da melhor forma, ou seja, representando uma maior quantidade de informação.

Devido à característica da matriz de covariância, um determinado número de autovalores são nulos [33]. É possível calcular o número de autovetores nulos pela Equação (4.11).

$$T = N - (M - 1) \quad (4.11)$$

onde:

$T$  : número de autovetores nulos;

$N$  : dimensão do espaço original;

$M$  : dimensão do número de imagens de treinamento.

O método a seguir mostra um caminho de se reduzir o número de autovetores, perdendo o mínimo de informação.

### 4.3 Método m

Normalmente é desejável reduzir o espaço de características, tanto quanto possível, e assim obter uma redução de dimensão. Geralmente, para cada aplicação deve-se ter um compromisso entre a redução da dimensão e a perda de informação que ela pode acarretar. Para se obter uma maior quantidade de informação, devem ser selecionados aqueles autovetores que possuem os maiores autovalores associados. Uma medida da quantidade de informação mantida pelos autovetores pode ser estimada por

$$I_K = 100 \frac{\sum_{i=1}^K \lambda_i}{\sum_{i=1}^N \lambda_i} \quad (4.12)$$

onde:

$\lambda_i$  : representa a intensidade do  $i$ -ésimo maior autovalor;

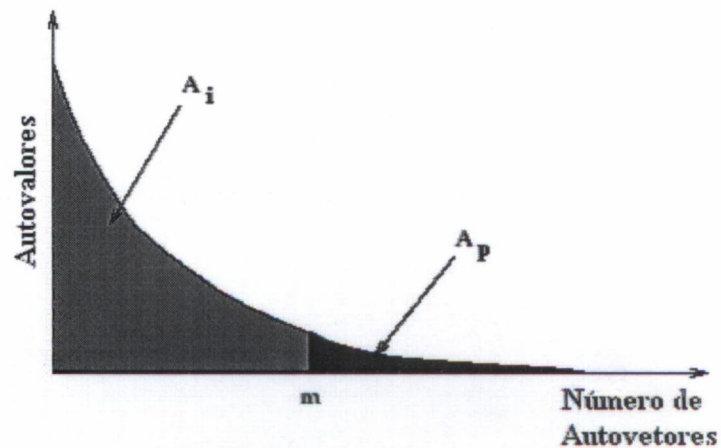
$K$  : representa o número de autovalores a serem utilizados;

$N$  : representa a dimensão total;

$I_K$  : representa o percentual de informação (variância) que é mantida na transformação.



A Equação (4.12) pode ser melhor entendida através da Figura 4.3, onde a intensidade dos autovalores ou o montante de informação, representada no eixo vertical, pode ser associada com o número de autovetores no eixo horizontal.



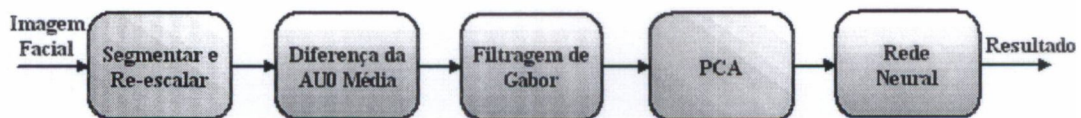
**Figura 4.3:** Autovalores *versus* número de autovetores

Na Figura 4.3,  $m$  é um limiar. A partir desse ponto os autovetores são ignorados, pois seu autovalor associado possui uma energia muito baixa. A área  $A_i$  representa a informação mantida na transformação e  $A_p$ , representa a informação perdida na transformação. Quanto maior o valor de  $m$ , maior é a informação mantida na transformação. Esta técnica é chamada de método  $m$  [33].

# Capítulo 5

## Sistema de Reconhecimento de Expressões

A Figura 5.1 mostra a estrutura do sistema proposto.



**Figura 5.1:** Estrutura do sistema proposto

Na Figura 5.1, a imagem facial é aquela obtida do banco de dados, cada bloco representa uma etapa do pré-processamento que é descrita a seguir e o resultado é a classificação da ação facial através dos neurônios de saída da rede neural.

A imagem capturada diretamente da câmera contém muita informação desnecessária para o reconhecimento das expressões, tal como o fundo e as legendas. Para excluir tais informações como também definir uma região de interesse, as imagens são segmentadas em face superior e inferior e um aproximado alinhamento das faces é efetuado via um programa aplicativo listado no Apêndice B.

As Figuras 5.2 e 5.3 mostram as imagens segmentadas, para a AU 0, correspondente à imagem neutra e para o caso de uma expressão de espanto, contendo as AUs 1, 2, 5 e 27.



**Figura 5.2:** Face superior segmentada da (a) imagem neutra e (b) com uma expressão de espanto.



**Figura 5.3:** Face inferior segmentada da (a) imagem neutra e (b) com uma expressão de espanto.

O alinhamento, não muito criterioso das imagens, se deve ao fato da dependência da posição dos olhos, nariz e queixo para a definição da janela de corte. As imagens segmentadas diferem no seu tamanho entre os quadros de uma seqüência para cada expressão e indivíduo. A regra para o local exato do corte e a forma aproximada do alinhamento das imagens é apresentada no Apêndice B.

Para a formação da matriz de covariância, é necessário que as imagens possuam um mesmo tamanho. Dessa forma, as imagens foram re-escaladas para um tamanho comum.

O tamanho médio das imagens segmentadas é de  $105 \times 190$  e  $133 \times 192$  pixels para a parte superior e inferior da face, respectivamente. Apesar da face inferior ser 28% maior do que a face superior, foi decidido torná-las com o mesmo tamanho de  $30 \times 45$  pixels para ser possível comparar a eficácia do sistema para ambas as regiões faciais. A forma de dezimação escolhida foi a de *vizinhança mais próxima* sem a prévia aplicação de nenhum filtro anti-recobrimento. O cálculo dos parâmetros do filtro anti-

recobrimento é uma tarefa difícil de ser efetuada visto que as imagens segmentadas possuem tamanhos diferentes e, portanto, o fator de escala, necessário para a determinação das características do filtro, é variável. A escolha de uma imagem de dimensão maior aumenta o custo computacional do sistema devido ao acréscimo de operações para a filtragem, determinação dos componentes principais e projeções no espaço de características. A Figura 5.4 mostra as imagens da parte superior e inferior da face segmentadas e re-escaladas.



**Figura 5.4:** Imagens segmentadas e re-escaladas da (a) face superior e (b) face inferior.

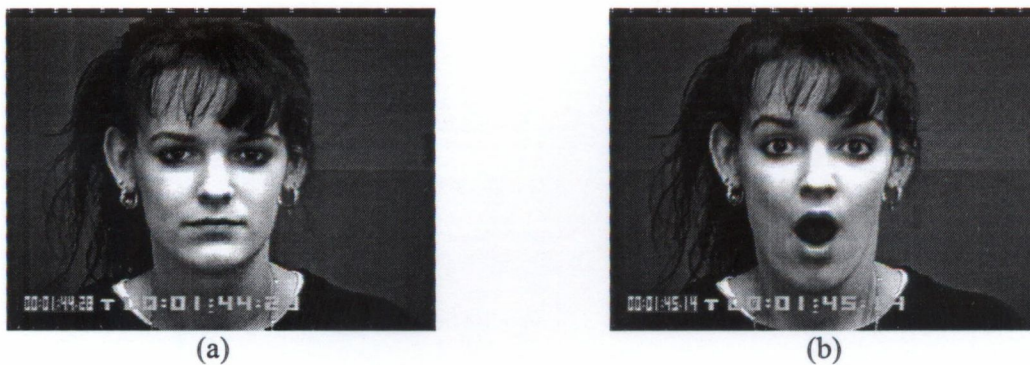
As imagens são subtraídas da face neutra (AU 0) média e então são filtradas por um conjunto de filtros de Gabor com diversas orientações. As imagens filtradas são projetadas nos componentes principais gerados pelo PCA. Em seguida, os coeficientes resultantes da projeção são usados como entrada para a rede neural que fará a classificação da ação facial. Nas próximas seções deste capítulo, é analisado com mais detalhe cada passo do pré-processamento.

## 5.1 Banco de Dados

Nas experiências e testes apresentados nestes trabalho foram utilizadas imagens da Base de Dados Cohn-Kanade [18]. Esta base contém 480 seqüências de imagens de 94 indivíduos heterogêneos com idades entre 18 e 50, de ambos os sexos, de diversos grupos raciais, como ibero-americanos, afro-americanos e orientais. Os indivíduos são orientados a executarem uma única AU ou combinações de AUs, eles estão localizados

diretamente à frente da câmera e efetuam, involuntariamente, pequenos movimentos com a cabeça na direção frontal e lateral à câmera. As seqüências de imagens começam com uma face neutra e são digitalizadas com uma resolução de  $640 \times 480$  pixels com 8 bits na escala de cinza a uma taxa de 30 quadros/s.

A Figura 5.5 apresenta alguns exemplos de imagens pertencentes ao banco de dados.










**Figura 5.5:** (a) Primeiro quadro da seqüência e (b) o último quadro de um indivíduo apresentando uma expressão de espanto.

As expressões presentes no banco de dados foram classificadas previamente por especialistas no FACS.

Após uma análise minuciosa nos arquivos do banco de dados, no qual foi verificada a integridade das imagens na seqüência e também se as ações faciais realmente correspondiam à classificação feita pelos autores do banco de dados, foram selecionados 7 padrões de ação facial da parte superior e 7 padrões da parte inferior da face. A escolha dos padrões foi baseada no fato deles aparecerem com maior frequência.

A Tabela 5.1 mostra o número do padrão, uma amostra da imagem a ser classificada, as AUs presentes no padrão de acordo com o FACS, e o número de amostras encontradas para cada expressão no banco de dados, perfazendo um total de 229 expressões para a face superior.








**Tabela 5.1:** Padrões de ação facial da parte superior da face

| Padrão | Imagem de Amostra   | Unidades de Ação | Amostras |
|--------|---|------------------|----------|
| 1      |    | 0                | 70       |
| 2      |    | 1+2              | 26       |
| 3      |    | 1+2+5            | 43       |
| 4      |    | 4+7              | 14       |
| 5      |    | 4+6+7+9          | 12       |
| 6      |   | 4+7+9            | 20       |
| 7      |  | 6                | 44       |

A Tabela 5.2 apresenta as ações faciais selecionadas da face inferior, totalizando 282 expressões.

Tanto as imagens da face inferior como superior têm suas expressões neutras chamadas de AU 0. Podemos notar, das imagens das Tabelas 5.1 e 5.2, que existem variações de luminosidade, cor da pele, sexo e até mesmo a presença de cabelo e batom em algumas amostras, confirmando a heterogeneidade do conjunto de dados que é usado nesse trabalho.

**Tabela 5.2:** Padrões de ação facial da parte inferior da face

| Padrão | Imagem de Amostra   | Unidades de Ação | Amostras |
|--------|---|------------------|----------|
| 1      |    | 0                | 55       |
| 2      |    | 25               | 39       |
| 3      |    | 26               | 49       |
| 4      |    | 27               | 45       |
| 5      |    | 12+25            | 53       |
| 6      |   | 15+17            | 24       |
| 7      |  | 20+25            | 17       |

## 5.2 Diferença da AU 0 Média

As vantagens do uso da diferença da face neutra no reconhecimento de ações faciais e emoções incluem robustez às mudanças na luminosidade, remoção na variação da superfície entre indivíduos e ênfase nos aspectos dinâmicos da seqüência de imagens [34].

Em [2], Donato *et. al.* propõem o uso da diferença entre o último quadro da seqüência de vídeo, expressão com a máxima intensidade, de todos os quadros anteriores à seqüência, começando com o quadro neutro. Eles demonstraram que esta técnica pode melhorar o desempenho do sistema de reconhecimento. Ainda em [2], foi

mostrado ser necessário efetuar um apurado alinhamento dos olhos e boca em cada seqüência de quadros para deixá-los rigorosamente na mesma posição a fim de tornar a aplicação direta de métodos tais como PCA e ICA mais eficientes.

Fasel e Luetin [10] extraíram as características faciais subtraindo a imagem com sua máxima intensidade da imagem neutra para um mesmo indivíduo. A base de dados utilizada foi adquirida em condições controladas, isto é, as faces mantiveram uma posição constante ao longo de todas as seqüências sem mudanças na escala, rotação e luminosidade. A principal desvantagem desta técnica é a dificuldade na determinação da face neutra para um mesmo indivíduo para aplicações em tempo real.

Neste trabalho apresentamos uma técnica robusta e de reduzida complexidade computacional. Propomos a análise da diferença, *pixel a pixel*, entre a imagem de máxima intensidade de ação facial da média de todas as imagens neutras (AU 0) do banco de dados. Nos experimentos descritos no Capítulo 6, esta aproximação sustenta as vantagens do uso da diferença com a imagem neutra para um mesmo indivíduo proposto por Fasel e Luetin [10], eliminando a análise e a definição prévia da face neutra para um dado indivíduo, tornando a técnica proposta mais eficiente e adequada para operar em sistemas que atuem em tempo real. A face neutra média pode ser determinada na etapa de treinamento do sistema.

A Figura 5.6 mostra um exemplo para a face superior com ação facial de máxima intensidade, a face neutra média e a diferença entre elas.

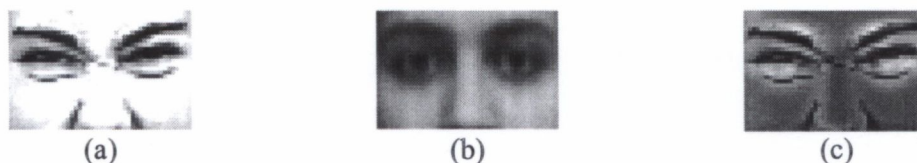
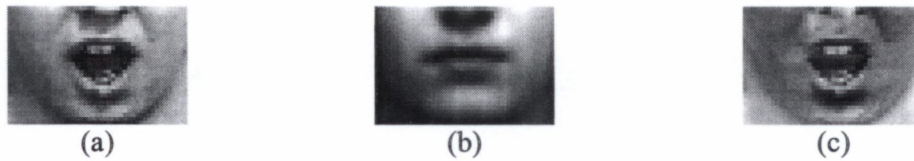


Figura 5.6: Ação facial da face superior: (a) imagem original (b) imagem neutra média e (c) imagem diferença



Do mesmo modo que a figura anterior, a Figura 5.7 mostra as imagens para a parte inferior da face.



**Figura 5.7:** Ação facial da face inferior: (a) imagem original (b) imagem neutra média e (c) imagem diferença







Pode-se notar que a diferença entre as imagens, original e neutra, realça justamente a área em que houve um maior movimento da face com relação à imagem neutra, facilitando a detecção via wavelets de Gabor para posterior classificação através da rede neural.

### 5.3 Aplicação das Wavelets de Gabor

Os experimentos foram realizados aplicando as wavelets de Gabor na imagem diferença. Várias configurações de filtros foram utilizadas como pode ser visto na Tabela 6.1. Foi utilizado o desvio padrão  $\sigma = \pi$ , o mesmo aplicado por [18][19].

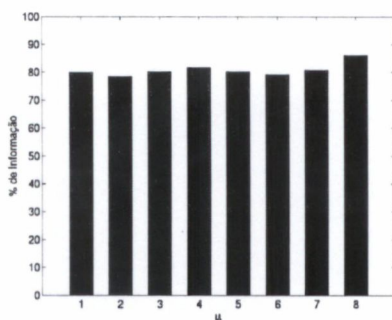
A Tabela 5.3 mostra a aplicação de dois filtros de Gabor com frequências espaciais  $k_v = \pi/2$  e  $\pi/(2\sqrt{2})$  e orientação  $\theta_\mu = \frac{\pi}{8}$  na imagem original e na diferença.

**Tabela 5.3:** Filtros de Gabor aplicados à imagem original e a imagem diferença.

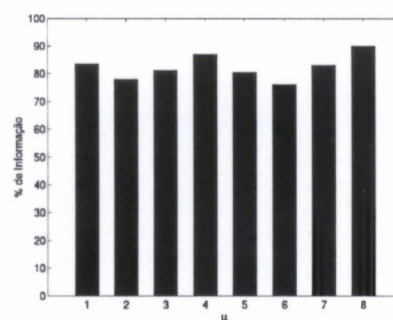
|           | Sem Gabor   | $k_v = \pi/2$<br>$\theta_\mu = \frac{\pi}{8}$                                     | $k_v = \pi/(2\sqrt{2})$<br>$\theta_\mu = \frac{\pi}{8}$                            |
|-----------|---|---|--|
| Original  |  |  |  |
| Diferença |  |  |  |

### 5.4 Aplicação do PCA

Foram extraídas, separadamente, os componentes principais de 229 imagens do conjunto de ações faciais da parte superior e 282 ações faciais da parte inferior da face. Através do método m, trinta componentes principais foram selecionadas para cada configuração de filtro. A quantidade de informação estimada foi de aproximadamente 80% e 82% para a parte superior e inferior, respectivamente. A Figura 5.8 mostra a distribuição do percentual de informação mantido, segundo o método m, para os filtros de Gabor com frequência espacial  $k_v = \frac{\pi}{2}$  e diversas orientações  $\theta_\mu = \mu \frac{\pi}{8}$ , onde  $\mu = 1 \dots 8$ .



(a)

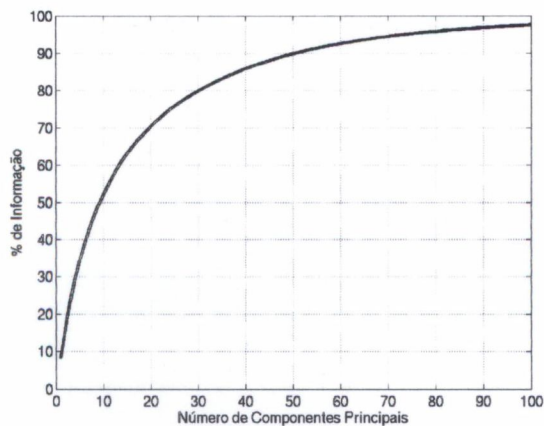


(b)

**Figura 5.8:** Distribuição do percentual da quantidade de informação mantida para os 30 autovetores da (a) face superior (b) face inferior.

Para um maior percentual de informação, é necessário um número maior de componentes principais, diminuindo o ganho de dimensão. A Figura 5.9 mostra a distribuição da quantidade de informação estimada através do método m em relação ao número de componentes principais para a frequência espacial  $k_v = \frac{\pi}{2}$  e orientação

$$\theta_\mu = \frac{\pi}{8}.$$




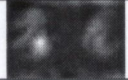


**Figura 5.9:** Percentual da quantidade de informação em relação ao número de componentes principais para a face superior

A Figura 5.9 mostra que para mantermos um percentual de informação de 90%, seria necessário 50 componentes principais, como será visto na Seção 5.5, o número de neurônios na camada de entrada da rede neural é diretamente proporcional ao número de componentes principais e o número de filtros utilizados. Portanto, um acréscimo significativo no número de componentes principais aumentaria a complexidade do treinamento da rede neural.

A Tabela 5.4 mostra dois componentes principais mais representativos da imagem original e da imagem diferença após a aplicação do filtro de Gabor com frequência espacial  $k_v = \pi/2$  e orientação  $\theta_\mu = \pi/8$  para a parte superior da face.

**Tabela 5.4:** Dois componentes principais para a imagem original e a imagem diferença

| Componente Principal | Original  | Diferença  |
|----------------------|---|--|
| 1                    |  |  |
| 2                    |  |  |

## 5.5 Rede Neural

Duas redes neurais do tipo *MultiLayer Perceptron* (MLP) foram projetadas para a parte superior e inferior da face. Ambas foram treinadas utilizando os algoritmos *backpropagation* rápidos como *Weight Decay* [35], disponíveis no simulador SNNS (*Stuttgart Neural Network Simulator*).

O SNNS é um programa aplicativo para simulação de redes neurais desenvolvido pelo Instituto de Sistemas Paralelos e Distribuídos de Alto Desempenho (IPVR) da Universidade de Stuttgart – Alemanha. O objetivo do projeto SNNS é criar um ambiente eficiente e flexível de simulação para pesquisadores e aplicações em redes neurais. Escolhemos o simulador SNNS devido à vasta documentação, e ainda por estar sendo utilizado em diversos centros de pesquisa, e possuindo uma lista de discussão, criada em 1994, com inúmeros comentários e sugestões.

O número de neurônios na camada de entrada é dado por:

$$N_{in} = N_{cp} \cdot N_{fg} \quad (5.1)$$

onde:

$N_{in}$ : número de neurônios da camada de entrada;

$N_{cp}$  : número de componentes principais;

$N_{fg}$  : número de filtros de Gabor.

O número de neurônios da camada escondida dependerá da estrutura do experimento a ser utilizado. Como é visto no Capítulo 5, alguns experimentos foram realizados utilizando filtros de Gabor com 2 frequências espaciais e 8 orientações, nesse caso, o número de neurônios na camada de entrada é  $N_{in} = 30.2.8 = 480$ .

Na camada escondida foi utilizada a função de ativação sigmoideal. Definimos o número de neurônios da camada escondida através da seguinte relação [36].

$$N_h = \frac{2}{3}(N_{in} + N_{out}) \tag{5.2}$$

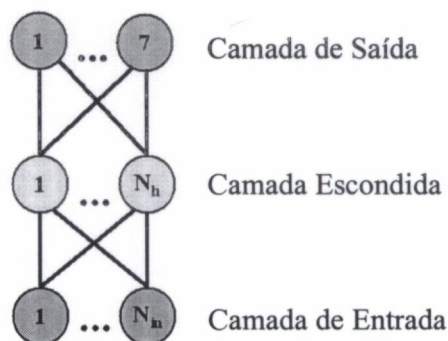
onde:

$N_h$ : número de neurônios da camada escondida;

$N_{in}$ : número de neurônios da camada de entrada;

$N_{out}$ : número de neurônios da camada de saída.

O número de neurônios na camada de saída é determinado pelo número de expressões a serem classificadas, ou seja, 7 neurônios. A função de ativação é a mesma da camada escondida. A Figura 5.10 ilustra a disposição dos neurônios e conexões da rede neural.



**Figura 5.10:** Estrutura da rede neural

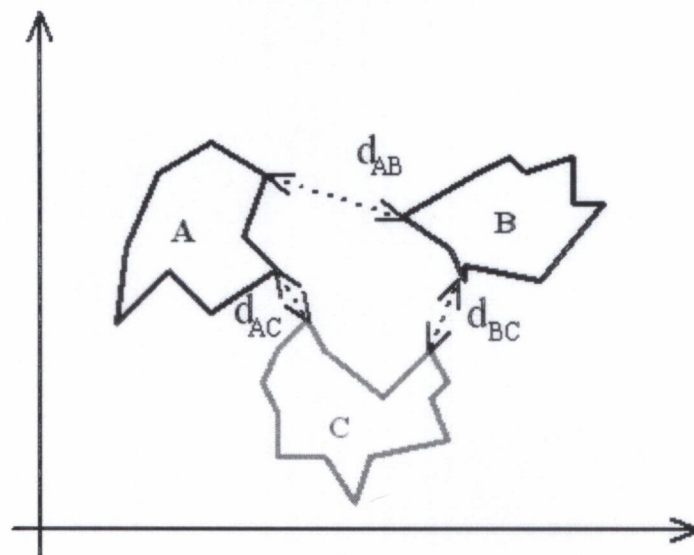
Para criar um conjunto de treinamento homogêneo, ou seja, todos os padrões de expressões com o mesmo número de amostras, novos vetores foram criados através da adição de um ruído branco aos já existentes. A amplitude do ruído é tal que não haja possibilidade do vetor mudar de classe (cada classe corresponde a um dos padrões de expressão descritos na Tabela 5.1). Isso é garantido através do cálculo da distância Euclidiana mínima ( $d_{\min}$ ) entre uma classe e todas as outras.

A distância mínima é definida como

$$d_{\min} = \min\{d_{ij}\} \quad (5.3)$$

onde,  $d_{ij}$  é a distância Euclidiana entre a classe  $i$  e  $j$ .

A Figura 5.11 mostra um exemplo.



**Figura 5.11:** Representação de 3 classes de dados

Sejam as classes A, B e C, para criarmos um novo vetor pertencente a classe B, acrescentaremos um ruído branco a um vetor desta classe tal que a magnitude não exceda  $d_{BC}/2$ . Cada elemento do vetor ruído branco é randômico, porém seu módulo

não deve exceder um determinado valor para garantir a independência entre as classes.

A Equação (5.4) define o valor máximo do  $i$ -ésimo elemento do vetor ruído.

$$x_i = \frac{d_{\min}}{2\sqrt{N}} \quad (5.4)$$

onde:

$x_i$  :  $i$ -ésimo elemento do vetor ruído branco;

$d_{\min}$  : distância Euclidiana mínima entre duas classes;

$N$ : dimensão do vetor ruído branco.

A Tabela 5.5 apresenta o número de amostras por padrão de expressão (Tabela 5.1) obtidos para os conjuntos de treinamento e teste para a parte superior da face.

**Tabela 5.5:** Dimensão do conjunto de treinamento e teste para a parte superior da face

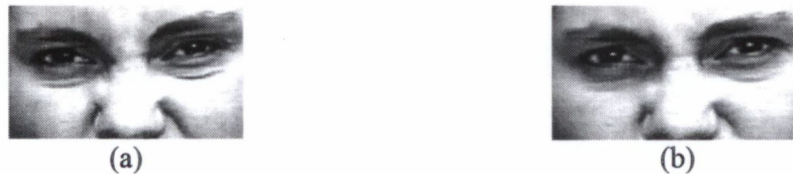
|               | Total de Amostras | Amostra/Padrão |
|---------------|-------------------|----------------|
| <b>Treino</b> | 371               | 53             |
| <b>Teste</b>  | 119               | 17             |
| <b>Total</b>  | <b>490</b>        | <b>70</b>      |

A Tabela 5.5 apresenta o número de amostras por padrão de expressão (Tabela 5.2) obtidos para os conjuntos de treinamento e teste para a parte inferior da face.

**Tabela 5.6:** Dimensão do conjunto de treinamento e teste para a parte inferior da face

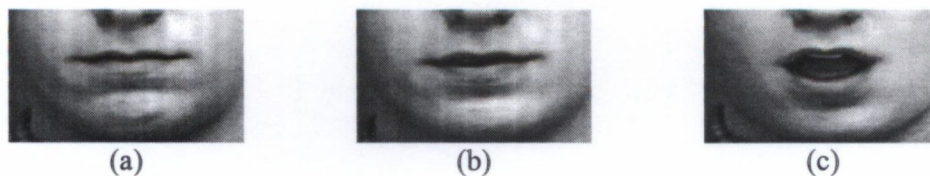
|               | Total de Amostras | Amostra/Padrão |
|---------------|-------------------|----------------|
| <b>Treino</b> | 287               | 41             |
| <b>Teste</b>  | 98                | 14             |
| <b>Total</b>  | <b>385</b>        | <b>55</b>      |

Devido à grande dificuldade das redes, chamadas redes primárias, de distinguir os padrões 5 e 6 da parte superior e os padrões 1, 2 e 3 da parte inferior da face, foram criadas redes secundárias especializadas para cada rede primária, a fim distinguir tais padrões. A Figura 5.12 mostra os padrões 5 e 6 da face superior.



**Figura 5.12:** Exemplos do (a) padrão 5, AU4+6+7+9; (b) padrão 6, AU 4+7+9

A Figura 5.13 mostra os padrões 1, 2 e 3 da face inferior.



**Figura 5.13:** Exemplos do (a) padrão 1, AU 0; (b) padrão 2, AU 25; (c) padrão 3, AU 26

As Figuras 5.12 e 5.13 mostram que mesmo nós seres humanos não treinados em FACS temos dificuldade para distinguir os padrões mostrados nessas figuras. Buscando melhorar a capacidade de discriminação, são criadas redes secundárias. O número de neurônios na camada de entrada, a função de ativação e o algoritmo de treinamento são os mesmos da rede primária. Na camada escondida, utilizamos o mesmo número de neurônios da camada de entrada. O número de neurônios da camada de saída corresponde ao número de padrões que se deseja distinguir, ou seja, dois para a parte superior da face e três para a parte inferior. Para a parte superior, a segunda rede somente é aplicada quando o neurônio 5 é ativado na rede primária. Para a parte inferior



da face, a segunda rede é aplicada quando qualquer um dos três primeiros neurônios for ativado.

# Capítulo 6

## Resultados Experimentais

Os resultados experimentais para diversas configurações dos filtros utilizados são avaliados para a face superior na Seção 6.1 e na Seção 6.2 para a face inferior. Os resultados obtidos de simulação do sistema operando em tempo real são apresentados na Seção 6.3.

### 6.1 Face Superior

Foi comparado o desempenho do sistema de reconhecimento em 16 experimentos. A Tabela 6.1 mostra as categorias dos filtros empregados, o tipo de diferença de quadro aplicada e a taxa de reconhecimento obtida. A segunda coluna da Tabela 6.1 mostra o tipo, a frequência espacial e o número de orientações usados em cada filtro. Nos experimentos 1, 4, 13 e 15 não foi utilizada nenhuma diferença de quadros, o PCA foi aplicado diretamente nas imagens. Nos experimentos 2 e 5 foi realizada a diferença do último quadro de maior intensidade da expressão com o primeiro quadro neutro daquela seqüência. Nos outros experimentos, o quadro com a expressão mais intensa foi subtraído do quadro neutro (AU 0) médio de todo o banco de dados. Nos experimentos 13 e 14 utilizamos o filtro de Sobel [37] que é um operador muito utilizado pela comunidade científica de processamento de imagens. O filtro de Sobel é capaz de detectar regiões na imagem que possuam componentes de alta frequência.

Tabela 6.1: Estrutura e resultados dos experimentos

| Experim. | Categoria do Filtro                                  | Diferença de Quadro | Taxa de Reconhecimento |
|----------|--|---------------------|------------------------|
| 1        | Gabor $k_v = \pi/2$ e $\mu = 1...8$                  | Não                 | 67,22 %                |
| 2        | Gabor $k_v = \pi/2$ e $\mu = 1...8$                  | Face neutra         | 62,18 %                |
| 3        | Gabor $k_v = \pi/2$ e $\mu = 1...8$                  | Face neutra média   | 83,19 %                |
| 4        | Gabor $k_v = \pi/2, \pi/2\sqrt{2}$ e $\mu = 1...8$   | Não                 | 56,30 %                |
| 5        | Gabor $k_v = \pi/2, \pi/2\sqrt{2}$ e $\mu = 1...8$   | Face neutra         | 52,94 %                |
| 6        | Gabor $k_v = \pi/2, \pi/2\sqrt{2}$ e $\mu = 1...8$   | Face neutra média   | 63,22 %                |
| 7        | Gabor $k_v = \pi/\sqrt{2}$ e $\mu = 1...8$           | Face neutra média   | 63,86 %                |
| 8        | Gabor $k_v = \pi/2\sqrt{2}$ e $\mu = 1...8$          | Face neutra média   | 49,57 %                |
| 9        | Gabor $k_v = \pi/4$ e $\mu = 1...8$                  | Face neutra média   | 43,69 %                |
| 10       | Gabor $k_v = \pi/2$ e $\mu = 2, 4, 6$ e 8            | Face neutra média   | 35,29 %                |
| 11       | Gabor $k_v = \pi/2$ e $\mu = 4$ e 8                  | Face neutra média   | 31,09 %                |
| 12       | Gabor $k_v = \pi/2$ e $\mu = 1...16$                 | Face neutra média   | 76,47 %                |
| 13       | Sobel $\mu = 4$ e 8                                  | Não                 | 47,89 %                |
| 14       | Sobel $\mu = 4$ e 8                                  | Face neutra média   | 54,62 %                |
| 15       | Sem filtragem (PCA aplicado diretamente nas imagens) | Não                 | 54,62 %                |
| 16       | Sem filtragem (PCA aplicado diretamente nas imagens) | Face neutra média   | 62,18 %                |

Comparando o resultado obtido nos experimento 1, 2, 3 e 4, 5, 6 é possível perceber que um número reduzido de filtros e a utilização da técnica de diferença do quadro neutro médio melhoram significativamente a taxa de reconhecimento. Comparando os experimentos 3, 7, 8, 9 podemos verificar que para frequências espaciais maiores e com mesmo número de orientações, o filtro de Gabor torna-se mais sensível as características faciais, porém a escolha de frequências espaciais muito altas

como  $k_v = \pi/\sqrt{2}$  provoca perda de informações faciais, devido a destacar apenas transições mais abruptas na imagem, excluindo transições mais suaves, ou seja, frequências menores, importantes para o reconhecimento. Comparando os experimentos 3, 10, 11, 12 nota-se que quando o número de orientações aumenta, o resultado melhora. No entanto, deve-se considerar que o treinamento da rede neural torna-se mais complexo quando é aumentado o número de neurônios, por isso a taxa de reconhecimento obtida no experimento 12 foi inferior ao esperado em relação ao experimento 3.

Nos experimentos 13 e 14 podemos notar que a taxa de reconhecimento, quando são aplicados os filtros de Sobel [37], está muito abaixo das obtidas pelos experimentos 3 e 12. Esse resultado é devido ao filtro de Sobel ser menos seletivo e também por possuir somente 2 orientações espaciais  $\frac{\pi}{2}$  e  $\pi$ . A Figura 6.1 mostra a distribuição no domínio de Fourier do filtro de Sobel com o orientações espaciais  $\frac{\pi}{2}$  e  $\pi$ , para uma janela de dimensão  $90 \times 90$ .



**Figura 6.1:** Distribuição no domínio de Fourier do filtro de Sobel para orientação espacial: (a)  $\frac{\pi}{2}$  e (b)  $\pi$

Os experimentos 15 e 16 mostram que a aplicação do PCA diretamente nas imagens, sem aplicação do filtro de Gabor, não apresentou ainda uma taxa de reconhecimento satisfatória, porém suficiente para superar alguns resultados obtidos nos

experimentos utilizando o filtro de Gabor com um número menor de orientações e com baixas frequências espaciais.

Foi testada a aplicação do classificador baseado na distância Euclidiana, ao invés da Rede Neural, no experimento 3, porém a taxa de reconhecimento foi insatisfatória, alcançando somente 10,71 %. Outro tipo de classificador aplicado para identificar faces efetuando expressões faciais foi recentemente proposto por Martinez [38]. Ele utilizou a técnica de fluxo ótico para designar um peso maior aos *pixels* mais representativos da imagem e com isso obteve uma taxa de reconhecimento um pouco superior à aplicação do PCA seguido do classificador baseado na distância Euclidiana.

A Tabela 6.1 também mostra que o uso da diferença entre o quadro de máxima intensidade do quadro neutro estimado de uma mesma seqüência, experimentos 2 e 5, não obtiveram uma adequada taxa de reconhecimento, mesmo quando comparados com os experimentos 1 e 4. Esta baixa taxa de reconhecimento deve-se a aplicação de uma segmentação aproximada e nenhum alinhamento dos olhos no sistema proposto. Para a obtenção de melhores resultados é necessário um estágio de pré-processamento mais apurado, aumentando a complexidade computacional.

O emprego da face neutra média nos experimentos 3 e 6 mostram que esta técnica é eficiente. O uso do quadro neutro médio aumenta a robustez do pré-processamento de imagens não perfeitamente alinhadas, mantendo a vantagem do uso da diferença das imagens.

A matriz de confusão obtida no experimento 3 é mostrada na Tabela 6.2, onde as linhas são os padrões de entrada corretos e as colunas são os padrões de saída classificados pela rede neural primária. É possível notar que o padrão 6 se confunde com o padrão 5. Realmente, eles são muito similares, como mostrado na Figura 5.12.

**Tabela 6.2:** Matriz de confusão do experimento 3

| Seq. | 1  | 2  | 3  | 4  | 5  | 6  | 7  |
|------|----|----|----|----|----|----|----|
| 1    | 14 | 1  | 0  | 1  | 0  | 0  | 1  |
| 2    | 3  | 12 | 2  | 0  | 0  | 0  | 0  |
| 3    | 1  | 2  | 14 | 0  | 0  | 0  | 0  |
| 4    | 0  | 0  | 0  | 15 | 0  | 2  | 0  |
| 5    | 0  | 0  | 0  | 0  | 17 | 0  | 0  |
| 6    | 0  | 0  | 0  | 0  | 6  | 11 | 0  |
| 7    | 1  | 0  | 0  | 0  | 0  | 0  | 16 |

Desta forma, foi aplicado a segunda rede neural, somente quando o neurônio 5 é ativado, para distinguir o padrão 5 e 6. A matriz de confusão final é mostrada na Tabela 6.3.

**Tabela 6.3:** Matriz de confusão para o experimento 3 após a aplicação da segunda rede neural

| Seq. | 1  | 2  | 3  | 4  | 5  | 6  | 7  |
|------|----|----|----|----|----|----|----|
| 1    | 14 | 1  | 0  | 1  | 0  | 0  | 1  |
| 2    | 3  | 12 | 2  | 0  | 0  | 0  | 0  |
| 3    | 1  | 2  | 14 | 0  | 0  | 0  | 0  |
| 4    | 0  | 0  | 0  | 15 | 0  | 2  | 0  |
| 5    | 0  | 0  | 0  | 0  | 17 | 0  | 0  |
| 6    | 0  | 0  | 0  | 0  | 2  | 15 | 0  |
| 7    | 1  | 0  | 0  | 0  | 0  | 0  | 16 |

Notamos que a segunda rede neural foi capaz de distinguir corretamente o padrão 5 do padrão 6, como um humano treinado no FACS.

O resultado final para o experimento 3 é mostrado na Tabela 6.4.

**Tabela 6.4:** Resultado final do experimento com a face superior

|                                   |                |
|-----------------------------------|----------------|
| <b>Amostras de Teste</b>          | 119            |
| <b>Classificados Corretamente</b> | 103            |
| <b>Erros</b>                      | 16             |
| <b>Taxa de Reconhecimento</b>     | <b>86.55 %</b> |

A Tabela 6.4 apresenta o número de amostras em teste, o número de amostras corretamente classificadas e a taxa de reconhecimento obtida.

## 6.2 Face Inferior

Com base nos resultados obtidos dos testes com a porção superior da face, aplicamos a mesma estrutura do experimento 3 (Tabela 6.1) para a parte inferior da face. A taxa de reconhecimento obtida, somente com a aplicação da primeira rede neural, foi de 78,57 %. A Tabela 6.5 mostra a matriz de confusão para o experimento com a face inferior utilizando a diferença do quadro com expressão de maior intensidade e o quadro neutro médio.

**Tabela 6.5:** Matriz de confusão para o experimento com a face inferior

| Seq. | 1 | 2  | 3 | 4  | 5  | 6  | 7  |
|------|---|----|---|----|----|----|----|
| 1    | 5 | 6  | 0 | 1  | 0  | 2  | 0  |
| 2    | 1 | 12 | 1 | 0  | 0  | 0  | 0  |
| 3    | 0 | 6  | 7 | 1  | 0  | 0  | 0  |
| 4    | 0 | 0  | 3 | 11 | 0  | 0  | 0  |
| 5    | 0 | 0  | 0 | 0  | 14 | 0  | 0  |
| 6    | 0 | 0  | 0 | 0  | 0  | 14 | 0  |
| 7    | 1 | 0  | 0 | 0  | 0  | 0  | 14 |

É possível notar na Tabela 6.5, a confusão entre os padrões 1, 2 e 3 e de fato eles são muito similares como mostrado na Figura 5.13. Buscando uma melhor distinção entre tais padrões pelo sistema proposto, foi usada uma rede secundária especializada com a mesma estrutura daquela empregada na parte superior da face. A nova matriz de confusão é apresentada na Tabela 6.6.

**Tabela 6.6:** Matriz de confusão após a aplicação da segunda rede neural

| Seq. | 1 | 2  | 3 | 4  | 5  | 6  | 7  |
|------|---|----|---|----|----|----|----|
| 1    | 6 | 5  | 0 | 1  | 0  | 2  | 0  |
| 2    | 0 | 13 | 1 | 0  | 0  | 0  | 0  |
| 3    | 0 | 5  | 8 | 1  | 0  | 0  | 0  |
| 4    | 0 | 0  | 3 | 11 | 0  | 0  | 0  |
| 5    | 0 | 0  | 0 | 0  | 14 | 0  | 0  |
| 6    | 0 | 0  | 0 | 0  | 0  | 14 | 0  |
| 7    | 1 | 0  | 0 | 0  | 0  | 0  | 14 |

Podemos notar que a segunda rede pouco contribuiu para a distinção entre os padrões 1, 2 e 3 devido a sua grande similaridade. O resultado final do experimento, considerando os 7 padrões, é mostrado na Tabela 6.7.

**Tabela 6.7:** Resultado final do experimento com a face inferior

|                                   |                |
|-----------------------------------|----------------|
| <b>Amostras de Teste</b>          | 98             |
| <b>Classificados Corretamente</b> | 80             |
| <b>Erros</b>                      | 18             |
| <b>Taxa de Reconhecimento</b>     | <b>81,63 %</b> |

















Donato *et. al.* em [2] obtiveram uma taxa de reconhecimento de 95,5 % na classificação de 6 AUs da parte superior e 6 AUs da parte inferior, usando 40 filtros de Gabor, imagens perfeitamente alinhadas e com diferença do quadro neutro para um mesmo indivíduo. As AUs classificadas eram bem distintas e a técnica de classificação utilizada foi a distância Euclidiana.

Zhang *et. al.* em [19] obtiveram uma taxa de reconhecimento de 90,1 % para reconhecer 6 expressões de emoção (alegria, tristeza, surpresa, irritação, desgosto e medo), além da expressão neutra, combinando 18 filtros de Gabor e o método de características geométricas. Os coeficientes foram extraídos de 34 pontos específicos selecionados manualmente e o banco de dados usado era composto somente por indivíduos do sexo feminino tipo oriental. Para a classificação foi utilizada uma rede neural com 612 neurônios de entrada, tanto para os filtros de Gabor como para o método de características geométricas.

No sistema proposto nesse trabalho, se combinássemos todas as AUs da parte superior e inferior teríamos um total de 49 expressões, o que vai muito além das 6 expressões propostas por Zhang [19]. A Tabela 6.8 mostra como exemplos, as 6 expressões de emoção (alegria, tristeza, surpresa, irritação, desgosto e medo) mais a neutra geradas a partir da combinação das ações faciais, classificadas neste trabalho, da parte superior e inferior da face.

**Tabela 6.8:** Expressões de emoção resultantes da combinação de AUs

| Face Superior  | Face Inferior  | Expressão de Emoção |
|--|--|---------------------|
|   |   | Neutra              |
|   |   | Espanto             |
|   |   | Alegria             |
|   |   | Irritação           |
|   |   | Desgosto            |
|   |   | Medo                |
|  |  | Tristeza            |

Além da possibilidade da geração de 49 distintas expressões, combinando as 7 ações da parte superior com as 7 ações da parte inferior, foi utilizado um banco de dados heterogêneo, com movimentos de cabeça e variação na luminosidade. Não foi necessário definir a face neutra para um mesmo indivíduo, uma vez que a imagem neutra média já foi previamente definida. Também, não foi realizado um rigoroso alinhamento e nem uma definição manual dos pontos específicos. Nós obtemos um reconhecimento de 86,55 % para a face superior e 81,63 % para a face inferior, distinguindo ações faciais muito similares entre si e com um reduzido custo computacional.

### 6.3 Desempenho em Tempo Real

O desempenho do sistema para operação em tempo real foi estimado para a configuração que obteve o melhor resultado, experimento 3 da Tabela 6.1, tanto para a parte superior, quanto para a inferior, o número de filtros de Gabor e a estrutura da rede neural são as mesmas, portanto, a estimativa é obtida para ambas as partes consideradas (inferior e superior). Utilizando um computador Pentium IV – 1.6GHz e 256MB de memória RAM e executando os seguintes passos:

- Re-escala para uma janela de dimensão  $30 \times 45$  pixels;
- Diferença com a AU0 média;
- Passagem dos 8 filtros de Gabor;
- Projeção nos 30 componentes principais;
- Operações da rede neural.

A Tabela 6.9 apresenta o tempo médio para a determinação da ação facial, em 1 quadro de uma seqüência, utilizando 2 tipos de linguagens.

**Tabela 6.9:** Desempenho do Sistema

| Linguagem | Tempo (ms) |
|-----------|------------|
| C++       | 8.00       |
| JAVA      | 10.05      |

A Tabela 6.9 apresenta o tempo de processamento para efetuar todas as operações é baixo. A aplicação do sistema proposto em seqüências de vídeo geradas a uma taxa de 30 quadros/s, normalmente empregado por câmeras digitais, é viável, visto que ainda restariam 25ms para a etapa de definição da janela de segmentação. Maiores

detalhes sobre o código do programa, utilizado nesse teste de desempenho, encontram-se no Apêndice C.

É importante salientar que a etapa de localização dos pontos na face e segmentação não foram levados em conta para estimar o tempo de processamento apresentado na Tabela 6.9, pois esta etapa foi realizada através do aplicativo mostrado do Apêndice B. O desenvolvimento de um sistema automático para localização da janela corte e segmentação estão fora do escopo deste trabalho.

# Capítulo 7

## Conclusões

Este trabalho apresenta o uso das wavelets de Gabor seguida da Análise de Componentes Principais associado a um passo de pré-processamento baseado na subtração da imagem com a face neutra média de toda a base de dados.

As imagens originais foram dezimadas através de um algoritmo simples, sem nenhuma aplicação prévia de um filtro anti-recobrimento, o que diminui os efeitos indesejáveis decorrentes do processo de dezimação, porém aumenta o tempo de processamento. As imagens de teste para a parte superior e inferior da face apresentam uma dimensão, aproximadamente, 17 vezes menor que as imagens segmentadas.

Comparando a utilização das wavelets de Gabor com o filtro de Sobel, foi possível notar que os filtros de Gabor obtiveram uma taxa de reconhecimento *significativamente maior, visto que o filtro de Sobel é menos seletivo do que o filtro de Gabor com  $k_v = \pi/2$*  e também pelo número limitado de orientações.

O uso do PCA se mostrou eficiente, foram escolhidos 30 componentes principais, o que possibilitou que fosse mantido um percentual de quantidade de informação estimada de 80% para cada filtro. A projeção das imagens nos espaços característicos resultou em um vetor de 240 coeficientes que serviu de entrada para a rede neural.

Para classificar 6 expressões de emoção mais a neutra Zhang *et. al.* [19] utilizaram uma rede neural com 612 coeficientes de entrada, tanto para o método de características geométricas quanto para o uso das wavelets de Gabor, sendo

posteriormente combinadas em uma saída comum. Para classificar a presença na face de 9 AUs, Tian *et. al.* [17] utilizaram uma rede com 495 neurônios de entrada, tanto para o método de características geométricas quanto para com wavelets de Gabor, que posteriormente foram combinadas em uma saída comum. Utilizamos uma rede neural com 240 neurônios de entrada para classificar 7 padrões de ações faciais da face superior e outra rede com as mesmas características para classificar padrões da face inferior.

Utilizamos imagens selecionadas de um banco de dados com indivíduos heterogêneos que incluíam movimentos involuntários da cabeça e mudança na luminosidade. Uma rede neural especializada proporcionou um aumento da taxa de reconhecimento para 86,55 % e 81,63 % para a parte superior e inferior, respectivamente, usando apenas 8 filtros de Gabor com frequência espacial  $k_v = \pi/2$  e 30 componentes principais para cada conjunto de imagens filtradas. O uso da diferença da imagem neutra média aumentou a robustez do método, permitindo que as imagens segmentadas e alinhadas de maneira aproximada fossem processadas com sucesso pela rede neural. O alinhamento e segmentação aproximados reduzem a complexidade no estágio de pré-processamento, uma vez que não é necessário estimar a imagem neutra daquele indivíduo, tornando a implementação em tempo real mais fácil do que as aproximações apresentadas na literatura atual. O tempo para se estimar uma ação facial em um computador Pentium IV – 1.6GHz e 256MB de memória RAM foi de 8 ms, restando ainda 25 ms, para definição da janela de segmentação, antes de se atingir o tempo médio de 33 ms entre quadros, utilizado em câmeras comerciais.

Para trabalhos futuros, sugerimos o uso da técnica aqui proposta para a classificação de um número maior de ações faciais, incluindo variação de intensidade, como também aplicar diferentes formas e estruturas da rede neural. Outra proposta é

---

conceber um sistema de Interação Homem-Computador amigável que integraria este método com outros que tratasse a fala ou gestos manuais.

## Referências

- [1] P. Eckman and E.L. Rosenberg, “What the Face Reveals: Basic and Applied Studies of Spontaneous Expression using the Facial Action Coding System (FACS)”, *Oxford Univ. Press*, New York, 1997.
- [2] G. Donato, M. S. Bartlett, J. C. Hager, P. Eckman, and T. J. Sejnowski, “Classifying Facial Action”, *IEEE Transaction on Pattern Analysis and Machine Intelligence*, 21 (10):974-989, October 1999.
- [3] P. Eckman and W. Friesen, “Facial Action Coding System: A Technique for the Measurement of Facial Movement”, *Consulting Psychologists Press*, Palo Alto, Calif., 1978.
- [4] Scherer, K., and Ekman, P., “Handbook of Methods in Nonverbal Behavior Research”, *Cambridge University Press*, Cambridge, UK, 1982.
- [5] K.D. Craig, S.A. Hyde and C.J. Patrick, “Genuine, Suppressed, and Faked Facial Behaviour During Exacerbation of Chronic Low Back Pain”, *Pain*, vol. 46, pp. 161-172, 1991.
- [6] M. Heller and V. Haynal, “The Faces of Suicidal Depression” (Translation Les Visages de la Depression de Suicide), *Kahiers Psychiatriques Genevois (Medicine et Hygiene Editors)*, vol. 16, pp. 107-117, 1994.
- [7] K. Mase, “Recognition of Facial Expression from Optical Flow”, *IEICE Transactions*, vol. E74, no. 10, pp. 3, 474-3, 483 1991.
- [8] Y. Yacoob and L. Davis, “Recognizing Human Facial Expression from Long Image Sequences Using Optical Flow”, *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 16, no. 6, pp. 636-642, June 1994.



- [9] M. Rosenblum, Y. Yacoob, and L. Davis, "Human Expression Recognition from Motion Using a Radial Basis Function Network Architecture", *IEEE Trans. Neural Networks*, vol. 7, no. 5, pp. 1, 121-1, 138, 1996.
- [10] I. Essa and A. Pentland, "Coding, Analysis, Interpretation, and Recognition of Facial Expressions", *IEEE Trans. Patterns Analysis and Machine Intelligence*, vol. 19, no. 7, pp. 757-763, July 1997.
- [11] A. Lanitis, C. Taylor, and T. Cootes, "Automatic Interpretation and Coding of Face Images Using Flexible Models", *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 19, no. 7, pp. 743-756, July 1997.
- [12] Z. Zhang, "Feature-Based Facial Expression Recognition: Sensitivity Analysis and Experiments with a Multi-Layer Perceptron", *International Journal of Pattern Recognition and Artificial Intelligence*, 13(6):893-911, 1999.
- [13] G. Cottrell and J. Metcalfe, "Face, Gender and Emotion Recognition Using Holons", *Advances in Neural Information Processing Systems*, D. Touretzky ed., vol.3, pp. 564-571, San Mateo, Calif., 1991.
- [14] C. Padgett and G. Cottrell, "Representing Face Images for Emotion Classification", *Advances in Neural Information Processing Systems*, M. Mozer, M. Jordan, and T. Petsche, eds., vol. 9, Cambridge, Mass.:MIT Press, 1997.
- [15] D. Terzopoulos and K. Waters, "Analysis and Synthesis of Facial Image Sequences Using Physical and Anatomical Models", *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 15, no. 6, pp. 569-579, 1993.
- [16] H. Li, P. Roivainen, and R. Forchheimer, "3-D Motion Estimation in Model-Based Facial Image Coding", *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 15, no. 6, pp. 545-555, 1993.
- [17] Y. Tian, T. Kanade, and J. Cohn, "Evaluation of Gabor-wavelet-based facial action unit recognition in image sequences of increasing complexity", *Proceedings of the Fifth*

*IEEE International Conference on Automatic Face and Gesture Recognition*, pp. 229 – 234, May, 2002.

[18] T. Kanade, J. Cohn, and Y. Tian. “Comprehensive database for facial expression analysis”, *Proceedings of International Conference on Face and Gesture Recognition*, pp. 46-53, March, 2000.

[19] Z. Zhang, M. Lyons, M. Schuster, and S. Akamatsu. “Comparison between geometry-based and gabor-wavelets-based facial expression recognition using multi-layer perceptron”, *International Workshop on Automatic Face and Gesture Recognition*, pp. 454-459, 1998.

[20] Beat Fasel, Juergen Luetttin, “Recognition of Asymmetric Facial Action Unit Activities and Intensities”, *Proceedings of International Conference on Pattern Recognition*, 5100-5103, Barcelona, Spain, 2000.

[21] M. Lyons, S. Akamatsu, M. Kamachi, and J. Gyoba, “Coding facial expressions with gabor wavelets”, *Proceedings of International Conference on Face and Gesture Recognition*, 1998.

[22] Ekman, P., Friesen, W. V., & Ellsworth, P., “Emotion in the human face: Guidelines for research and a review of findings”, *Pergamon Press*, New York, 1972.

[23] Frois-Wittmann, J., “The judgment of facial expression”, *Journal of Experimental Psychology*, pp. 113-151, 13, 1930.

[24] Ekman, P., Friesen, W. V. and Tomkins, S. S., “Facial affect scoring technique (FAST): A first validity study”, *Semiotica*, 3 (1), pp. 37-58, 1971.

[25] D. H. Hubel and T. N. Wiesel, “Functional architecture of macaque monkey visual cortex”, *Proc. Royal Soc. B (London)*, vol. 198, pp. 1-59, 1978.

[26] S. Marcelja, “Mathematical description of the responses of simple cortical cells,” *J. Optical Soc. Am.*, vol. 70, pp. 1297-1300, 1980.

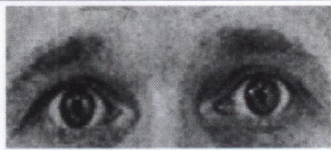





- [27] J. G. Daugman, "Two dimensional spectral analysis of cortical receptive field profile", *Vision Research*, vol. 20, pages 847-856, 1980.
- [28] D. Gabor, "Theory of communication," *J. IEE*, vol. 93, pages 429-459, 1946.
- [29] T. S. Lee, "Image representation using 2-d Gabor wavelets," *IEEE Trans. On Pattern Analysis Pattern Analysis and Machine Intelligence*, vol. 18, no.10, October, 1996.
- [30] B.S. Manjunath, R. Chellappa, and C. von der Malsburg, "A FeatureBased Approach to Face Recognition", *Proc. of International Conf. on Computer Vision*, 1992.
- [31] I.T. Jolli, "Principal Component Analysis", Springer-Verlag, 1986.
- [32] M. Kendall, "Multivariate Analysis" *Charles Gri\_n&Co.*, 1975.
- [33] Birk, Henrik and Thomas B. Moeslund, "Recognizing Gesture from the Hand Alphabet Using Principal Component Analysis", *Master's thesis*, Aalborg University, 1996.
- [34] J. R. Movellan, "Visual Speech Recognition with Stochastic Networks", *Advances in Neural Information Processing Systems*, G. Tesauro, D.S. Touretzky, and T. Leen, eds., vol. 7, pages 851-858, Cambridge, Mass.: MIT Press, 1995.
- [35] P. Werbos, "Backpropagation: Past and future", *Proceedings of the IEEE International Conference on Neural Networks*, pages 343-353. IEEE Press, 1988.
- [36] Documentation of NeuralWorks v5.5 software application. NeuralWare Company, Carnegie, USA.
- [37] R. Gonzalez and R. Woods, "*Digital Image Processing*", Addison Wesley, 1992, pp 414 - 428.

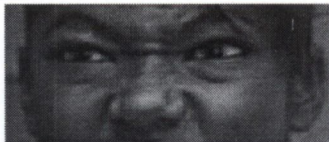





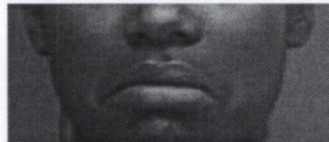




- 
- [38] A.M. Martinez, "Recognizing expression variant faces from a single sample image per class", *Computer Vision and Pattern Recognition*, June, 2003.









# Apêndice A



A seguir são apresentados as unidades de ação, uma breve descrição do movimento facial e os músculos estimulados.

**Tabela A.1:** Unidades de ação facial e os músculos que são estimulados

| AU | Descrição                              | Músculos Atuantes  | Imagem   |
|----|--|--|--|
| 1  | Parte interna da sobrancelha levantada | <i>Frontalis, Pars Medialis</i>                                      |    |
| 2  | Parte externa da sobrancelha levantada | <i>Frontalis, Pars Lateralis</i>                                     |   |
| 4  | Sobrancelha abaixada                   | <i>Depressor Glabellae;<br/>Depressor Supercilli;<br/>Corrugator</i> |  |
| 5  | Pálpebra superior levantada            | <i>Levator Palpebrae Superioris</i>                                  |  |
| 6  | Bochechas levantadas                   | <i>Orbicularis Oculi, Pars Orbitalis</i>                             |  |
| 7  | Pálpebras levemente apertadas          | <i>Orbicularis Oculi, Pars Palpebralis</i>                           |  |
| 8  | Lábios na direção de um a outro        | <i>Orbicularis Oris</i>  |  |

|    |  |   |  |
|----|--|---|--|
| 9  | Nariz ondulado                         | <i>Levator Labii Superioris,<br/>Alaeque Nasi</i>                 |    |
| 10 | Lábio superior levantado               | <i>Levator Labii Superioris,<br/>Caput Infraorbitalis</i>         |    |
| 11 | Nasolabial aprofundado                 | <i>Zygomatic Minor<br/>Deepener</i>                               |    |
| 12 | Borda dos lábio puxadas                | <i>Zygomatic Major</i>  |    |
| 13 | Bochechas cheias                       | <i>Caninus</i>  |    |
| 14 | Cavidades                              | <i>Buccinator</i>   |   |
| 15 | Borda dos lábios abaixada              | <i>Triangularis</i>   |  |
| 16 | Lábio inferior abaixado                | <i>Depressor Labii</i>  |  |
| 17 | Queixo abaixado                        | <i>Mentalis</i>   |  |
| 18 | Lábios enrugados<br>(em forma de bico) | <i>Incisivii Labii Superioris;<br/>Incisivii Labii Inferioris</i> |  |
| 20 | Lábio esticado                         | <i>Risorius</i>   |  |

|    |                             |  |  |
|----|-----------------------------|--|--|
| 22 | Lábio canalizado            | <i>Orbicularis Oris</i>  |    |
| 23 | Lábio apertado              | <i>Orbicularis Oris</i>  |    |
| 24 | Lábio pressionado           | <i>Orbicularis Oris</i>  |    |
| 25 | Lábios separados            | <i>Depressor Labii, ou Relaxamento do Mentalis ou Orbicularis Oris</i> |    |
| 26 | Mandíbula solta ou relaxada | <i>Masseter; Temporal e Interno Pterygoid</i>                          |    |
| 27 | Boca aberta                 | <i>Ptergoids; Digastric</i>  |   |
| 28 | Lábio sugado                | <i>Orbicularis Oris</i>  |  |
| 38 | Narina dilatada             | <i>Nasalis, Pars Alaris</i>  |  |
| 39 | Narina comprimida           | <i>Nasalis, Pars Transversa e Depressor Septi Nasi</i>                 |  |
| 41 | Pálpebra caída ou relaxada  | <i>Relaxamento do Levator Palpebrae Superioris</i>                     |  |
| 42 | Olhos rachados ou partidos  | <i>Orbicularis Oculi</i>   |  |

|    |  |  |  |
|----|--|--|--|
| 43 | Olhos fechados                                   | <i>Relaxamento do Levator Palpebrae Superioris</i>   |  |
| 44 | Olhar com os olhos parcialmente abertos          | <i>Orbicularis Oculi, Pars Palpebralis</i>   |  |
| 45 | Piscar os olhos                                  | <i>Relaxamento do Levator Palpebrae e Contração do Orbicularis oculi, Pars Palpebralis</i> |  |
| 46 | Piscar os olhos de forma disfarçada ou piscadela | <i>Orbicularis Oculi</i>   |  |

A Tabela A.2 mostra uma descrição dos movimentos faciais de outras AUs menos comuns. Nenhuma das AUs descritas nesta tabela está presentes no Banco de Dados Cohn-Kanade.

**Tabela A.2:** Descrição de AUs pouco comuns

| AU | Descrição            |
|----|----------------------|
| 19 | Língua para fora     |
| 21 | Nuca apertada        |
| 29 | Maxilar empurrado    |
| 30 | Maxilar para o lado  |
| 31 | Maxilar apertado     |
| 32 | Mordida no lábio     |
| 33 | Bochecha inflada     |
| 34 | Bochecha de assopro  |
| 35 | Bochecha sugada      |
| 36 | Língua estufada      |
| 37 | Lábios sendo secados |



## Apêndice B

A Figura B.1 mostra a tela do aplicativo desenvolvido para realizar a segmentação de imagens.



**Figura B.1:** Aplicativo utilizado para a segmentação de imagens

O operador deste aplicativo deve carregar todos os quadros da seqüência. O primeiro quadro é selecionado pelo aplicativo automaticamente e o operador deve ativar os pontos descritos na Tabela B.1.

**Tabela B.1:** Localização dos pontos na imagem a ser cortada

| Ponto | Localização                                   |
|-------|---|
| 1     | Na parte mais alta de alguma das sobrancelhas |
| 2     | Na parte interna do olho direito              |
| 3     | Na parte interna do olho esquerdo             |
| 4     | Na parte mais alta de alguma narina           |
| 5     | Na parte mais baixa do queixo                 |

Ao clicar nos pontos descritos, o aplicativo gerará a janela de corte (retângulos na Figura B.1) para a parte superior e inferior da face com base nos pontos citados na Tabela B.1. A expressão do retângulo superior é a seguinte:

$$l_s = \frac{7}{5} d_{x23} \quad (C.1)$$

$$h_s = d_{y14} + 5 \quad (C.2)$$

onde:

$l_s$  é a largura do retângulo superior;

$h_s$  é a altura do retângulo superior;

$d_{xij}$  : distância entre os componentes horizontais do ponto  $i$  ao ponto  $j$

$d_{yij}$  : distância entre os componentes verticais do ponto  $i$  ao ponto  $j$

O retângulo inferior é gerado a partir da seguinte expressão:


$$l_i = l_s \quad (C.3)$$


$$h_i = d_{y45} + 5 \quad (C.4)$$

onde:

$l_i$ : largura do retângulo inferior;

$h_i$ : altura do retângulo inferior.

Ao definir os pontos, o operador deve ativar o botão , que seguirá para o próximo quadro. Automaticamente é feito o rastreamento na vertical e na horizontal (se a opção “*Horizontal Tracking*” estiver selecionada) dos pontos definidos no primeiro quadro da seqüência. O operador deve verificar se os pontos estão nos devidos lugares (ver Tabela B.1), caso contrário deve ativar a caixa “*Change Mode*”, selecionar o ponto desejado e selecioná-lo novamente na localização correta.

Ao final do último quadro da seqüência, o operador deve ativar o botão  que salvará as imagens cortadas em face superior e inferior em um diretório definido previamente. Paralelamente será criado um arquivo que constará o nome das imagens e a localização dos cinco pontos.

# Apêndice C

A seguir é apresentado o código, em C++, que simula a execução dos passos de pré-processamento.

```
#include <stdio.h>
#include <stdlib.h>
#include <math.h>
#include <time.h>

void main() {

clock_t start, stop;
int i=0, j=0, k=0, z=0, w=0, t=0;
float a = 0;
float b = 0;
double var = 0.0;

int Orig[150][225]; //supondo img. original com este tamanho
int Media[30][45]; //AU0 média
int M[30][45]; //nova imagem

start = clock();
//executado 1000 vezes p/ evitar prob. c/ precisao.
for (t=0; t<1000;t++) {
//////////////////// RE-ESCALA e Dif AU0 Media //////////////////////////////////////
//re-escala: método do vizinho mais próximo
for (i=0; i<30; i++) {
    for (j=0; j<45; j++) {
        M[i][j] = Orig[(i-1)*3][(j-1)*3];
        M[i][j] -= Media[i][j]; //aproveita o mesmo loop
                                //para fazer a dif AU0
    }
}
//////////////////////////////////// Filtros de Gabor e PCA //////////////////////////////////////
// tamFiltro = 10x10
// N_Filtros = 8

for (k=0; k<8; k++) {
    for (i=0; i<30; i++) {
        for(j=0; j<45; j++) {
            for (z=0; z<100; z++) {
                // cada ponto da imagem terá 100 multi-
                // plicacoes e 99 somas
                M[i][j]*=M[i][j]; //simulacao das mult.
                M[i][j]+=M[i][j]; //simulacao das somas+1
            }
            // a seguir será aproveitado o mesmo loop para
            // as projeções PCA.
            // Cada ponto sera multiplicado pelas 30 C.P.
            for (w=0; w<30; w++) {
                M[i][j]*=M[i][j];
            }
        }
    }
}
```

```
        M[i][j]+=M[i][j]; //somando os coeficientes
    }
}
}
/////////////////////////////////////////////////////////////////
///////////////////////////////////////////////////////////////// REDE NEURAL ///////////////////////////////////////////////////////////////////
//Nin = 240
//Nhide = 160
//Nout = 7
for (i=0; i<240; i++) {
    for (j=0; j<160; j++) {
        a*=b; //multiplicando os pesos
        a+=b; //somando (LBF)
    }
    var = 1/(1+exp(-a)); //funcao sigmoid (simulacao)
}
for (i=0; i<160; i++) {
    for (j=0; j<7; j++) {
        a*=b;
        b+=b;
    }
    var = 1/(1+exp(-a));
}
/////////////////////////////////////////////////////////////////
}
}
stop = clock();
float resul = (float) ((stop-start)/CLOCKS_PER_SEC);
printf("Tempo: %f",resul);
}
```