Universidade Federal da Bahia
Escola Politécnica / Instituto de Matemática
Programa de Pós-Graduação em Mecatrônica

CAROLINE PACHECO DO ESPÍRITO SILVA

RECONHECIMENTO DE EXPRESSÕES FACIAIS UTILIZANDO REDES NEURAIS ARTIFICIAIS

Salvador
2012

CAROLINE PACHECO DO ESPÍRITO SILVA

RECONHECIMENTO DE EXPRESSÕES FACIAIS UTILIZANDO REDES NEURAIS ARTIFICIAIS

Dissertação apresentada ao Programa de Pós-Graduação em Mecatrônica da Universidade Federal da Bahia como requisito para obtenção do grau de Mestre em Mecatrônica.

Orientador: Dr. Leizer Schnitman

Co-orientador: Dr. Luciano Rebouças de Oliveira

Salvador
2012

S586 Silva, Caroline

Reconhecimento de expressões faciais utilizando redes neurais artificiais / Caroline Silva. – Salvador, 2012.

100 f. : il. color.

Orientador: Leizer Schnitman
Coorientador: Luciano Oliveira

Dissertação (mestrado) – Universidade Federal da Bahia. Escola Politécnica, 2012.

1. Expressão facial. 2. Redes neurais (computação). 3. Inteligência artificial. I. Schnitman, Leizer. II. Oliveira, Luciano. III. Universidade Federal da Bahia. IV. Título.

CDD: 006.3

TERMO DE APROVAÇÃO

CAROLINE PACHECO DO ESPÍRITO SILVA

RECONHECIMENTO DE EXPRESSÕES FACIAIS UTILIZANDO REDES NEURAIS ARTIFICIAIS

Dissertação aprovada como requisito parcial para obtenção do grau de Mestre em Mecatrônica, Universidade Federal da Bahia, pela seguinte banca examinadora:

Leizer Schnitman - Orientador _____
Doutor em Engenharia Eletrônica e Computação, Instituto Tecnológico de Aeronáutica (ITA)
Universidade Federal da Bahia

Luciano Rebouças de Oliveira – Co-orientador _____
Doutor em Engenharia Elétrica e de Computadores, Instituto de Sistemas e Robótica da Universidade de Coimbra, Portugal
Universidade Federal da Bahia

Ângelo Amâncio Duarte – Membro interno _____
Doutor em Ciência da Computação, Universitat Autònoma de Barcelona, Espanha
Universidade Federal da Bahia

Roberto Kawakami Harrop Galvão – Membro externo _____
Doutor em Engenharia Eletrônica e Computação, Instituto Tecnológico de Aeronáutica (ITA)
Instituto Tecnológico de Aeronáutica

Salvador, 18 de dezembro de 2012

AGRADECIMENTOS

Gostaria de aproveitar esta oportunidade para expressar minha gradidão ao meu orientador Leizer Schnitman, pelo apoio que foi muito importante para a conclusão desse trabalho e também ao meu co-orientador Luciano Oliveira que passou os conhecimentos necessários para a inicialização desse projeto. Agradeço aos meus pais Eduardo e Crispina e aos demais familiares e amigos que de maneira indireta me apoiaram para a realização deste sonho. Agradeço também a Andrews Sobral que sempre esteve ao meu lado e foi meu companheiro durante todo esse tempo. Ao meu sogro Ruivaldo Sobral pelas palavras e pela ajuda que foram muito importantes nesta caminhada. A minha cunhada Priscilla, e em especial ao professor Vitor Leão Filardi pela confiança no período inicial do mestrado. Finalmente, agradeço a todos aqueles que de alguma maneira torceram por mim.

"O conhecimento é adquirido e pode ser passado adiante e compartilhado por meio de palavras e outros símbolos. A compreensão é uma experiência imediata e só pode ser comentada (muito insatisfatoriamente) nunca compartilhada."

—ALDOUS HUXLEY

RESUMO

A análise automática de expressões faciais tem atraído cada vez mais a atenção de pesquisadores em diversas áreas como psicologia, ciência da computação, linguística, neurociência e áreas relacionadas. Nas últimas décadas, pesquisadores têm realizado muitos trabalhos e inúmeras abordagens promissoras para o reconhecimento automático de expressões faciais têm surgido. Este crescente interesse surgiu através do desenvolvimento de novos métodos de processamento de imagens, novas abordagens para detecção e reconhecimento facial, bem como o aumento da capacidade computacional. Nesta dissertação é proposto um sistema de reconhecimento automático de expressões faciais. O sistema proposto classifica sete diferentes expressões: felicidade, raiva, tristeza, surpresa, desgosto, medo e neutra. Utilizou-se as bases de dados MUG Facial *Expression* e *Face and Gesture Recognition Research Network* (FG-NET). Estas bases apresentam imagens com plano de fundo uniforme e não uniforme. As bases de dados também contém imagens de indivíduos que apresentam diferenças individuais tais como: barba, bigode e óculos. Os resultados experimentais demonstram que o sistema proposto baseado em redes neurais artificiais alcança uma taxa média de acerto de 97,62% para as sete diferentes expressões faciais definidas.

Palavras-chave: Classificação de Expressões Faciais, Redes Neurais Artificiais

ABSTRACT

The automatic analysis of facial expressions has drawn attention from researchers in different fields of study such as psychology, computer science, linguistics, neuroscience and related fields. In the last decades several researchers have released many studies and introduced a large amount of approaches and methods for human-face detection, recognition and analysis. The advances on image processing, computer vision and computing power also contributed to this success. This work proposes a system for automatic human-face expression recognition. The proposed system classifies seven different facial expressions: happiness, anger, sadness, surprise, disgust, fear and neutral. The proposed system was evaluated with two facial expression databases: MUG Facial Expression e Face and Gesture Recognition Research Network (FG-NET). These databases contain images with uniform and non-uniform background. The databases also contain images of people who have individual differences such as: beard, mustache and glasses. The experimental results demonstrate that the proposed system shows 97.62% of accuracy for the seven defined facial expressions using artificial neural networks.

Keywords: Facial Expression Classification, Artificial Neural Networks

LISTA DE FIGURAS

LISTA DE TABELAS

LISTA DE ABREVIATURAS

AAM	Active Appearance Model - Modelo Ativo de Aparência
AFA	Automatic Face Analysis - Análise Automática da Face
ANN	Artificial Neural Network - Rede Neural Artificial
ASM	Active Shape Model - Modelos de Forma Activa
A-SVMs	Adaptive Support Vector Machines - Máquinas de Vetores de Suporte Adaptativa
AUs	Action Units - Unidades de Ação
FG-NET	Face and Gesture Recognition Research Network
FACS	Facial Action Coding System - Sistema de Codificação de Animação Facial
FER	Facial Expression Recognition - Reconhecimento de Expressões Faciais
GPA	Generalized Procrustes Analysis - Análise Generalizada de Procrustes
HMM	Hidden Markov Model - Modelo Oculto de Markov
HSV	Hue, Saturation and Value - Matiz, Saturação e Valor
ICA	Independent Component Analysis - Análise de Componentes Independentes
KNN	K-Nearest Neighbor - K-Vizinho-mais-Próximo
LBP	Local Binary Pattern - Padrões Binários Locais
LBP-TOP	Local Binary Patterns from Three Orthogonal Planes
LDA	Linear Discriminant Analysis - Análise Discriminante Linear
LFA	Local Feature Analysis
LPQ	Local Phase Quantization - Quantização de Fase Local
LPQ-TOP	Local Phase Quantisation from Three Orthogonal Planes

MLR	Multinomial Logistic Ridge Regression - Regressão Logística Multinomial
MPL	Multi-Layer Perceptrons - Percepttron Multicamadas
PCA	Principal Component Analysis - Análise de Componentes Principais
PDM	Point Distribution Model - Modelo Pontual de Distribuição
SVM	Support Vector Machine - Máquinas de Vetor de Suporte

SUMÁRIO

CAPÍTULO 1

INTRODUÇÃO

Nos últimos anos, sistemas automáticos para a análise de expressões faciais tornaram-se cada vez mais importantes na comunidade de pesquisa de visão computacional. Em muitos casos, observa-se que as expressões faciais descrevem os estados emocionais de uma pessoa, com maior precisão do que as palavras. O ser humano possui essa capacidade de interação com seu semelhante independente da utilização de uma linguagem em comum através da expressão. Consequentemente, sistemas automáticos de reconhecimento de expressões faciais podem trazer melhorias no que diz respeito à interação homem-máquina, de modo que, por exemplo, máquinas possam perceber o comportamento humano e reagirem proativamente. Outras possíveis aplicações encontram-se na área de segurança, realidade aumentada, diagnóstico médico, sistema de percepção, entre outras (SHAN, 2008; HESSE, 2011).

O naturalista britânico Charles Darwin foi o precursor em relacionar as expressões faciais com o ser humano, evidenciando a importância da face como um indicativo social para a garantia da sobrevivência do homem no processo de seleção natural. As ideias inovadoras de Darwin (1872) despertaram o interesse de muitos pesquisadores que começaram a desenvolver pesquisas neste campo. Um dos psicólogos que se destaca é Paul Ekman, autor do livro "*A Linguagem das Emoções*" e seu colega Friesen, que em 1971, aprofundando os estudos de Darwin, concluíram que existem emoções básicas universais e transculturais, ou seja, que não precisam ser ensinadas e que não variam entre grupos sociais. Essas expressões básicas são: felicidade, raiva, tristeza, surpresa, desgosto e medo; que são inatas na natureza humana e, independem da cultura, do país, da raça ou dos aprendizados adquiridos.

A fim de padronizar o processo de reconhecimento das expressões faciais, foram desenvolvidos diversos sistemas de codificação da ação facial. Dentre estes se destaca o

Facial Action Coding System (FACS) criado por Ekman e Friesen em 1978 que descreve as expressões faciais em unidades de ação (*Action Units* - AU's). Das 44 AU's definidas, 30 são anatomicamente relacionadas com a contração de músculos faciais específicos (12 AU's para a parte superior da face, e 18 AU's para a parte inferior da face), que podem corresponder a um músculo específico ou a um grupo muscular. O FACS vem sendo bastante utilizado como uma ferramenta descritiva e tem sido empregado em vários trabalhos científicos relacionados a micro-expressões.

Segundo Fasel e Luettin (2003), uma das primeiras investigações sobre reconhecimento automático de expressões faciais foi apresentada por Suwa et al. (1978). Os autores analisaram as expressões faciais, rastreando o movimento de 20 pontos identificados em uma sequência de imagens. A partir deste trabalho, muitos progressos têm sido realizados, devido ao desenvolvimento de novos métodos de processamento de imagens, novas abordagens para detecção e reconhecimento facial, bem como o aumento da capacidade computacional.

Apesar dos avanços obtidos em análise automática da expressão facial desde o primeiro trabalho em 1978, muitas questões permanecem sem solução. Segundo Tian et al. (2005) e Zhan et al. (2006a), o movimento da cabeça, baixa resolução das imagens, oclusão, variação de iluminação e as diferenças individuais estão entre os fatores que dificultam uma correta análise da expressão facial. Devido a isto, muitas pesquisas continuam sendo realizadas para solucionar tais problemas visando tornar os sistemas de reconhecimento automático de expressões cada vez mais robustos.

O presente trabalho propõe o desenvolvimento de um sistema automático de reconhecimento de sete expressões faciais que classificam as expressões universais definidas por Ekman e Friesen (1971): felicidade, raiva, tristeza, surpresa, desgosto, medo; incluindo a expressão neutra. As aplicações do sistema proposto são vastas, podendo ser útil em áreas como interação homem-máquina, jogos e robótica móvel. O sistema proposto foi treinado e avaliado através de dois métodos de classificação (baseados em correspondência entre modelos e em redes neurais artificiais) utilizando as bases de dados MUG *Facial Expression* (AIFANTI et al., 2010) e *Face and Gesture Recognition Research Network* (FG-

NET) (WALLHOFF, 2006) em que as imagens apresentam plano de fundo não uniforme e neutro e em imagens em que os indivíduos apresentam diferenças individuais ou artefatos: barba, bigode e óculos. Além disso, as imagens utilizadas pelo sistema estão restritas a ambientes fechados.

1.1 OBJETIVO

Este trabalho apresenta um sistema que se propõe ao reconhecimento automático de expressões faciais. O objetivo é classificar sete diferentes estados emocionais: felicidade, raiva, tristeza, surpresa, desgosto, medo e neutra utilizando as abordagens baseadas em correspondência entre modelos e em redes neurais artificiais.

1.1.1 Objetivos Específicos

São objetivos específicos desta dissertação:

- Apresentar alguns dos principais sistemas de reconhecimento de expressões faciais disponíveis no estado da arte.
- Realizar a extração de um conjunto de características descritivas da face.
- Avaliar possíveis abordagens de como classificar as características extraídas para o reconhecimento de expressões faciais.
- Comparar os resultados obtidos pelas diferentes abordagens utilizadas.

1.2 PUBLICAÇÃO

O presente trabalho foi aceito para publicação no XIX Congresso Brasileiro de Automática (CBA) sob o título "Detecção de Landmarks em Imagens Faciais Baseada em Informações Locais"(SILVA et al., 2012).

1.3 ORGANIZAÇÃO DO TEXTO

Esta dissertação está organizada de maneira que, no Capítulo 2, são introduzidas metodologias e abordagens para o reconhecimento de expressão facial existentes na literatura com ênfase nos métodos de extração de características e de classificação. O sistema proposto é apresentado no Capítulo 3 e os resultados experimentais são descritos o Capítulo 4. Por fim, o Capítulo 5 provê a conclusão desta dissertação e algumas perspectivas para trabalhos futuros.

CAPÍTULO 2

REVISÃO DA LITERATURA

A expressão facial é uma manifestação visível do estado emocional, atividade cognitiva, intenção e personalidade de um determinado indivíduo. Mehrabian (1968), pioneiro na pesquisa de linguagem corporal, relatou que em toda comunicação interpessoal cerca de 7% da mensagem é verbal (somente palavras), 38% é vocal (incluindo tom de voz, inflexão e outros sons) e 55% é não verbal. Dentre as diversas formas de comunicação não verbal, pode-se destacar as expressões faciais, as quais desempenham um papel fundamental nas relações interpessoais, revelando características da pessoa ou mensagem sobre algo a expressar, se caracterizando como um importante canal na comunicação.

Nas últimas décadas muitas pesquisas têm sido realizadas para desenvolver novos ou melhorar os atuais sistemas automáticos de reconhecimento de expressões faciais (*Facial Expression Recognition - FER*), os quais tem se potencializado devido a avanços nas áreas de visão computacional, aprendizagem de maquina e processamento de imagem. Muitos sistemas de FER foram desenvolvidos e são similares no sentido de que a maioria inicialmente extrai características faciais, e em seguida as utilizam para classificar diferentes categorias de expressões. O diferencial destes sistemas encontra-se no tipo de informação discriminativa extraída da imagem e pelo procedimento de classificação utilizado. Sendo assim, neste capítulo a ênfase é dada aos principais métodos de extração de características e de classificação utilizados no reconhecimento de expressões faciais, incluindo recentes estudos sobre sistemas de FER.

2.1 EXTRAÇÃO DE CARACTERÍSTICAS FACIAIS

Para que a análise de expressão facial seja bem sucedida, uma das etapas essenciais é extrair informações da expressão a ser reconhecida. Durante a fase de extração, os dados

do *pixel* das imagens são convertidos em uma representação de alto nível, chamada de vetores de características, que em seguida são utilizados para classificação das expressões. Normalmente, extrair um conjunto de características não é uma tarefa trivial, pois estas devem ser descritivas e, preferencialmente, não correlacionadas. Segundo Tian et al. (2005), existem usualmente dois tipos de abordagens para a extração de características faciais: as baseadas em geometria e as baseadas em aparência. Entretanto, alguns trabalhos utilizam características híbridas, buscando se beneficiar de uma combinação entre as características baseadas na geometria com outras baseadas na aparência

As características baseadas em geometria normalmente são representadas por formas das regiões faciais (sobrancelhas, olhos, boca, etc.) ou relações geométricas (distância, ângulos, etc.) entre *landmarks*[1], podendo ser extraídas manualmente ou automaticamente a partir da imagem de entrada. Já as características baseadas em aparência geralmente representam as variações na aparência (textura da pele) da face, tais como rugas, sulcos, etc. (SOBOTTKA; PITAS, 1997; CHUANG; SHIH, 2006). Nas próximas subseções, são apresentados estudos recentes sobre sistemas FER, categorizando-os a partir do tipo de característica utilizada para classificação.

2.1.1 Características Baseadas em Geometria

Um dos algoritmos encontrado no estado da arte para detectar *landmarks* faciais é o *Active Shape Model* (ASM) proposto inicialmente por Cootes et al. (1995). Este algoritmo é comumente utilizado para extrair características da face (TORRE; COHN, 2011). ASMs consistem em modelos estatísticos que permitem estimar os parâmetros de posição, escala e forma de um objeto numa imagem. O ASM tem sido amplamente utilizado para rastreamento de deformação facial, porém, pode falhar quando ocorrem transformações significativas na expressão. Para resolver este problema, Chang et al. (2006) inicialmente definem um modelo da face composto por 58 *landmarks* faciais (conforme mostrado na Figura 2.1), e em seguida utilizam modelos específicos de ASM para cada *cluster* (con-

[1]De acordo com Ross (2005), *landmarks* podem ser definidos como um conjunto finito de pontos que descrevem a forma de um determinado objeto de maneira exata e correta.

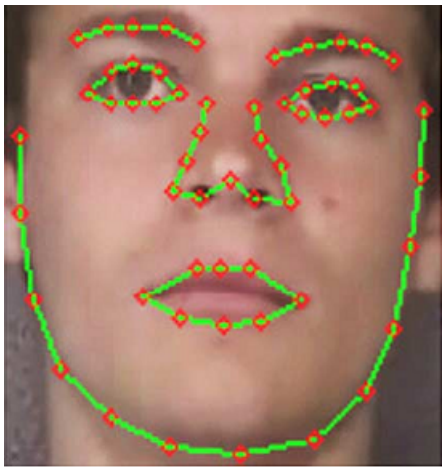


Figura 2.1: Modelo da face definidos por 58 *landmarks*. Fonte: (CHANG et al., 2006)

juntos de características para cada expressão). Um *cluster* consiste em um grupo de objetos semelhantes (ou relacionados) entre si e distintos de (ou não relacionados aos) outros objetos de outros grupos. No método, uma seleção *online* do modelo é realizada probabilisticamente de forma cooperativa com a classificação da expressão, melhorando a confiabilidade do rastreamento, ou seja, do acompanhamento da deformação facial obtida no inicio da sequência de imagens até o final da mesma.

Em Huang e Huang (1997), a face foi representada utilizando o *Point Distribution Model* (PDM), o qual foi criado a partir de um conjunto de *landmarks* extraídos de 90 imagens faciais de 15 indivíduos apresentando seis diferentes expressões. Nesse trabalho os autores utilizaram um detector de bordas para estimar a localização da face na imagem. A análise do valor da intensidade dos *pixels* entre os lábios e duas extremidades verticais simétricas representou os limites verticais exteriores da face, gerando uma estimativa da sua localização. No entanto, o método apresenta algumas limitações; o individuo não pode apresentar pêlos faciais nem óculos, não podem existir variações de luminosidade nem movimentos bruscos na posição da cabeça. A Figura 2.2a mostra o modelo gerado utilizando o PDM e a Figura 2.2b ilustra o ajuste do modelo PDM à face.

Segundo Torre e Cohn (2011), além de sistemas de FER que utilizam formas das regiões faciais, como anteriormente descritos, outros tipos de características amplamente utilizadas para compor o vetor de características baseado em geometria são: x_1^U distância

(a) **(b)**

Figura 2.2: (a) Modelo PDM (b) Ajuste do modelo à face. Fonte: (HUANG; HUANG, 1997)

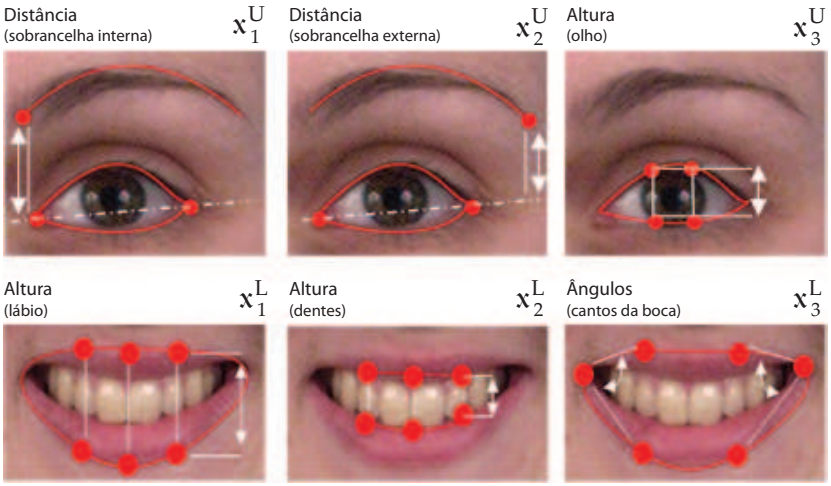


Figura 2.3: Seis diferentes características extraídas a partir de distâncias entre *landmarks*, altura das regiões faciais e ângulos entre os cantos da boca. Fonte: Adaptado de (TORRE; COHN, 2011).

entre a sobrancelha interna e o olho, x_2^U distância entre a sobrancelha externa e o olho, x_3^U altura do olho, x_1^L altura do lábio, x_2^L altura dos dentes e x_3^L o ângulo entre os cantos da boca, conforme ilustra a Figura 2.3.

No sistema proposto por Khandait et al. (2012) são utilizadas técnica de pré-processamento de imagens e segmentação para extrair informações das sobrancelhas, olhos, nariz e boca. Na extração foram obtidos 15 parâmetros para compor o vetor de característica, tais como: largura e altura das sobrancelhas, olhos, boca, e distâncias entre o centro do olho e da sobrancelha, etc. Na classificação o sistema apresentou uma taxa de acerto de 100% para o conjunto de dados treinados e precisão de 95,26% para o

(a) **(b)**

Figura 2.4: (a) Distâncias entre características (b) Imagem normalizada à esquerda e o mapa das bordas das caraceterísticas à direita. Fonte: (TIAN et al., 2003)

conjunto de teste. Em Sako e Smith (1996), os autores desenvolveram um sistema que classifica cinco diferentes expressões faciais. Basicamente, na etapa de extração de características, foram calculadas a largura e altura da boca e da face, e também a distância entre os olhos e as sobrancelhas para compor o vetor de características. Na etapa de classificação, os autores utilizaram o classificador *K-Nearest Neighbors (KNN)*, alcançando uma taxa de acerto de 71%.

Alguns sistemas de reconhecimento de expressões utilizam a combinação entre formas das regiões faciais e relações de distâncias e ângulos para compor o vetor de características. Tian et al. (2003) selecionam distâncias e características da forma ao longo das regiões dos olhos, das sobrancelhas e da boca, com intuito de extrair distâncias entre a linha dos olhos e das sobrancelhas, largura da boca, além da forma da boca que no trabalho deles é calculado a partir de histogramas e bordas. Assim um vetor contendo 17 características foi empregado como entrada de uma rede neural artificial para reconhecer as expressões: felicidade, neutra, raiva, surpresa e outras (incluindo o medo, a tristeza, e o desgosto). A Figura 2.4a ilustra a localização das características, onde L1, L2, L3, L4, L5 correspondem a distâncias entre seis características faciais e a Figura 2.4b mostra a face normalizada para o tamanho da face canônica com base na separação dos olhos à esquerda e o mapa das bordas das características à direita.

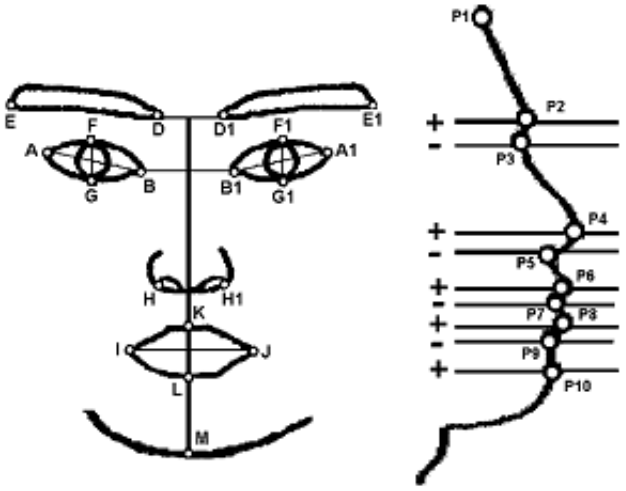


Figura 2.5: Modelos da face frontal e de perfil. Fonte: (PANTIC et al., 2000)

Pantic e Rothkrantz (2000) utilizaram modelos baseados em *landmarks* para representar duas visualizações diferentes da face: visão frontal e em perfil, com 30 e 10 localizações de *landmarks*, respectivamente. O modelo da posição frontal foi formado por 19 *landmarks* faciais, e as características restantes representaram quatro formas específicas da boca e uma do queixo. Na visão de perfil, os 10 *landmarks*, foram colocados de forma estratégica na curvatura que define o contorno lateral da face. Os dois modelos foram combinados para descrever 29 AU's diferentes. Os autores conseguiram uma taxa média de reconhecimento de 92% para AU's da face superior e 86% para AU's da face inferior. Segundo os autores 2% das imagens apresentaram falhas de reconhecimento. Pantic e Rothkrantz (2004) também utilizaram modelos da face com visualizações frontal e de perfil, porém com 19 e 10 *landmarks*, respectivamente (Figura 2.5). Entretanto, a praticidade da abordagem é limitada pela necessidade da câmera ser fixada sobre a cabeça do indivíduo, restringindo a liberdade de movimento, além de ser comumente desconfortável.

Além dos trabalhos relacionados nesta subseção, outros inúmeros trabalhos são encontrados na literatura que utilizam características geométricas para reconhecer expressões faciais. Entre estes, pode-se citar os Edwards et al. (1998), Wang et al. (1998), Cohn et al. (1999a), Tsapatsoulis et al. (2000), Bourel et al. (2002), Pardas et al. (2002), Vukadinovic e Pantic (2005), Kotsia et al. (2007), Seyedarabi et al. (2007), Hupont et al.

(2008a), Khanum et al. (2009), Rao et al. (2011), Pantic e Rothkrantz (2004), Pantic e Patras (2006), Valstar e Pantic (2006).

2.1.2 Características Baseadas em Aparência

Em oposição às características geométricas, as características de aparência geralmente codificam mudanças na aparência (textura da pele) da face, tais como rugas, saliências e sulcos. *Gabor wavelets* (DAUGMAN, 2003) é uma das técnicas bastante utilizadas para extrair mudanças na aparência como um conjunto de coeficientes de multi-escala e multi-orientação. Nos trabalhos de Ford (2002), Donato et al. (1999), Bartlett et al. (2001), foram extraídos coeficientes de *Gabor* em toda a imagem da face. Similarmente em Lyons et al. (1998a), Zhang et al. (1998), Tian et al. (2000), os filtros de *Gabor* são aplicados em regiões específicas da face.

O filtro de *Gabor* também pode ser aplicado em *landmarks* faciais, como acontece em Guo e Dyer (2005), trabalho no qual foi extraído um vetor de dimensão 612 (=3x6x34), sendo 3 escalas, 6 direções utilizadas pela filtragem de *Gabor* e 34 *landmarks* faciais. Este vetor foi empregado em diferentes classificadores, e os autores constataram que ao utilizar o classificador bayesiano o sistema apresentou uma taxa de reconhecimento de 63,3%. A Figura 2.6 mostra um exemplo de extração de características utilizando *Gabor wavelets*: no lado esquerdo são exibidas duas diferentes expressões faciais com suas correspondentes representações *Gabor* no lado direito (FASEL; LUETTIN, 2003).

Métodos como *Principal Component Analysis* (PCA)(TURK; PENTLAND, 1991), *Local Feature Analysis* (LFA)(DONATO et al., 1999), *Independent Component Analysis* (ICA)(DONATO et al., 1999), *Linear Discriminant Analysis* (LDA)(BELHUMEUR et al., 1997) também têm sido amplamente utilizados. Donato et al. (1999) compararam a utilização das técnicas citadas anteriormente, com o *Gabor wavelets* e *local principal components* (KAMBHATLA; LEEN, 1997). Segundo os autores os métodos que apresentaram melhores resultados, quando utilizado para reconhecer expressões faciais, foram os *Gabor wavelets* e *Independent Component Analysis*. No trabalho de Samad e Sawada

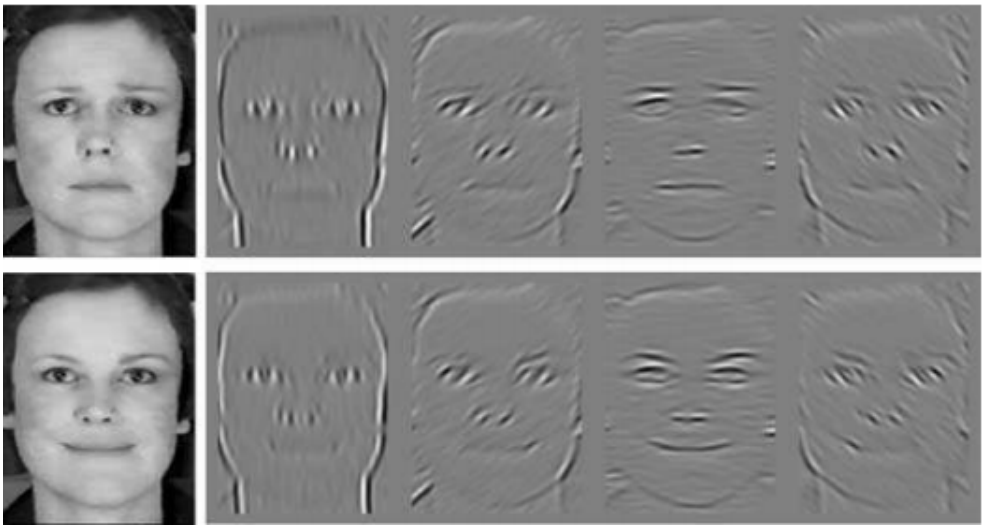


Figura 2.6: Extração de características utilizando *Gabor wavelets*. Fonte: (FASEL; LUETTIN, 2003)

(2011), foram extraídas características utilizando o filtro de *Gabor*, em seguida o *Principal Component Analysis* (PCA) foi utilizado para diminuir a quantidade de características extraídas pelo filtro de *Gabor*, e por fim, o classificador *Support Vector Machines* foi empregado para reconhecer as expressões. Ao utilizar a base de dados FG-NET (WALLHOFF, 2006), o sistema desenvolvido por eles, alcançou uma taxa de acerto de 81,7%.

Além dos métodos citados anteriormente, outras abordagens são utilizadas para extração de características baseadas em aparência. Em um estudo comparativo, Littlewort et al. (2002) analisam técnicas existente na literatura e classificam seis diferentes tipos de expressões faciais usando valores de intensidade de *pixel* e *Support Vector Machines (SVMs)*. Os autores obtiveram uma precisão em torno de 73%, quando os *pixels* foram extraídos a partir de toda a face. Apesar da precisão apresentada, os autores argumentaram que as características de intensidade dos *pixels* são importantes devido a simplicidade de extração.

Whitehill e Omlin (2006) utilizaram características do tipo *Haar* e o classificador *AdaBoost* (FREUND; SCHAPIRE, 1995) para reconhecer AU. Basicamente características do tipo *Haar* são características no formato retangular baseadas no estudo realizado previamente por Papageorgiou e Poggio (2000), no qual é proposto o uso de *Haar Wavelets* (MALLAT, 1989) para facilitar o treinamento de classificadores. No trabalho desenvolvido

Figura 2.7: Haar Wavelets. Fonte: (WHITEHILL; OMLIN, 2006)

pelos autores a velocidade e a precisão do sistema desenvolvido são comparados com o método de *Gabor* para extrair características e *Support Vector Machines* para classificar as expressões. Os resultados demonstraram que as técnicas *Haar* e *AdaBoost* apresentam resultados equivalentes ao *Gabor* + SVM em relação à precisão. Entretanto, a combinação de *Haar* e *AdaBoost* é duas vezes mais rápida que o método *Gabor* + SVM na velocidade de classificação. A Figura 2.7 ilustra um exemplo de três *Haar Wavelets* sobrepostas em uma imagem da face. Outros trabalhos que utilizam características do tipo *Haar* são o de Wang et al. (2004) e Jung et al. (2005).

Jiang et al. (2011) utilizaram descritores *Local Binary Pattern* (LBP)(OJALA et al., 1996) para detecção de unidades de ação. No sistema proposto por eles, inicialmente a face é localizada, em seguida variações na posição da cabeça são removidas da imagem. Depois a face é alinhada automaticamente e dividida em pequenos blocos utilizando *Local Binary Patterns from Three Orthogonal Planes* (LBP-TOP)(ZHAO; PIETIKAINEN, 2007) e *Local Phase Quantisation from Three Orthogonal Planes* (LPQ-TOP). Em seguida, as características LBP e *Local Phase Quantization* (LPQ)(OJANSIVU; HEIKKILä, 2008) são extraídas e selecionadas. Os histogramas LBP obtidos são concatenados para representar a face. Segundo os autores, ao utilizar o classificador SVM para detectar nove AU's da parte superior da face, o LPQ alcançou taxas de reconhecimento maiores em oposição ao LBP, e o LPQ-TOP superou as outras abordagens experimentadas.

Características LBP também foram utilizadas no trabalho de Shan e Gritti (2008) em que barras de histogramas de LBP são aprendidas, a fim de discriminar diferentes expressões. Segundo os autores, as barras representam uma compacta e discriminativa representação das faces expressivas, porém, como nem todas as barras contribuíram para

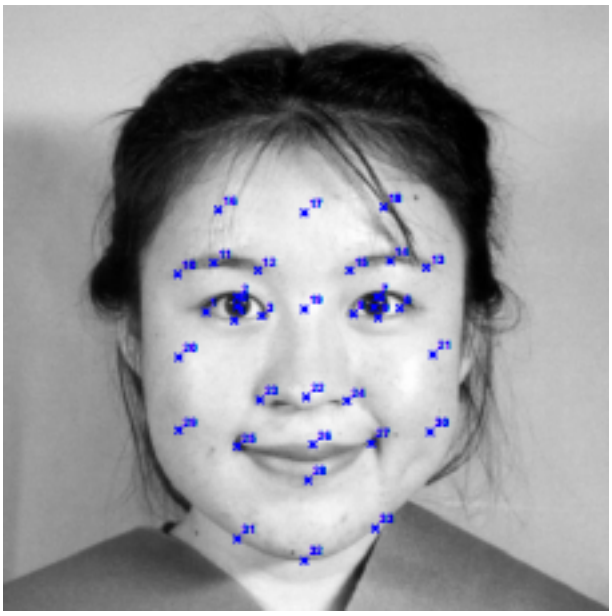


Figura 2.8: Representação Geométrica: 34 *landmarks* que representam a geometria facial. Fonte: (ZHANG et al., 1998)

o desempenho do reconhecimento, as mais relevantes foram selecionados através do classificador *AdaBoost*. Em Shan et al. (2009), inicialmente foram extraídas características LPB da imagem, em seguida o classificador *Boosting* foi utilizado para selecionar as características mais discriminantes e, por fim o *Support Vector Machine* foi empregado para classificar sete diferentes expressões. Ao utilizar a base FG-NET (WALLHOFF, 2006), o sistema alcançou uma taxa de reconhecimento de 82%.

2.1.3 Características Híbridas

Existem alguns sistemas de reconhecimento de expressões faciais que utilizam características híbridas, ou seja, combinam técnicas baseadas em aparência com as geométricas para compor o vetor de características. No sistema de Youssif e Asker (2011), por exemplo, são extraídos 19 *landmarks* em torno dos olhos, nariz e boca, além do histograma das regiões da borda da face representando características de aparência, as quais foram utilizadas como entrada para uma rede neural artificial. Os resultados experimentais demonstraram precisão de 93,5% ao classificar um conjunto de emoções básicas utilizando

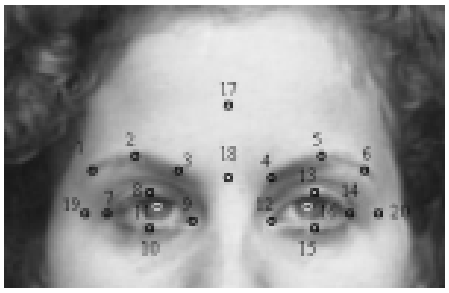
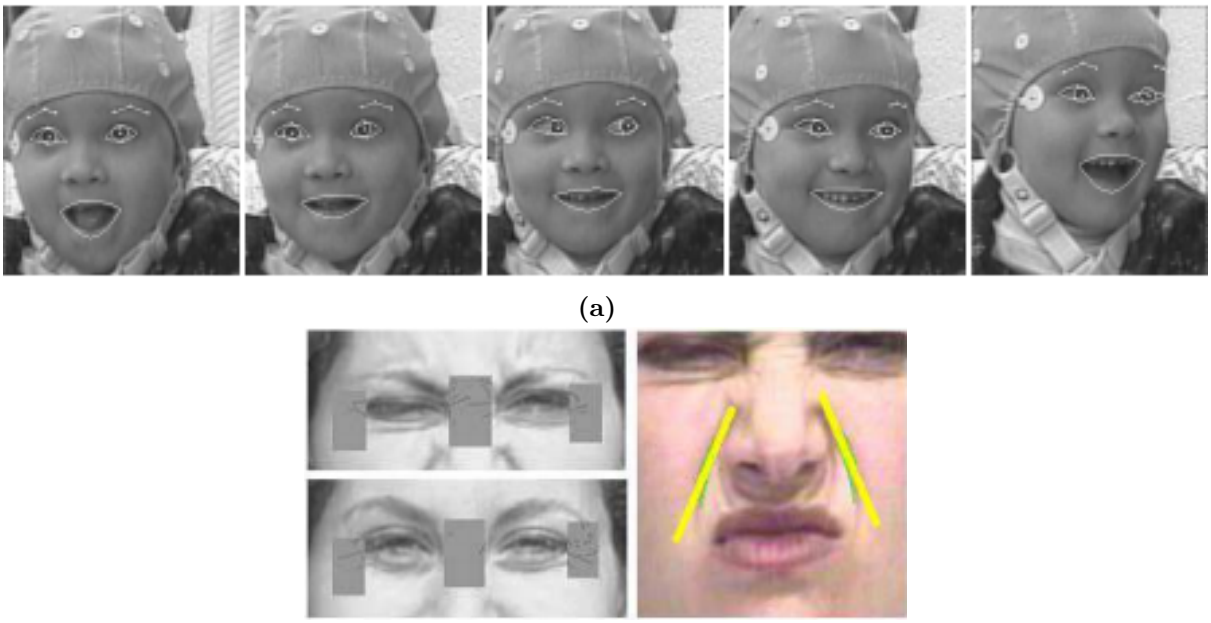


Figura 2.9: Locais para calcular coeficientes de *Gabor* na parte superior da face. Fonte: (TIAN et al., 2002)

a base de dados Cohen-Kanade (KANADE et al., 2000).

Zhang et al. (1998) utilizaram uma rede neural artificial para classificar uma combinação de 34 posições geométricas (Figura 2.8) e 612 coeficientes *Gabor wavelets* extraídos da imagem da face, a partir dos 34 *landmarks* selecionados manualmente. O sistema atingiu uma taxa de precisão de 63,5% ao classificar sete expressões distintas quando apenas as posições dos *landmarks* foram utilizadas para classificar. Ao utilizar apenas características de *Gabor*, o sistema consegue uma precisão de 89,6%. No entanto, o sistema combinado - coeficientes de *Gabor* e *landmarks*, apresentou taxa de acerto de 90,1%. Através dos resultados, pode-se perceber que o sistema ao utilizar características combinadas não apresentou melhores resultados que utilizando apenas *Gabor*, demonstrando que sistemas que utilizam características combinadas devem ser projetados com cuidado, a fim de obter benefícios em todas as características avaliadas.

Similarmente, Tian et al. (2002) desenvolveram um sistema utilizando uma rede neural artificial para classificar tanto características de *Gabor* quanto geométricas. Através de uma sequência de imagem, foram criados modelos das regiões faciais para extrair um grupo de 15 parâmetros que descreviam a forma, o movimento, o estado dos olhos, o movimento da testa e das bochechas, e os sulcos da parte superior da face. As características de movimento foram obtidas através de uma versão modificada do algoritmo de rastreamento chamado Lucas - Kanade (LUCAS; KANADE, 1981) que consiste basicamente em dividir a imagem em janelas para calcular o fluxo ótico (*optical flow*) em cada uma delas. Além disso, o *Gabor wavelets* foi utilizado para extrair as mudanças na aparência facial como um conjunto de coeficientes multiescala e multiorientação. Os autores segui-



(a)

(b)

Figura 2.10: (a) Características permanentes (b) Características transientes. Fonte: (TIAN et al., 2001)

ram Zhang et al. (1998), que utilizaram o filtro *Gabor* de uma forma seletiva, ao invés de aplicá-lo em toda a imagem. Foram calculados 800 coeficientes de *Gabor wavelets* em 20 locais diferentes da parte superior da face (Figura 2.9). Através dos experimentos, os autores concluíram que, ao utilizar apenas características geométricas, o sistema apresentou uma taxa de precisão de 87,6% e de 32% quando apenas características baseadas em aparência foram utilizadas. A melhor taxa de reconhecimento foi 92,7% obtida pela combinação de coeficientes de *Gabor* e características geométricas.

Em Lucey et al. (2010), os autores utilizaram o *Active Appearance Model* (AAM) proposto por Cootes et al. (1998) para extrair a forma e a aparência da face. Basicamente, os AAM's são modelos de aparência construídos através da combinação de um modelo estatístico da forma geométrica em conjunto com um modelo dos níveis de cinza de um determinado objeto. Após a aplicação do AAM, as características extraídas foram utilizadas no treinamento do classificador SVM para reconhecimento de expressões faciais. Segundo os autores, os experimentos demonstraram que a combinação entre as características da forma e aparência da face apresentaram resultados superiores em relação ao uso delas separadamente. Ao utilizar os dois tipos de características o sistema apresentou

uma taxa de classificação de 94,5%.

No trabalho apresentado por Tian et al. (2001) são desenvolvidos modelos multi-estados da regiões da face: três estados para o modelo do lábio (com os estados aberto, fechado e bem fechado), um modelo de dois estados para os olhos (como os estados aberto ou fechado), um estado para o modelo da testa e outro para a bochecha. Além disso, algumas características de aparência, tais como "pé-de-galinha", rugas na região nasal, e sulcos nasolabiais foram incorporadas a um modelo de dois estados (presente ou ausente). A Figura 2.10a ilustra as características permanentes (olhos, sobrancelhas, e boca) e a Figura 2.10b mostra as características transientes ("pé-de-galinha", rugas na região nasal, e sulcos, nasolabiais) detectadas pelos autores. Características transientes são linhas faciais e saliências que não se tornaram permanentes, por exemplo, com a idade, mas que são causadas pelas expressões. Em Zhang e Ji (2003), Zhang e Ji (2005) foi utilizado um conjunto de 26 características faciais em torno dos olhos, do nariz e da boca, além de um conjunto de características transientes da face similar ao trabalho de Tian et al. (2001).

No trabalho de Hupont et al. (2008b), os autores utilizaram um classificador baseado em regras para reconhecer sete diferentes expressões. Os experimentos foram divididos em três etapas. Primeiramente, os autores utilizaram apenas as distâncias entre alguns *landmarks* faciais e obtiveram uma taxa de reconhecimento de 71%. Na segunda etapa, os autores utilizaram o filtro de *Gabor* para detectar a presença de rugas na região nasal em conjunto com as características extraídas na primeira etapa e alcançaram 85% de acerto. Por fim, na terceira etapa, os autores associaram informações relativas a forma da boca em conjunto com as características extraídas da primeira etapa e obtiveram uma taxa de acerto de 91%.

Em Martin et al. (2008), os autores inicialmente aplicaram o *Active Appearance Model* para extrair características em imagens faciais para reconhecer expressões. Em seguida, três diferentes classificadores foram utilizados para fins de comparação. Os classificadores utilizados foram baseados em regras, *Support Vector Machines* e redes neurais artificiais multicamadas (*Multi-Layer Perceptrons* - MPL). Os autores perceberam que quanto mais características faciais foram utilizadas como entrada do classificador, maior

foi a taxa de reconhecimento. Eles também mostraram que a taxa de reconhecimento maior foi de 92% quando formas das regiões faciais em conjunto com a textura foram utilizadas como entradas para o classificador *Support Vector Machines*.

2.2 CLASSIFICAÇÃO DE EXPRESSÕES FACIAIS

Após extrair características, a classificação é realizada no último estágio de um sistema de análise de expressão facial, que consiste em um procedimento de decisão realizado geralmente por um classificador baseado nas características extraídas. As abordagens existentes na literatura normalmente descrevem mudanças nas expressões através de *expressões básicas universais* ou utilizando unidades de ação (AU's) do *Facial Action Coding System (FACS)*.

De acordo com a teoria de Ekman e Friesen (1971), existem seis expressões que são universais para os povos de diferentes nações e culturas, podendo ser chamadas de *expressões básicas universais* e consistem em: felicidade, raiva, tristeza, surpresa, desgosto e medo (ver Figura 2.11). Alguns autores consideram a face neutra (ausência de qualquer emoção) como uma sétima classe de expressão. Trabalhos como os de Black e Yacoob (1997), Huang e Huang (1997), Michel e Kaliouby (2003), Littlewort et al. (2004), Chuang e Shih (2006), Zeng et al. (2006) têm descrito mudanças nas expressões através de um conjunto ou subconjunto de *expressões básicas universais*.

Outra abordagem para descrever as expressões é através do *Facial Action Coding System (FACS)* desenvolvido por Ekman e Fiesen em 1978. Como definido no Capítulo 1, o sistema FACS descreve as expressões faciais em unidades de ação. Das 44 AU's definidas, 30 são anatomicamente relacionadas com a contração de músculos faciais específicos (12 AU's para a parte superior da face, e 18 AU's para a parte inferior), que podem corresponder a um músculo específico, ou a um grupo muscular. A Tabela 2.1 lista exemplos de algumas unidades de ação. Utilizando um conjunto de regras prescritas, um especialista em FACS pode descrever cada uma das expressões faciais através de combinações de AU's. Trabalhos como os de Cohn et al. (1999b), Donato et al. (1999), Pantic

Figura 2.11: Seis expressões faciais da esquerda para direita: felicidade, raiva, tristeza, surpresa, desgosto e medo.

e Rothkrantz (2000), Tian et al. (2002), Bartlett et al. (2006), Ryan et al. (2009), Jiang et al. (2011), entre outros, utilizam a presente abordagem para descrever as alterações faciais.

Segundo Tian et al. (2005) os métodos de classificação de expressões podem ser normalmente baseados em imagens estáticas ou baseados em sequência de imagens. Métodos baseados em imagens estáticas utilizam apenas o *frame* corrente, com ou sem uma imagem de referência (normalmente é utilizada a imagem de face neutra) para reconhecer as expressões apresentadas nas imagens. Por outro lado, os baseados em sequências de imagens utilizam informações temporais de pelo menos dois exemplos de uma face no mesmo estado emocional. A classificação baseada em imagens estáticas e sequências de imagens são apresentadas nas subseções subsequentes.

Tabela 2.1: Alguns exemplos de unidades de ação (EKMAN; FRIESEN, 1978)

Número da AU	Nome FACS	Base Muscular
1	Elevação da parte interior das sobrancelhas	Frontalis, Pars Medialis
6	Elevação da bochecha	Orbicularis Oculi, Pars Orbitalis
7	Compressão das pálpebras	Orbicularis Oculi, Pars Palebralis
12	Alongamento dos cantos da boca	Zygomatic Major
17	Elevação do queixo	Mentalis
20	Esticador dos lábios	Risorius

2.2.1 Classificação de Expressões Faciais em Imagens Estáticas

Métodos pertencentes a esta categoria utilizam informações da imagem de entrada com ou sem *frame* de referência para reconhecer a expressão, podendo ser apenas uma imagem ou um conjunto de imagens. Vários métodos podem ser encontrados na literatura para reconhecer expressões faciais em imagens estáticas, entre estes, os mais utilizados são os métodos baseados em redes neurais artificais, *Support Vector Machines (SVMs)* e regras. A seguir são relatados trabalhos que utilizam métodos de cada uma das três diferentes categorias citadas.

2.2.1.1 Métodos Baseados em Redes Neurais Artificiais

Uma Rede Neural Artificial (ANN) é um grupo interligado de neurônios artificiais que utilizam um modelo matemático para processamento de informações. Na maioria dos casos uma ANN é um sistema adaptativo que muda sua estrutura, ou pesos, com base na informação externa ou interna, que flui através da rede. A rede neural artificial desempenha um papel importante no reconhecimento de expressão facial (HAYKIN, 1998). Em Padgett et al. (1996), uma rede neural artificial foi utilizada para classificar seis expressões universais como definidas por Ekman e Friesen (1971). A taxa média de reconhecimento alcançado pelo sistema foi de 86%. Ainda em 1996, também utilizando redes neurais artificiais, Bartlett et al. (1996) alcançaram uma taxa de acerto de 89% de seis AU's da face superior utilizando FACS.

Kobayashi e Hara (1997) desenvolveram um sistema para reconhecimento de expressões empregando uma rede neural artificial utilizando o algoritmo de treinamento

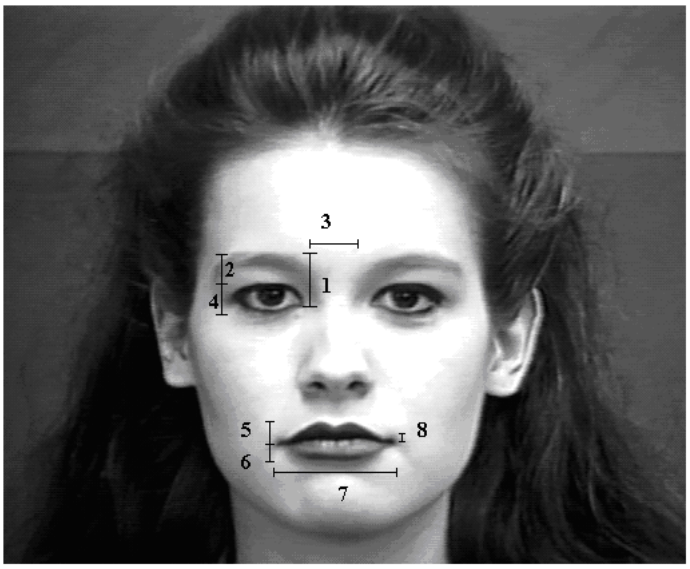


Figura 2.12: Medidas de valores reais de uma imagem da face representando a expressão neutra. Fonte: (SAKET et al., 2009)

back-propagation. As unidades da camada de entrada correspondiam a dados relacionados à distribuição do brilho extraído a partir da imagem de entrada da face enquanto que cada unidade da camada de saída equivalia a uma categoria da expressão. A rede neural artificial foi treinada em imagens, nas quais 15 indivíduos apresentavam seis expressões básicas. A taxa média de reconhecimento foi de 85%.

No trabalho realizado por Tian et al. (2001) foi desenvolvido um sistema automático de análise da face chamado *Automatic Face Analysis* (AFA) utilizando duas redes neurais artificiais (uma para a parte superior e a outra para a parte inferior da face). O sistema reconheceu mudanças na expressão facial em AU's do sistema FACS atingindo uma taxa de reconhecimento média de 95% para AU's da parte superior do rosto e 96,7% para AU's da parte inferior da face.

Saket et al. (2009) desenvolveram um sistema para classificar expressões faciais em imagens estáticas utilizando um comitê de redes neurais. Resumidamente, um comitê de redes neurais consiste em um arranjo de redes independentes, trabalhando em paralelo, mas no sentido de uma classificação única e consensual. No trabalho de Saket et al. (2009) dois tipos de parâmetros foram extraídos a partir de imagens da face de 97 indivíduos:

Figura 2.13: Medidas binárias a partir de exemplos de imagens da face com diferentes expressões. Fonte: (SAKET et al., 2009)

(1) parâmetros reais e (2) parâmetros binários. Os parâmetros reais dependiam do valor da distância em *pixels* encontrados entre as regiões faciais. Estes parâmetros reias são representados na Figura 2.12, sendo 1- distância da sobrancelha levantada, 2- distância entre a pálpebra superior e a sobrancelha, 3- distância entre sobrancelhas, 4- distância entre pálpebra superior e pálpebra inferior, 5- espessura do lábio superior, 6 -espessura do lábio inferior, 7- largura da boca e 8- abertura da boca. As medidas binárias correspondiam à presença (= 1) ou ausência (= 0) de determinadas características. Estes parâmetros binários são representados na Figura 2.13, sendo 1- dentes superiores visíveis, 2- dentes inferiores visíveis, 3- linhas da testa, 4- linhas da sobrancelha, 5- linhas do nariz, 6- linhas do queixo e 7- linhas nasolabiais. Ao todo, um total de 15 parâmetros (8 reais e 7 binários) foram obtidos. As redes neurais artificiais foram treinadas com todos os 15 parâmetros para reconhecer sete expressões distintas (felicidade, raiva, tristeza, surpresa, desgosto, medo e neutra). Segundo os autores o sistema conseguiu classificar aproximadamente 90,43% dos casos.

No estudo realizado por Zhao e Kearney (1996), foi empregada uma rede neural

artificial utilizando o algoritmo de treinamento *back-propagation* para classificar determinada imagem facial em uma das seis categorias de expressões básicas. Os dados de entrada da rede neural artificial consistiram em um conjunto de intervalos, que resultou do tratamento estatístico das distâncias normalizadas entre vários pontos da face. O sistema apresentou uma taxa de reconhecimento de 100% para indivíduos conhecidos, enquanto que os autores não informaram a taxa de acerto para indivíduos novos no sistema.

Muitos outros trabalhos são encontrados na literatura que utilizam redes neurais artificiais para reconhecer expressões faciais que podem ser empregados em diferentes aplicações, entre estes estão os de Zhang et al. (1998), Padgett e Cottrell (1996), Zhang e Zhang (1999), Yoneyama et al. (1997), Oliveira (2000), Dailey et al. (2000), Tian et al. (2001), Stathopoulou e Tsihrintzis (2004), Sebe et al. (2007), Youssif e Asker (2011), Giripunje e Bajaj (2012), Fasel et al. (2004), Sun et al. (2009).

2.2.1.2 Metodos Baseados em *Support Vector Machines*

Support Vector Machines (SVMs) consistem em uma técnica de aprendizagem de máquina para resolver problemas de reconhecimento de padrão. Foram introduzidas por Vapnik (1995) através da teoria estatística de aprendizagem. As SVMs se propõem a obter um hiperplano, como superfície de decisão, de maneira que a margem de separação entre exemplos positivos e negativos seja maximizada (HAYKIN, 1998).

A técnica de SVM tem sido aplicada com sucesso em inúmeras tarefas de classificação, entre elas, para o reconhecimento de expressões faciais. Como por exemplo, o estudo realizado por Tian et al. (2002), no qual primeiramente foram extraídas características de *Gabor* de imagens da face e, em seguida, o SVM foi treinado para reconhecer seis diferentes expressões. Ao utilizar a base de dados CMU (SIM et al., 2002) o sistema apresentou uma taxa média de acerto de 90%. Shan et al. (2005) também utilizaram o SVM com características LBP e obtiveram resultados semelhantes ao utilizar a mesma base de dados.

Chuang e Shih (2006) propuseram um sistema para classificação de expressão facial

que utiliza *Independent Component Analysis* (ICA) como método de extração de características e o classificador SVM para reconhecer AU's individuais (definidas no Capítulo 2 e na seção 2.2), ou combinações destas. Os autores realizaram uma comparação do método desenvolvido que apresentou taxas de reconhecimento de 97,06% (AU da parte superior da face), 97,13% (AU da parte inferior da face) e 100% (AU de toda a face neutra) com os trabalhos de Tian et al. (2002) e Donato et al. (1999), que acertaram respectivamente 95,6% (AU da parte superior da face) e 96,9% (AU da parte inferior da face), 86,55% e 81,63%.

Recentemente, trabalhos como os dos pesquisadores Nagpal e Garg (2011) utilizaram o *Active Appearance Model* (AAM) que são modelos de aparência construídos através da combinação de um modelo estatístico da forma geométrica, com um modelo dos níveis de cinza de um determinado objeto para extrair características e o SVM para classificar quatro distintas expressões. O sistema apresentou uma taxa de reconhecimento de 82,7%. Outro trabalho é de Visutsak (2012) no qual foi utilizado um vetor de deslocamento de características como entrada para o *Adaptive Support Vector Machines (A-SVMs)*. Os experimentos foram realizados através da base de dados JAFEE (LYONS et al., 1998b) sendo que o sistema atingiu uma taxa média de classificação de aproximadamente 75%.

O SVM também foi utilizado nos trabalhos de Bartlett et al. (2003), Anderson e McOwan (2006), Littlewort et al. (2002), Zhan et al. (2006b), Lucey et al. (2010), Martin et al. (2008), Samad e Sawada (2011).

2.2.1.3 Métodos Baseados em Regras

Recentemente Khanam et al. (2008) apresentaram um sistema nebuloso (*fuzzy*) do tipo Mamdani, baseado em regras, para reconhecer expressões faciais utilizando a base de conhecimento dividida em dois componentes principais: base de dados e base de regras. A primeira foi constituída pela entrada do sistema formada por diversos estados das características faciais, e a saída, que correspondia a cada uma das sete expressões. A segunda foi formada por um conjunto de regras *fuzzy* que foram divididas em dois grupos: as regras principais que classificaram as expressões faciais e as regras secundárias

que possibilitaram uma sobreposição entre expressões permitindo uma transição sutil entre as expressões básicas e tendo assim, um menor peso na classificação. A taxa média de reconhecimento foi de 87,5%.

Pantic e Rothkrantz (2004) propuseram um método para detectar unidades de ação através dos contornos das regiões faciais, tais como olho e boca, nos quais foram extraídos 19 *landmarks*. Nesse trabalho foram reconhecidas 32 unidades de ação e ao utilizar um classificador baseado em regras o reconhecimento atingido pelo sistema foi 86%.

Pantic e Rothkrantz (2000) estimaram alguns pontos principais da face com o intuito de obter um modelo para que, em seguida, fosse calculada a diferença entre as características do modelo com as da face neutra do indivíduo. As expressões foram classificadas através da deformação do modelo em conjunto com as unidades de ação, apresentando uma taxa de acerto de 92% para a parte superior da face e 86% para a inferior da face.

2.2.2 Classificação de Expressões Faciais em Sequências de Imagens

Métodos que pertencem a esta categoria utilizam informações temporais das sequências de imagens para reconhecer as expressões. Em experimentos realizado pelo psicológico Bassili (1979) conclui-se que a dinâmica das expressões é importante ao interpretá-las. Para utilizar a informação temporal, algumas dos métodos mais empregados na análise de expressão facial são os probabilísticos, tais como *Hidden Markov Model (HMM)* e as redes bayesianas dinâmicas, além dos métodos baseados em correspondência entre modelos. A seguir são relacionados alguns trabalhos que utilizaram as técnicas mencionadas para reconhecer expressões faciais em sequencias de imagens.

2.2.2.1 Métodos Probabilísticos

Muitos campos de conhecimento têm aplicado métodos probabilísticos com sucesso. Nas áreas em que se tenta simular a percepção ou o comportamento humano os métodos métodos probabilísticos são bastante utilizados. Na análise de expressão facial, os dois

métodos probabilísticos normalmente utilizados são o *Hidden Markov Model* (HMM) e as Redes Bayesianas Dinâmicas. A seguir são descritos alguns trabalhos que utilizam cada uma dessas técnicas.

Hidden Markov Model

Hidden Markov Model (HMM) é uma ferramenta probabilística utilizada para modelagem de séries temporais, e tem sido amplamente utilizada no reconhecimento de voz (PATEL; RAO, 2010). Entretanto, o método HMM também tem sido explorado para capturar comportamentos temporais exibidos através de expressões faciais. Oliver et al. (2000) empregaram HMM para reconhecer expressões a partir do rastreamento da deformação das formas da boca em tempo real, sendo que cada uma das expressões foi associada com um HMM treinado. As expressões faciais foram identificadas a partir do cálculo da probabilidade máxima da verossimilhança da sequência de imagens de entrada, em relação a todos os HMMs treinados. Os autores afirmaram que a taxa de acerto se aproximou de 100%.

Yeasin et al. (2004) apresentaram uma abordagem em duas etapas para classificar seis emoções básicas. Primeiramente, um conjunto de classificadores lineares foi aplicado para cada *frame* para produzir uma assinatura temporal. Na segunda etapa, as assinaturas temporais foram calculadas para treinar HMMs discretos para aprender os modelos de cada emoção.

O Hidden Markov Model também foi aplicado em Cohen et al. (2003), Bartlett et al. (2001) Otsuka e Ohya (1998), Lien et al. (2000) e Cohn et al. (1999b).

Redes Bayesianas Dinâmicas

Redes Bayesianas Dinâmicas são modelos probabilísticos que codificam as dependências entre um conjunto de variáveis aleatórias ao longo do tempo. As redes Bayesianas normalmente tem a capacidade de contabilizar incertezas no reconhecimento da expressão facial, representando relações probabilísticas entre diferentes ações (SHAN et al., 2009). Zhang e Ji (2005) utilizaram a técnica de fusão de informações multissensorial com redes

Bayesianas Dinâmicas para modelar comportamentos temporais de expressões faciais em sequências de imagens.

Cohen et al. (2003) desenvolveram um sistema de reconhecimento de expressões no qual inicialmente são extraídas características de movimento que são utilizadas como entrada para um classificador de redes Bayesianas. A capacidade de aprendizagem com dados não rotulados, a possibilidade de inferência do rotulo da classe mesmo quando algumas características não estão presentes e a viabilidade de estender o sistema para outras modalidades adicionando sub-redes, foram os motivos que motivaram esses autores a utilizar as redes Bayesianas.

Outros trabalhos que tratam o reconhecimento de expressões faciais envolvendo redes Bayesianas Dinâmicas são relatados em Gu e Ji (2004), Hoey e Little (2004), Tong et al. (2006).

2.2.2.2 Métodos Baseados em Correspondência entre Modelos

Os métodos baseados em correspondência entre modelos geralmente criam um modelo de referência (modelo médio) para cada expressão. Em seguida, normalmente se compara o modelo da expressão facial da imagem de entrada com os modelos de referência. A melhor correspondência decide a categoria da expressão realizada (PANTIC et al., 2000). Em trabalhos como os de Cohn et al. (1998), funções discriminantes foram aplicadas separadamente para análise do movimento das regiões faciais, tais como sobrancelhas, olhos e boca. Os autores utilizaram duas funções discriminantes para três ações na área das sobrancelhas, duas na área dos olhos, e cinco funções discriminantes para nove ações faciais na região do nariz e da boca. Em seguida, foi calculado o deslocamento dos pontos entre a imagem inicial e a imagem corrente, gerando-se grupos separados de variância-convariância, utilizados para classificação.

Essa e Pentland (1997) extraem características espaciais e temporais de uma sequencia de imagens para criar um modelo espaço-temporal para cada uma das seis expressões sendo, duas para ações faciais (sorriso e elevação das sobrancelhas) e quatro para expressões representativas do estado emocional (surpresa, tristeza, raiva e desgosto). As

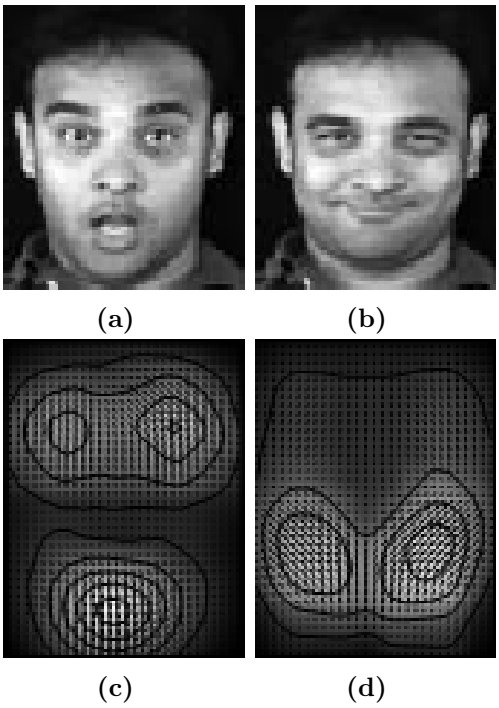


Figura 2.14: Determinação de expressões em sequências de imagens. Fonte: (ESSA; PENTLAND, 1997)

Figuras 2.14(a) e (b) ilustram as expressões de surpresa e felicidade e em (c) e (d) são mostrados exemplos da representação dos modelos espaço-temporal associados ao movimento facial das expressões surpresa e felicidade, tomando-se como referência a imagem da face neutra.

2.3 QUESTÕES E DESAFIOS

Como descrito nas seções anteriores, vários métodos têm sido propostos para reconhecimento de expressão facial. No entanto, algumas questões e alguns desafios ainda permanecem no desenvolvimento de um sistema de reconhecimento de expressão facial. Algumas destas questões e desafios, conforme Tian et al. (2005) e Zhan et al. (2006b), são descritos nos itens seguintes.

2.3.1 Base de Dados

Muitas vezes os bancos de dados utilizados para reconhecimento de expressões faciais são relativamente limitados, fato que não favorece a avaliação de desempenho dos diferentes sistemas. Na maioria das bases de dados utilizadas pela literatura atual, apenas algumas expressões faciais realizadas por um número limitado de indivíduos são consideradas. Os indivíduos que compõem a base de dados frequentemente apresentam a mesma idade e origem étnica, e as condições de gravação são controladas. O reconhecimento de expressão facial, na prática, normalmente não pode ser alcançado exclusivamente com base nesses dados de treinamento os quais, usualmente, se mostram sensíveis a variações nas expressões, contextos e propriedade da imagem. Assim, pesquisas sobre reconhecimento de expressões faciais, bancos de dados bem definidos e critérios de avaliação para validar as bases de dados ainda são objetos de estudo.

2.3.2 Resolução da Face

Imagens faciais em baixa resolução, muitas vezes podem fornecer menos informação sobre as características faciais. Segundo Tian et al. (2003), a maioria dos métodos existentes para detecção da face trabalham com resolução de 36 x 48 *pixels* ou superior. Sistemas utilizando características geométricas normalmente não são capazes de alcançar um desempenho relativamente bom quando as resoluções da face são menores que 72 x 96 *pixels*, enquanto que o limite de resolução para os métodos baseados em aparência é de 36 x 48 *pixels*. Apenas poucos trabalhos tentam reconhecer expressões através de faces que apresentam baixa resolução, tais como Tian e Chen (2012), Smith et al. (2001), Liao et al. (2006) e, Lien et al. (2006).

Figura 2.15: Imagens apresentando variações na iluminação. Fonte: (FEI-FEI et al., 2007)

2.3.3 Variação do Ambiente

A variação no ambiente é um dos problemas que continua sendo pesquisado no campo de reconhecimento de expressão facial. As variações em plano de fundo complexo, presença de outras pessoas e as condições de iluminação não controladas, frequentemente dificultam o reconhecimento das expressões. A maioria dos trabalhos apresenta um conjunto de dados de treinamento em imagens de plano de fundo neutro ou apenas uma pessoa encontra-se presente na cena. A Figura 2.15 ilustra algumas imagens que evidenciam tais limitações. Métodos que apresentam bom desempenho em ambiente experimental com condições de iluminação controladas pode ter um desempenho limitado em ambientes que apresentam iluminação natural. Para evitar a influência das variações do ambiente, os pesquisadores normalmente utilizam em suas pesquisas imagens que apresentam plano de fundo organizado e iluminação controlada, embora tais condições não retratem os ambientes na prática. Trabalho como o de Lu et al. (2007) apresentou taxa de sucesso satisfatória mesmo sob condições variáveis de luminosidade, porém não são relatadas informações em relação a robustez quanto à rotação e artefatos faciais adicionais, tais como óculos, barbas, etc.

2.3.4 Posição da Cabeça

Levando em conta que a maioria dos sistemas utiliza câmera fixa, as restrições são muitas vezes impostas sobre a posição e orientação da cabeça com relação à câmera para garantir que a imagem de entrada da face apareça em posição frontal ou quase frontal. No entanto, na prática, as rotações da cabeça ocorrem com frequência, por isso é necessário

Figura 2.16: Imagens apresentando variações na posição da cabeça. Fonte: (SIM et al., 2002)

Figura 2.17: Imagens apresentando variações na aparência. Fonte: (MILBORROW et al., 2010)

o desenvolvimento de métodos de reconhecimento de expressões invariantes a rotações. As rotações da cabeça podem ser parcialmente resolvidas pela normalização antes da extração de características faciais. Pantic e Rothkrantz (2000) foram os primeiros a utilizar duas câmeras, sendo uma delas colocadas em frente à face e a outra do lado direito da face. Nesse trabalho as câmeras estão em movimento juntamente com a cabeça para eliminar efeitos de escala e variação na orientação das imagens da face adquirida. Alguns trabalhos lidam com variação da rotação sem utilizar múltiplas câmeras, tais como Xiao et al. (2002), Shih e Chuang (2004). A Figura 2.16 apresenta algumas imagens que ilustram variações na posição da cabeça.

2.3.5 Diferenças Individuais

Características faciais, tais como a forma, a textura e a cor, variam de acordo com o gênero, etnia e idade, causando impactos no reconhecimento de expressões da face. Tais impactos podem ser observados na análise facial entre um asiático e um europeu, pois as diferenças quanto à abertura dos olhos e contraste entre a íris e a esclera geralmente

afetam a robustez dos métodos de extração de características faciais. Outros fatores que podem afetar o reconhecimento das expressões é a presença das barbas, dos óculos, ou de maquiagens que podem ofuscar as características da face. Segundo Tian et al. (2005) poucas pesquisas foram realizadas para solucionar tais problemas com exceção dos autores Zlochower et al. (1998) que inicialmente aplicaram o algoritmo de fluxo ótico otimizado em faces de adultos e quando empregaram o mesmo algoritmo em faces de bebês observaram que os resultados foram inferiores. A textura da pele das crianças, a falta de sulcos, pelos faciais entre outras características podem ter contribuído para os diferentes resultados apresentados entre crianças e adultos. Portanto para desenvolver algoritmos que sejam robustos às diferenças individuais, é fundamental um conjunto considerável de amostras de indivíduos de diferentes etnias, idade e sexo que apresentem cabelo facial, joias, óculos, maquiagens, etc.

2.4 RESUMO

Neste capítulo, foram descritos alguns dos importantes trabalhos encontrados na literatura, tendo em vista os principais métodos de extração de características e de classificação utilizados no reconhecimento de expressões faciais. Alguns trabalhos relatam que os métodos baseados em características geométricas são muitas vezes sensíveis a variação da forma e a resolução da imagem, enquanto que os baseados em aparência podem conter informações redundantes (ZHAN et al., 2006b). Segundo estudos realizados por Bartlett et al. (1999) sistemas que utilizam características baseadas em aparência geralmente apresentam melhor desempenho. Entretanto, recentemente Pantic e Patras (2006) demonstraram que, na maioria dos casos, sistemas que empregaram características geométricas consomem menos capacidade computacional do que os utilizam características baseadas em aparência, já que lidam com alguns pontos sobre a face, ao invés de toda a imagem da face.

No presente capítulo, os métodos de classificação foram categorizados de acordo com imagens estáticas e sequências de imagens. Vários métodos de classificação foram men-

cionados, porém, muitos outros classificadores clássicos para reconhecimento de padrões tem sido aplicado a expressões faciais, tais como *Gaussian Mixture Model* (GMM), *Multinomial Logistic Ridge Regression* (MLR), *Naive Bayes*. Cada classificador tem suas vantagens e desvantagens e muitas vezes a escolha do método de classificação depende do contexto da aplicação.

Independente do tipo de características utilizadas, estas devem ser discriminativas e preferencialmente não correlacionadas. Além do mais, é fundamental que as informações das expressões exibidas pelo indivíduo sejam preservadas. Já no processo de classificação, o desempenho de cada classificador depende de muitos fatores como, por exemplo, o conjunto dos dados de treinamento e as características utilizadas.

Os trabalhos apresentados neste capítulo, de uma maneira geral, apresentaram sucesso ao serem experimentados no banco de dados escolhido pelos seus respectivos pesquisadores. No entanto, para que sistemas de reconhecimento de expressões faciais apresentem sucesso também na prática, os métodos desenvolvidos devem lidar com problemas que ocorrem em ambientes reais, tais como resolução da imagem, variação do ambiente, posição da cabeça, diferenças individuais, etc. No próximo capítulo é apresentado o sistema proposto neste trabalho, o qual visa prover o reconhecimento de sete diferentes expressões (felicidade, raiva, tristeza, surpresa, desgosto, medo e neutra), em imagens nas quais os indivíduos apresentam algumas diferenças individuais ou artefatos, tais como óculos, bigode e barba em planos de fundo não uniforme e neutro. Além disso, as imagens utilizadas pelo sistema, estão restritas a ambientes fechados.

As bases de dados escolhidas para treinar e testar o sistema foram: MUG *Facial Expression* (AIFANTI et al., 2010) e *Face and Gesture Recognition Research Network* (FG-NET) (WALLHOFF, 2006).

CAPÍTULO 3

SISTEMA PROPOSTO

Neste capítulo, é apresentado o sistema automático de reconhecimento de expressão facial desenvolvido no presente trabalho. O sistema proposto classifica cada imagem apresentada em uma das sete classes de expressões faciais: felicidade, raiva, tristeza, surpresa, desgosto, medo e neutra. A Figura 3.1 ilustra uma visão geral do sistema que consiste nos seguintes módulos: detecção da face, detecção das regiões faciais, extração de características e classificação de expressões. O sistema proposto neste trabalho inicialmente localiza a face e regiões faciais (os olhos, as sobrancelhas e a boca) em uma determinada imagem, em seguida extrai características da face (*landmarks* faciais), e por fim, reconhece a expressão apresentada na imagem.

Nas próximas seções, os requisitos do sistema de reconhecimento de expressões faciais desenvolvido neste trabalho são especificados. São descritos também os procedimentos realizados em cada um dos módulos do sistema.

Figura 3.1: Arquitetura do sistema proposto.

3.1 ESPECIFICAÇÕES DO SISTEMA

O processo de reconhecimento, incluindo as etapas (detecção da face, detecção das regiões faciais, extração das características, e classificação de expressões faciais), deve ser

executado automaticamente em um conjunto de imagens. O sistema deve ser treinado e testado em imagens que apresentem iluminação uniforme, planos de fundo não uniforme e neutro e variações na aparência, tais como óculos, bigode e barba. As imagens utilizadas pelo sistema estão restritas a ambientes fechados.

Neste trabalho, primeiramente, a face deve ser localizada nas imagens de entrada. Entre os métodos existentes na literatura, o método adotado neste trabalho é baseado em *Haar-like-features* como extrator de características e *AdaBoost* como classificador (VIOLA; JONES, 2001). Este método foi escolhido, devido ao seu baixo custo computacional e sua precisão na detecção (ZHAN et al., 2006b). Uma descrição detalhada do método pode ser encontrada no Apêndice A. Em trabalhos como os dos autores Bartlett et al. (2001), Vukadinovic e Pantic (2005), Zhan et al. (2006b), Lu et al. (2007), Cao e Tong (2008) o método proposto por Viola e Jones (2001) também foi utilizado para detectar a face.

O método desenvolvido por Viola e Jones (2001) também foi utilizado para localizar as regiões faciais (olhos e boca), com a finalidade de obter vantagens quanto à baixa taxa de processamento computacional do *Haar-like features* (ZHAN et al., 2006b).

A abordagem utilizada neste trabalho para extração das características faciais foi baseada em características geométricas. Conforme relacionado no Capítulo 2 (subseção 2.4), métodos baseados em aparência geralmente ultrapassam o desempenho dos métodos baseados em geometria (BARTLETT et al., 1999). Entretanto, recentemente Pantic e Patras (2006) e Valstar e Pantic (2006) demonstraram que em alguns casos, as características geométricas superam as baseadas em aparência. Além do mais, métodos geométricos frequentemente possuem menor custo computacional que os métodos que utilizam características baseados em aparência, pois lida apenas com determinados pontos na imagem (SHAN; BRASPENNING, 2010). Neste trabalho, assim como em Khandait et al. (2012) são aplicadas diferentes técnicas de pré-processamento e segmentação para extrair características. Para representar as expressões foi extraído um conjunto de 20 *landmarks faciais*, que foram escolhidos devido a serem considerados uns dos pontos mais discriminativos da face, sendo possível através destes obter informações como largura e altura das regiões faciais (os olhos, as sobrancelhas e a boca).

Como discutido no Capítulo 2, mudanças nas expressões podem ser descritas através de unidades de ação (AU's) ou *expressões básicas universais*. Pesquisadores como Pantic et al. (2000) e Fasel e Luettin (2003) consideram que AU's apresentam melhor desempenho para classificar classes de expressões, uma vez que normalmente podem descrever quase todas as possíveis mudanças faciais, especialmente as sutis. No entanto, a codificação de expressões através de AU's normalmente é um trabalho árduo, pois é realizada manualmente seguindo um conjunto de regras prescritas, além disso, é comumente necessário um especialista treinado em *Facial Action Coding System* (FACS) para codificar as mudanças faciais através de AU's. Sendo assim, neste trabalho as alterações nas expressões são descritas através de um conjunto de *expressões básicas universais*.

Ainda no Capítulo 2, os sistemas de *Facial Expression Recognition* (FER) podem ser categorizados como baseados em imagens estáticas ou baseados em sequência de imagens. O presente trabalho se desenvolve no contexto de imagens estáticas. Apesar de alguns trabalhos disponíveis na literatura demonstrarem que métodos baseados em sequência apresentam taxas de reconhecimento satisfatório, como é o caso de Essa e Pentland (1997), Oliver et al. (2000), e Yeasin et al. (2004), as informações de uma única imagem usualmente são suficientes para reconhecer a expressão. Além disso, em muitas aplicações de multimídia e interface homem-máquina, as informações temporais detalhadas das expressões não estão disponíveis, mas apenas uma única imagem estática. Assim sendo, constata-se que apesar da existência de abordagens baseadas em sequência de imagens, a abordagem baseada em imagens estáticas ainda apresenta amplo espectro de aplicações (GAO et al., 2003).

Nas seções seguintes, o algoritmo proposto para cada módulo do sistema é descrito em detalhes. Para os módulos de detecção da face e detecção das regiões faciais (olhos e boca), o método utilizado é descrito em detalhes no Apêndice A.

3.2 DETECÇÃO DA FACE

A primeira etapa do sistema de reconhecimento de expressões proposto neste trabalho é localizar a face na imagem. Normalmente detectar a face inicialmente é vantajoso, pois a busca pelas características fica limitada a apenas a região da face. Para localizá-la foi utilizado um detector baseado em *Haar-like-features* como extrator de características e *AdaBoost* como classificador (VIOLA; JONES, 2001). Como visto no Capítulo 2 (subseção 2.3.2) é normalmente complicado detectar a face em imagens que apresentam baixas resoluções, pois além da face, as regiões faciais podem tornar-se menos detectáveis. Neste trabalho as imagens utilizadas para detecção da face apresentam resoluções de 896x896 *pixels* e 640x480 *pixels*.

3.3 DETECÇÃO DAS REGIÕES FACIAIS

A região da face detectada representa uma região de interesse na imagem para que em seguida as características sejam extraídas. Com o intuito de reduzir ainda mais a região de busca pelas características, as áreas como sobrancelhas, olhos e boca também foram localizados após a detecção da face. Para isto, foi utilizado também o método de Viola e Jones (2001) para localizar as regiões faciais (olhos e boca) dentro da área da face detectada. A região da sobrancelha foi encontrada a partir da localização da região dos olhos. Para isto, foi definido um retângulo apresentando resolução em *pixels* de 17x5 para cada uma das sobrancelhas. O algoritmo de Viola e Jones (2001) encontra-se disponível na biblioteca OpenCV para detecção da face, dos olhos e da boca.

O método de Viola e Jones também pode ser usado para a detecção da sobrancelha, porém neste trabalho optou-se por detectar a sobrancelha dada a região dos olhos por ser uma alternativa prática. A Figura 3.2 ilustra da esquerda para direita as etapas utilizadas para detectar as regiões faciais: face original, detecção da face pelo método de Viola-Jones e detecção das regiões de interesse são detectados.

Após a detecção das regiões de interesse (ROIs) são extraídas características da

face. Este procedimento é descrito na próxima seção.

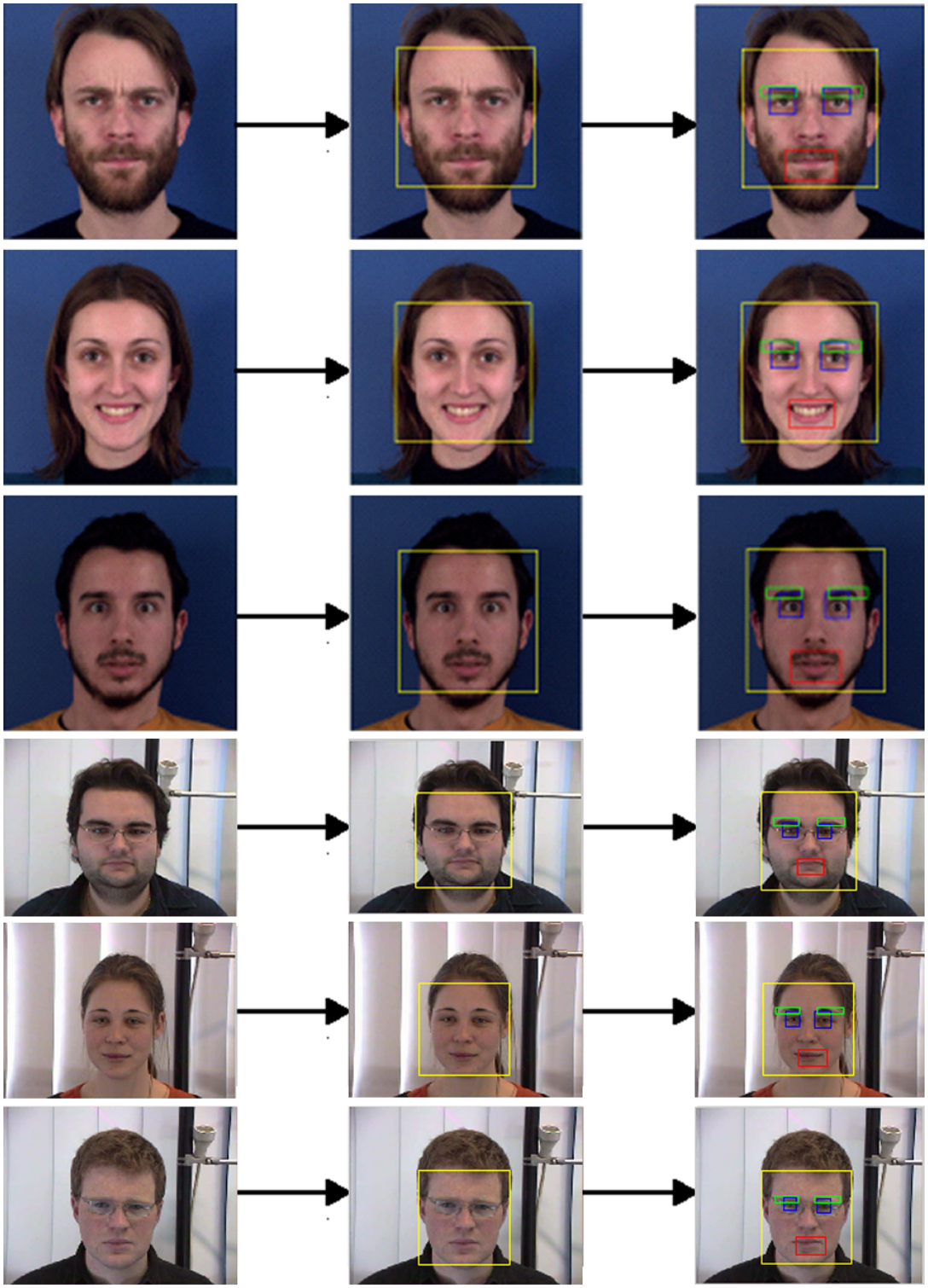


Figura 3.2: Da esquerda para direita: face original, detecção da face pelo método de Viola-Jones e detecção das regiões de interesse.

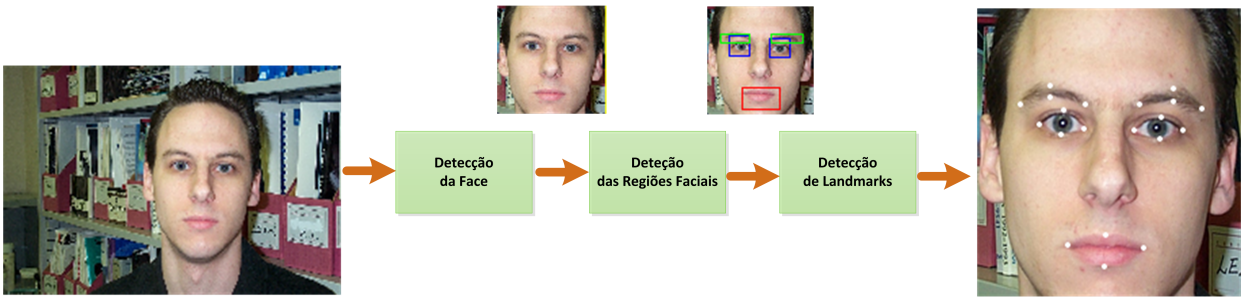


Figura 3.3: Processo para localização dos *landmarks* faciais.

3.4 EXTRAÇÃO DE CARATERÍSTICAS FACIAIS

Neste trabalho o tipo de característica extraída para representar cada uma das expressões foi um conjunto de *landmarks* faciais localizados na região da face. A Figura 3.3 ilustra o processo para localização de 20 *landmarks* faciais distribuídos por toda a face para descrever a formas das regiões faciais (das sobrancelhas, dos olhos e da boca). Estes *landmarks* foram extraídos a partir de um conjunto de imagens.

Para localizar os *landmarks* na face, inicialmente aplicou-se um conjunto de técnicas de processamento de imagem para melhorar a qualidade da imagem de entrada e realizar sua respectiva segmentação com o objetivo de extrair informações relevantes da imagem. As técnicas utilizadas foram escolhidas, depois de uma série de experimentos incluindo técnicas que foram testadas sem sucesso. Após os experimentos, as técnicas de processamento de imagens utilizadas neste trabalho foram: (a) equalização de histograma; (b) um filtro gaussiano; (c) ajuste de contraste; (d) métodos de segmentação: limiarização e operações morfológicas, e um (e) detector de bordas. As técnicas utilizadas e a busca pelos 20 *landmarks* faciais em cada uma das regiões de interesse são descritas detalhadamente a seguir.

3.4.1 Pré-processamento da Região do Olho

Para melhorar a qualidade da imagem, e tornar o contorno do olho mais destacado, inicialmente foi calculado o histograma e os limites de intervalo de intensidades da imagem

original, e em seguida foi realizado um ajuste de intensidade. O resultado deste ajuste pode ser um clareamento ou um escurecimento na imagem. Neste trabalho foi realizado um clareamento na imagem. Como a região do olho é composta também por uma parte da cor da pele, foi preciso eliminar esta última. Para isto, a imagem é convertida para escala de cinza e, em seguida limiarizada, conforme:

$$L(x,y) = \begin{cases} 1 & \text{se } I(x,y) \geq T. \\ 0 & \text{se } I(x,y) < T. \end{cases} \tag{3.1}$$

sendo $I(x,y)$ o nível de cinza do *pixel* nas coordenadas (x,y) da imagem e T o limiar adotado para obtenção do valor binário $L(x,y)$. O resultado é uma imagem binária, uma matriz composta por 0's e 1's, onde os elementos com valor 1 correspondem aos *pixels* com valores acima do limiar T e os elementos com valor 0 aos *pixels* com valores abaixo do limiar T. A limiarização foi realizada com base na cor da pele por ser uniforme, ou seja, os *pixels* com valor 1 correspondem à região da pele, enquanto que os *pixels* com valor 0 correspondem à região do olho. Neste trabalho o limiar foi definido como $T = 0.5$. Este valor foi definido por validação experimental.

Segundo Gonzalez e Woods (2008) para determinar se dois *pixels* p e q estão conectados, verifica-se se eles têm alguma relação de adjacência e se seus níveis de cinza obedecem algum critério de similaridade. Normalmente existem duas maneiras para definir conectividade entre *pixels*: conectividade-de-4 e conectividade-de-8. Na conectividade-de-4 considera-se os quatro vizinhos horizontais e verticais do *pixels*, enquanto que na conectividade-de-8, consideram-se também os quatro vizinhos diagonais. Em relação ao critério de similaridade, normalmente é observado se *pixels* conectados-de-4 ou conectados-de-8 possuem o mesmo nível de cinza. Neste trabalho, procurou-se por um conjunto de objetos conectados-de-8.

Após os *pixels* conectados serem detectados na imagem, algumas propriedades foram obtidas, tais como o número de objetos conectados e a quantidade de *pixels* de cada região do objeto. Em seguida, procurou-se pelo objeto de maior área quando comparado com os outros objetos contido na imagem. A região de maior área foi considerada como

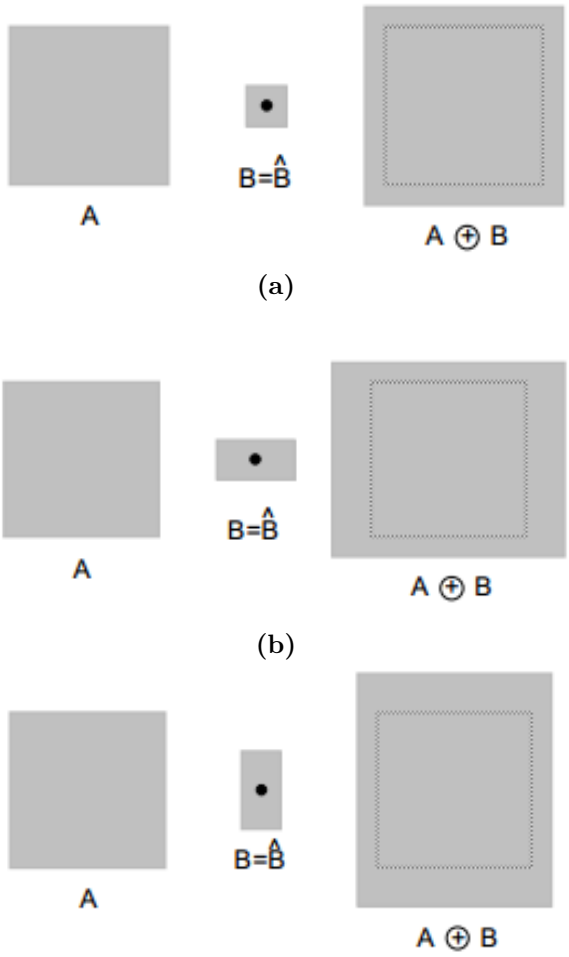


Figura 3.4: Dilatação. Fonte: (GONZALEZ; WOODS, 2008)

sendo o olho, enquanto que os demais objetos foram eliminados. Em seguida, foram aplicadas duas operações morfológicas: dilatação e preenchimento de lacunas. A dilatação é uma operação que "aumenta" objetos em uma imagem binária. Com A e B conjuntos de Z^2 (onde Z denota o conjunto dos números inteiros), a dilatação de A por B é denotada por $A \oplus B$, é definida como

$$A \oplus B = \{z|(\hat{B})_z \cap A \neq \varnothing\} \tag{3.2}$$

em que $\hat{B}$ é a reflexão de B em torno de sua origem e $(\hat{B})_z$ a translação dessa reflexão por z. O elemento estruturante B é representado por uma matriz binária de tamanho predefinido

cuja forma geométrica é representada pelos elementos não nulos [1] podendo assumir o formato de um quadrado, retângulo, círculo, entre outros. Então, a dilatação de A por B é o conjunto de todos os deslocamentos, z, de maneira que $\hat{B}$ e A se sobreponham pelo menos por um elemento. O elemento estruturante utilizado foi uma reta com tamanho de 10 *pixels*. A Figura 3.4 ilustra os efeitos da dilatação de um conjunto A usando três elementos estruturantes (B) distintos. Observa-se que as operações morfológicas são sempre referenciadas a um elemento do conjunto estruturante (neste caso, o elemento central).

Após a dilatação, algumas lacunas foram encontradas na imagem. Para preencher estas lacunas foi utilizado um algoritmo baseado em dilatação, complemento e intersecção de conjuntos. Seja A um conjunto cujos elementos são fronteiras[2] com conectividade-de-8, cada uma delas englobando uma região de fundo (ou seja, uma lacuna). Dado um ponto em cada lacuna, o objetivo é preencher todos estes com valor 1.

O algoritmo começa pela formação de um arranjo matricial, X_0 de 0s (o mesmo tamanho que o arranjo que contém A), exceto nas posições em X_0 correspondentes ao ponto dado em cada uma das lacunas, que foi definido como 1. Depois o procedimento a seguir preenche as lacunas com 1's.

$$X_n = (X_{n-1} \oplus B) \cap A^c \therefore n = 1, 2, 3 \ldots \tag{3.3}$$

sendo B o elemento estruturante e A^c o complemento de A. O algoritmo termina na interação n quando $X_n = X_{n-1}$. O conjunto X_n contém todas as lacunas preenchidas. A união de X_n e A contém todas as lacunas preenchidas e suas fronteiras. Este procedimento é ilustrado na Figura 3.5. Na parte (a) tem-se o conjunto original A, cujo complemento é mostrado em (b). A Figura 3.5c mostra o elemento estruturante utilizado. A parte (d) indica o resultado obtido após a sexta iteração (a última que ainda produziu alguma diferença em relação à iteração anterior), em que os números indicam que iteração con-

[1] Um pixel nulo é aquele cujo respectivo valor é igual a zero.

[2] Seja R um subconjunto de pixels numa imagem, R é chamado de uma região da imagem se R é um conjunto ligado. A fronteira da região R é o conjunto de pixels na região que tem um ou mais vizinhos que não se encontram em R (GONZALEZ; WOODS, 2008).

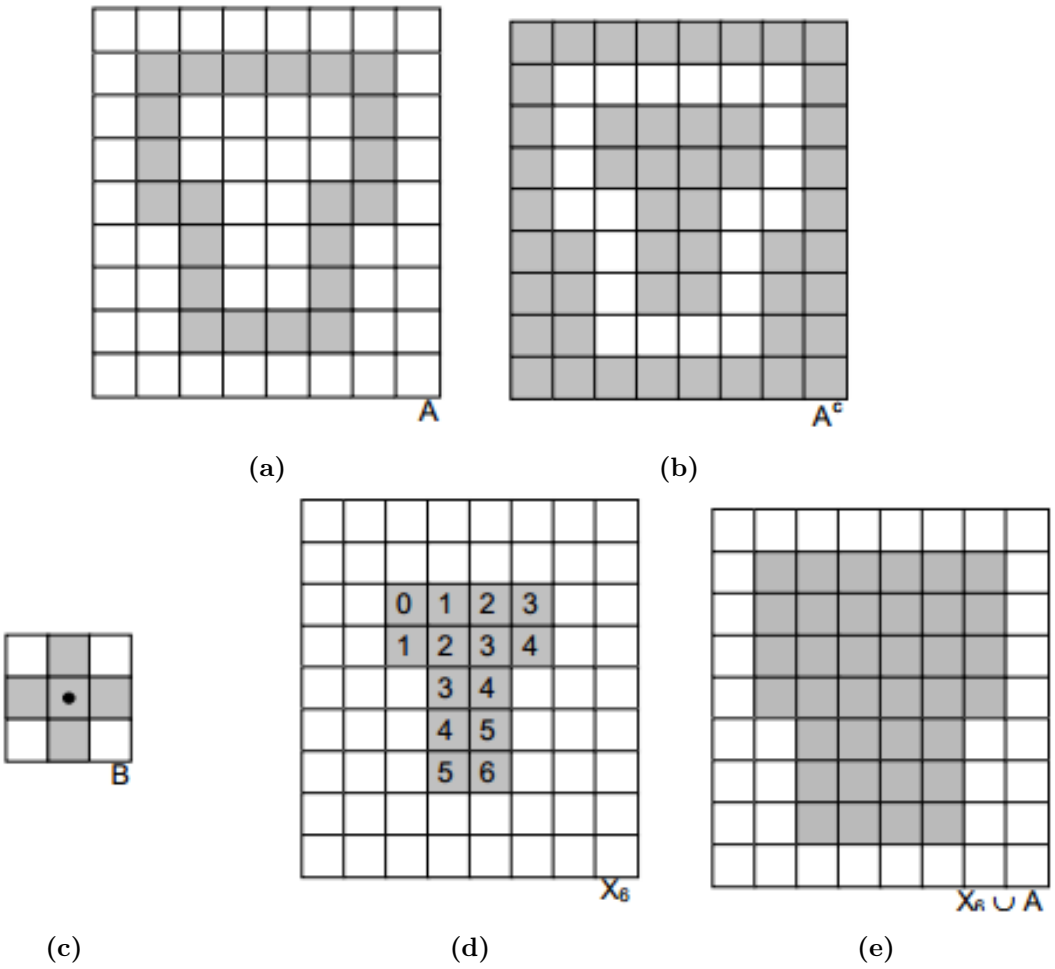


Figura 3.5: Preenchimento de lacunas. Fonte: (GONZALEZ; WOODS, 2008)

tribuiu para o surgimento de quais *pixels* no resultado parcial. Finalmente, o resultado da união do conjunto Figura 3.5d com o conjunto original é mostrado na parte (e).

Nas Figura 3.6, os passos propostos para a detecção do olho são ilustrados da esquerda para direita: olho original, ajuste de intensidade, limiarização, seleção da área de interesse, dilatação e preenchimento de lacunas.

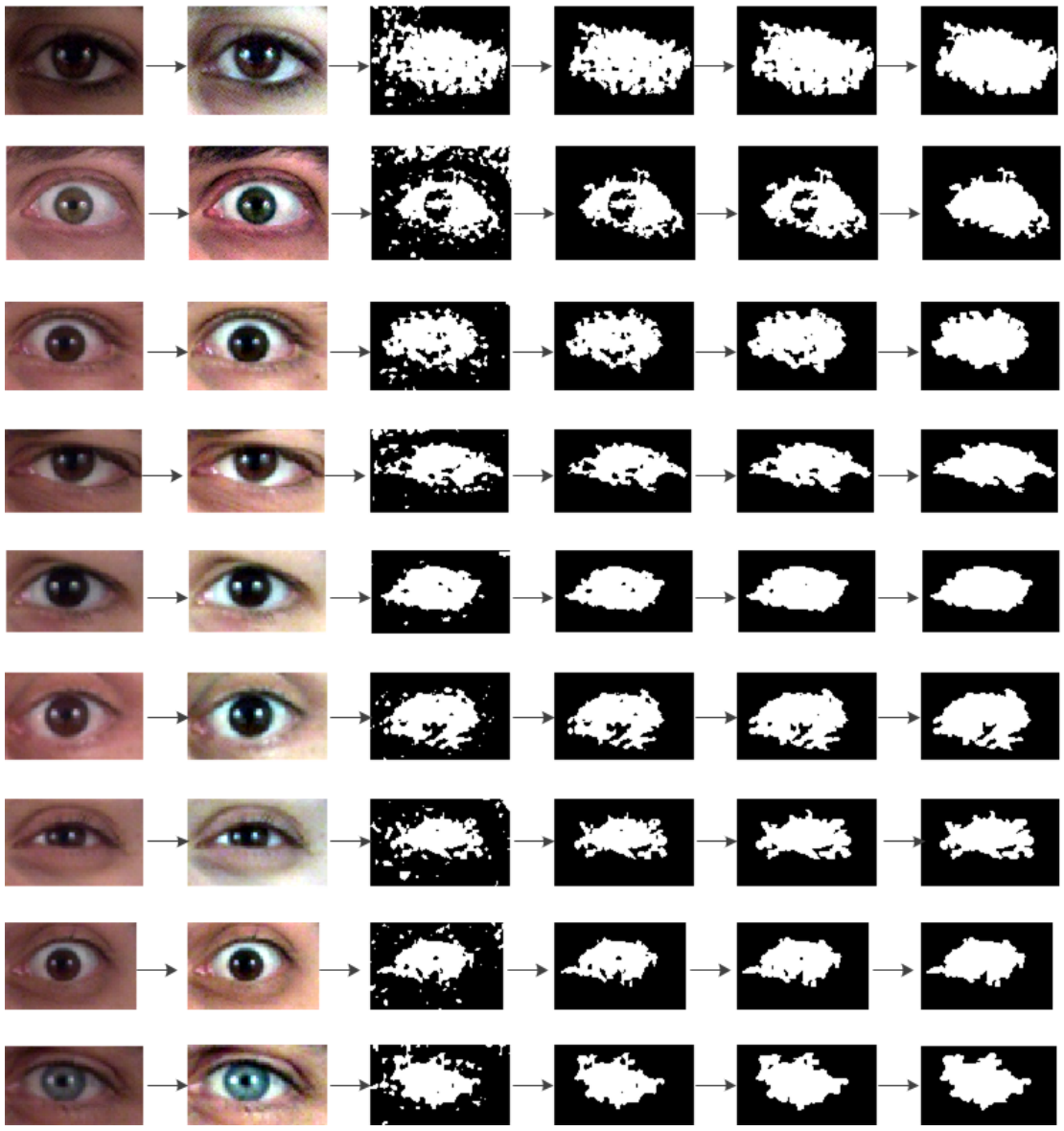


Figura 3.6: Detecção do olho.

3.4.2 Pré-processamento da Região da Sobrancelha

Após a região das sobrancelhas serem estimadas, a partir da região dos olhos, esta foi convertida para escala de cinza e o histograma foi equalizado utilizando a Eq. (3.4). A equalização de histograma consiste numa transformação $\Gamma(r_k)$ em que a imagem original resulte numa imagem onde os níveis de intensidade são uniformemente distribuídos. Normalmente para equalizar o histograma de uma determinada imagem, calcula-se inici-

almente o seu histograma original utilizando a seguinte expressão

$$p_r(r_j) = \frac{n_j}{n} \tag{3·4}$$

sendo $p_r(r_j)$ a função de distribuição de probabilidade ou uma função de distribuição de frequência do j-ésimo nível de cinza, n é o número total de *pixels* na imagem e n_j é o número de *pixels* cujo nível de cinza corresponde a j. Em seguida calcula-se o histograma acumulativo

$$s_k = \Gamma(r_k) = \sum_{j=0}^{k} p_r(r_j) \tag{3·5}$$

onde s_k é a função de distribuição acumulada, $0 \leq r_k \leq 1$ (nível de cinza normalizado) e $k = 0, 1, \ldots, C-1$ (C é o número de níveis de cinza). O resultado desta função gera uma escala probabilística uniformemente distribuída entre 0 e 1. Para converter os valores probabilísticos para valores em níveis de cinza utiliza-se

$$r_k = \Gamma^{-1}(s_k) \tag{3·6}$$

onde $0 \leq r_k, s_k \leq 1$. Como a geometria da região de interesse da sobrancelha é retangular, uma parte da cor da pele também está presente na imagem. Para eliminá-la, a imagem foi limiarizada utilizando a Eq. (3.1). O valor do limiar para esta região foi $T = 0.46$. Este valor foi determinado por validação experimental.

Depois da operação de limiarização a imagem foi dilatada. Assim como na região do olho o elemento estruturante utilizado na operação de dilatação foi uma reta com tamanho de 10 *pixels*. Em seguida a Eq.(3.2) foi utilizada para preencher os espaços vazios. Nas Figura 3.7, os passos propostos para a detecção da sobrancelha são ilustrados da esquerda para direita: sobrancelha original, sobrancelha convertida para escala de cinza, equalização de histograma, dilatação e preenchimento de lacunas.

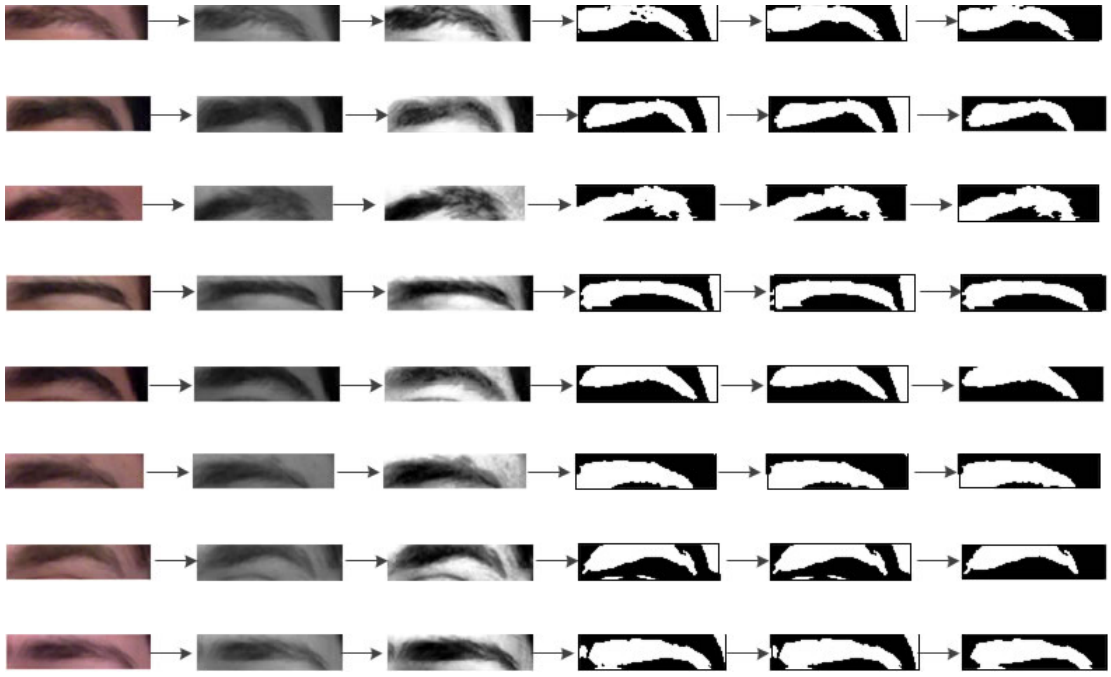


Figura 3.7: Detecção da sobrancelha.

3.4.3 Pré-processamento da Região da Boca

Para detectar características na região da boca é preciso levar em conta a variabilidade das formas que a boca pode apresentar. Com efeito, complexidades são adicionadas quando a boca está aberta ou os dentes são visíveis entre o lábio superior e inferior, devido ao sorriso ou qualquer outro tipo de expressão. Estas duas situações fornecem região escura e brilhante, respectivamente, no contorno da boca e faz com, que o processo de extração de características se torne bastante complexo (SAYEED et al., 2006).

Para lidar com os problemas citados acima, inicialmente foi aplicado um filtro gaussiano 2D que é basicamente uma operação de convolução, utilizada para suavizar uma imagem com o propósito de remover ruídos (NIXON; AGUADO, 2008). A imagem da boca inicialmente foi convertida para escala de cinza, porém como não apresentou resultados favoráveis, outro espaço de cor foi escolhido, o HSV (*Hue, Saturation and Value*) em que a cor é dividida em três componentes: *matiz* (H), *saturação* (S) e *valor* (V) (GONZALEZ; WOODS, 2008).

Em seguida, foi aplicada uma operação morfológica de abertura (erosão seguida

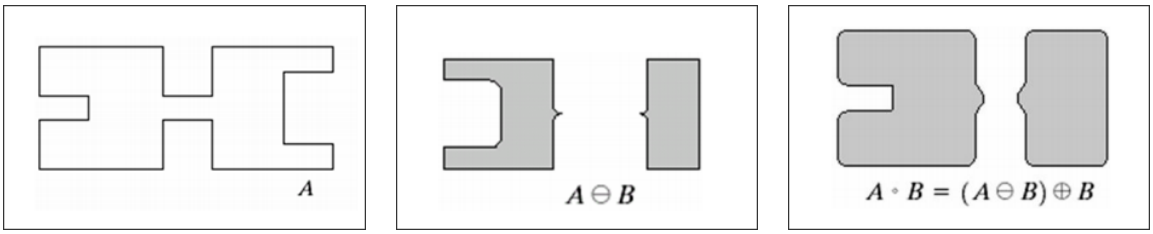


Figura 3.8: (a) Imagem original, (b) após erosão e (c) após a dilatação (abertura). O elemento estruturante utilizado foi em forma de disco. Fonte: Adaptado de (GONZALEZ; WOODS, 2008)

de dilatação) utilizando o mesmo elemento estruturante na componente H da imagem. A Figura 3.8 ilustra um exemplo da operação de abertura. A operação de abertura foi utilizada para abrir pequenos espaços vazios entre objetos próximos na imagem. Neste trabalho o elemento estruturante utilizado foi um disco de dimensão 5. A Abertura de um conjunto A por um elemento estruturante B, indicada por $A \circ B$ é definida como

$$A \circ B = (A \ominus B) \oplus B \tag{3.7}$$

Assim, a abertura de A por B é a erosão de A por B, seguida de uma dilatação do resultado por B. Em oposição à dilatação, a operação de erosão $\ominus$ "diminui" objetos em uma imagem binária. Após a aplicação das operações morfológicas, a imagem foi limiarizada utilizando a Eq. (3.1). O valor do limiar para região da boca foi $T = 0.5$. Em seguida foram encontrados na imagem os componentes conectados-de-8.

Depois que os componentes conectados foram detectados na imagem, algumas propriedades foram obtidas, tais como o número de objetos conectados e a quantidade de *pixels* de cada área do objeto. Em seguida, procurou-se pelo objeto de maior área quando comparado com os outros objetos contido na imagem. Neste caso, a região de maior área foi considerada como sendo boca, enquanto que os demais objetos foram eliminados. Nas Figura 3.9, os passos propostos para detecção da boca são ilustrados da esquerda para direita: boca original, aplicação do Filtro Gaussiano 2D, imagem convertida para espaço de cor HSV, limiarização e operação de abertura na componente H da imagem, dilatação e seleção da área de interesse.

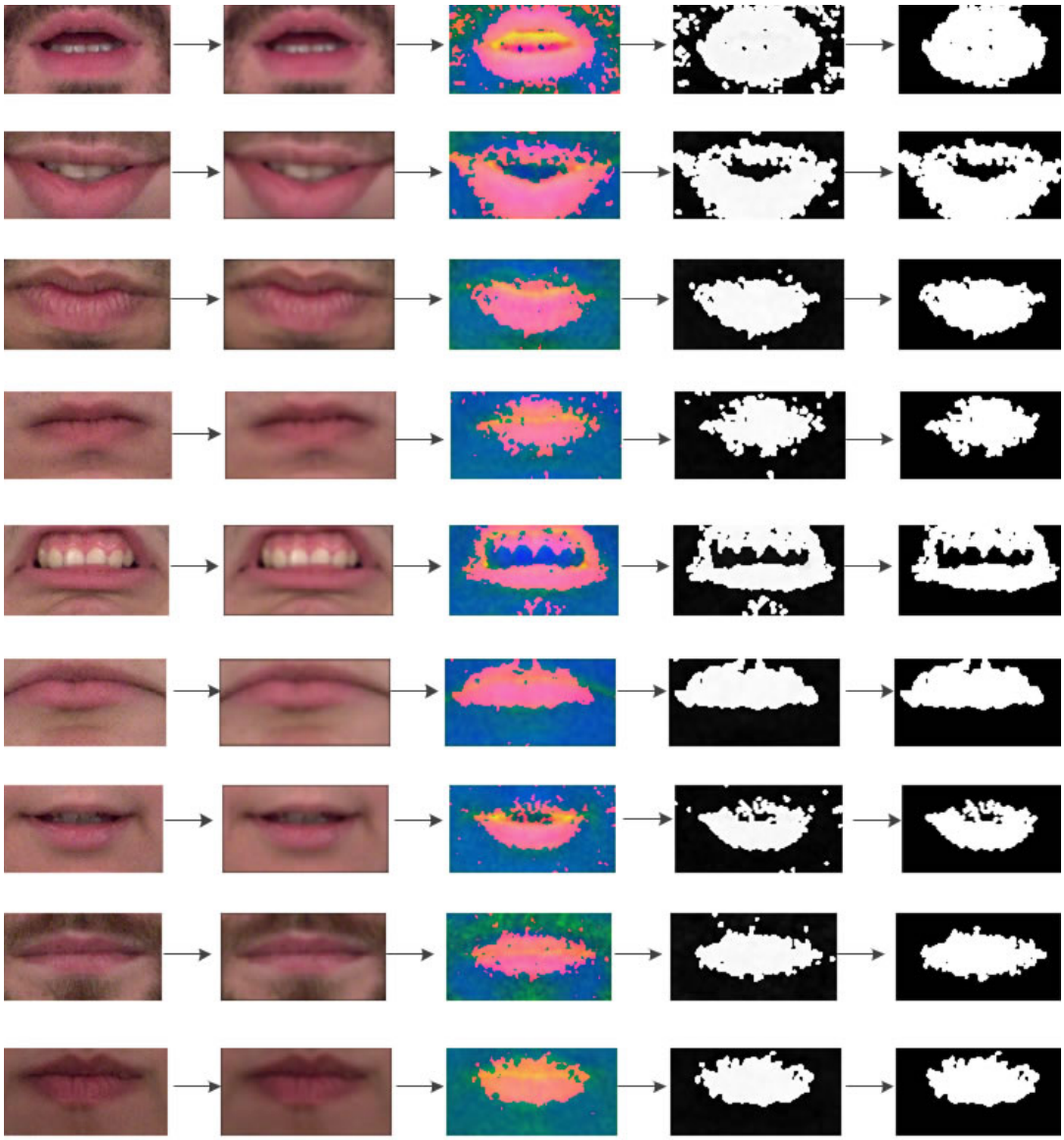
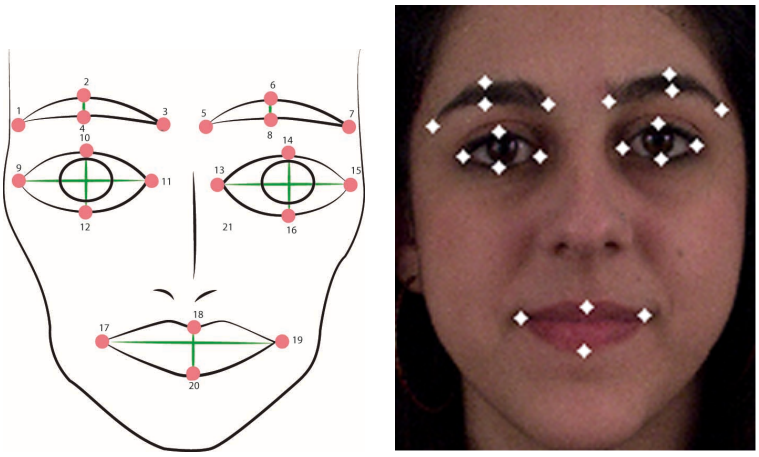


Figura 3.9: Detecção da boca.

3.4.4 Detecção dos Landmarks

Após serem aplicadas técnicas de processamento de imagens em cada uma das regiões faciais foi utilizado um detector de bordas para extrair os contornos. Dois dos principais métodos comumente utilizados são *Sobel* (GONZALEZ; WOODS, 2008) e *Canny* (CANNY, 1986). Este último método foi aplicado neste trabalho, pois através dos experimentos percebeu-se que este conseguiu não só localizar mais bordas, mas também mais detalhes da imagem. Para localizar os 20 *landmarks* nas regiões dos olhos, das sobrance-



(a) **(b)**

Figura 3.10: Detecção dos 20 *landmarks* faciais.

lhas e da boca, as bordas resultantes após aplicação do *Canny* são divididas em quatro partes iguais (Fig. 3.10 (a)).

Após as bordas de cada uma das regiões serem divididas, foram extraídos os *landmarks* que estavam localizados nas extremidades esquerda e direita da borda em cada uma das regiões detectadas. Foram localizados 4 *landmarks* para cada uma das regiões das sobrancelhas, olhos e boca, totalizando 20 *landmarks* detectados em torno da face. As Figuras 3.10 (a) e (b) ilustram os 20 *landmarks* detectados nas regiões das sobrancelhas, olhos e da boca.

3.5 CLASSIFICAÇÃO DE EXPRESSÕES

Antes de realizar a classificação das expressões, um método chamado *Generalized Procrustes Analysis* (GPA) desenvolvido por Gower (1975), e aprimorado por Berge (1977) foi aplicado às características extraídas. O GPA é um algoritmo estatístico que elimina os efeitos de escala, rotação e translação dos objetos, conseguindo quantificar as variações entre os *landmarks* correspondentes, buscando o ajuste para a melhor sobreposição das configurações geométricas (conforme ilustração na Figura 3.11). A configuração consiste em um conjunto de *landmarks* de um determinado objeto (DRYDEN; MARDIA, 1998). Depois da aplicação do método GPA, foram empregadas duas aborda-

gens para classificação: as baseadas em correspondência entre modelos e em redes neurais artificiais. No método baseado em correspondência entre modelos, o GPA também foi utilizado para estimar o modelo médio para cada uma das sete expressões definidas neste trabalho. Uma descrição do método GPA é realizada na subseção 3.5.1. A seguir, as abordagens de classificação utilizadas no presente trabalho são descritas.

3.5.1 Classificação Baseada em Correspondência entre Modelos

Na primeira etapa, é criado um modelo médio para cada uma das seguintes expressões: felicidade, raiva, tristeza, surpresa, desgosto, medo e neutra. O modelo médio é estimado através do método estatístico de *Generalized Procrustes Analysis* desenvolvido por Gower (1975), e aprimorado por Berge (1977), conforme ilustração na Figura 3.12. Na segunda etapa é calculado o grau de similaridade entre os modelos estimados utilizando distância de *Procrustes* com o intuito de classificar as expressões.

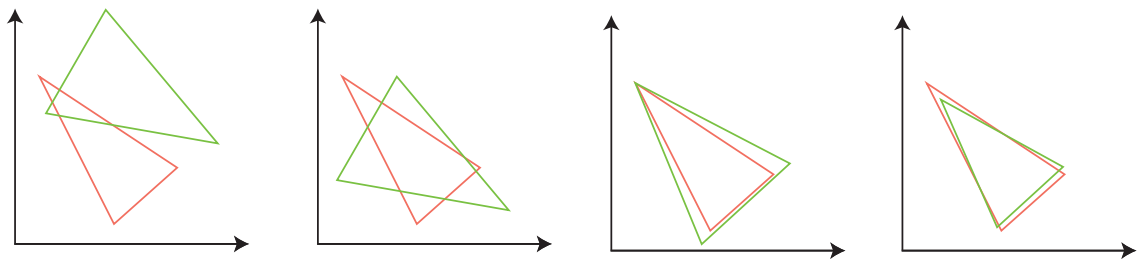
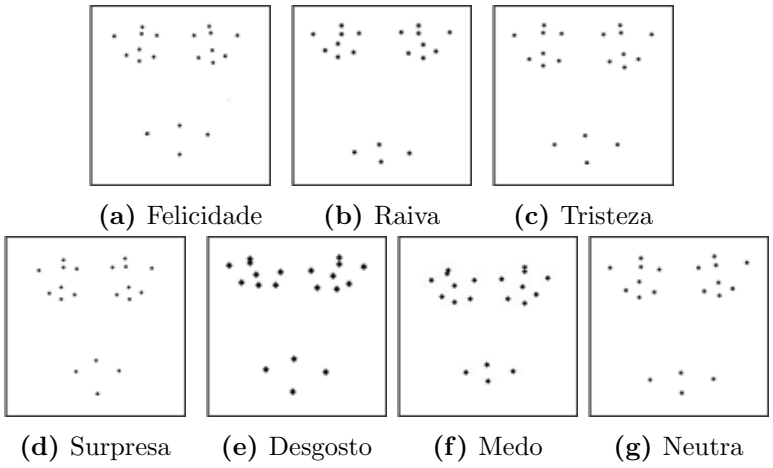


Figura 3.11: Análise de *Procrustes* da esquerda para a direita: Configuração inicial, translação, rotação e escala.

3.5.1.1 Generalized Procrustes Analysis

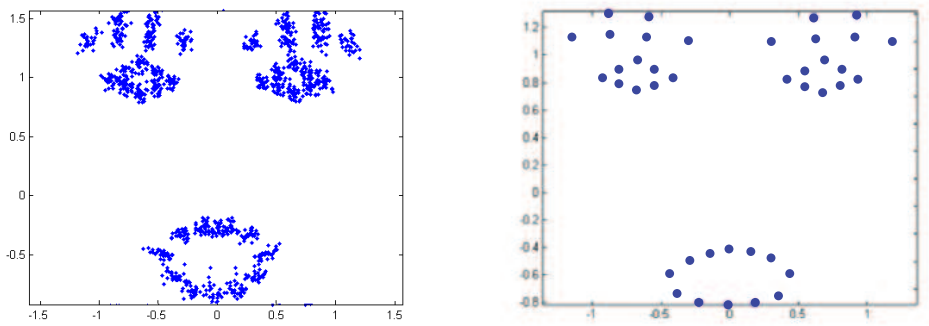
O *Generalized Procrustes Analysis* (GPA) baseia-se em alinhar sequencialmente pares de formas utilizando uma forma de referência (a forma média) e alinha as outras formas para a forma média. Inicialmente uma forma aleatória do conjunto é escolhida para ser a forma média. Após o alinhamento (ver ilustração na Figura 3.11), uma nova estimativa para a forma média é calculada, e novamente, as formas são alinhados para a forma média. Este procedimento é realizado repetidamente até que a forma média não



(a) Felicidade **(b)** Raiva **(c)** Tristeza

(d) Surpresa **(e)** Desgosto **(f)** Medo **(g)** Neutra

Figura 3.12: Modelo médio de sete expressões faciais.



(a) **(b)**

Figura 3.13: (a) Exemplo de dados brutos extraídos de uma sequência de imagem de uma única expressão e (b) a forma média após a aplicação do GPA.

mude significativamente durante as interações (BOOKSTEIN, 1996). A Figura 3.13 mostra os resultados deste procedimento de alinhamento. O lado esquerdo da imagem mostra exemplo de dados brutos extraídos de uma sequência de imagem de uma única expressão e o lado direito corresponde a forma média após do GPA.

Após os modelos médios serem estimados, a técnica de distância de *Procrustes* é utilizada para calcular o grau de similaridade entre a forma média (de cada uma das expressões) e a forma obtida da imagem de entrada. O modelo que apresentar a menor distância de *Procrustes* é considerado a expressão apresentada na imagem. O valor da distância de *Procrustes* varia entre 0 e 1, ou seja, quando mais próximo de 0 mais similar é a forma comparada. O método de distância *Procrustes* realiza uma transformação linear

em uma forma para encontrar a melhor correspondência entre duas formas. O objetivo deste método é encontrar o grau de similaridade entre duas formas (DRYDEN; MARDIA, 1998).

3.5.2 Classificação Baseada em Redes Neurais Artificiais

Para realizar a classificação através de redes neurais artificiais, utilizou-a rede *Multi-Layer Perceptrons* (MPL) do tipo *feed forward*. Tipicamente uma rede multicamadas (*Multilayer*) é constituída por uma camada de entrada, uma camada de saída e uma ou mais camadas ocultas (HAYKIN, 1998).

Neste trabalho a camada de entrada possui 40 neurônios (sendo oito parâmetros da boca, oito para cada olho e oito para cada sobrancelha). Para a camada oculta foram escolhidos de 10 a 16 neurônios com intuito de avaliar o desempenho do classificador. A camada de saída possui 7 neurônios, um para cada classe de expressão (felicidade, raiva, tristeza, surpresa, desgosto, medo e neutra). O algoritmo de treinamento utilizado foi CGP (*Conjugate gradient backpropagation with Polak-Ribiére updates*) proposto por Polak e Ribiere (1969). Nos experimentos preliminaresf realizados, este algoritmo obteve melhor desempenho que o tradicional *backpropagation*. Todos os neurônios foram configurados com a função de ativação sigmóide (optou-se por esta função, devido ser a mais comumente utilizada).

CAPÍTULO 4

RESULTADOS EXPERIMENTAIS

Neste capítulo são apresentados os resultados experimentais do sistema proposto. Para isto, foram realizados alguns experimentos com o intuito de obter as taxas de detecção da face e das regiões faciais, a precisão das características extraídas, e por fim as taxas de reconhecimento para as duas abordagens de classificação utilizadas neste trabalho. Além disso, as ferramentas computacionais e as bases de dados utilizadas para testar e treinar o sistema são descritas.

4.1 FERRAMENTAS COMPUTACIONAIS UTILIZADAS

O sistema de reconhecimento de expressões faciais desenvolvido neste trabalho tem como característica a portabilidade do código, o que possibilita a sua execução em várias plataformas computacionais, com diferentes sistemas operacionais. Os algoritmos foram desenvolvidos em (MATrix LABoratory) e C/C++ utilizando a biblioteca OpenCV. O MATLAB é uma ferramenta computacional com inúmeras funções de análise numérica, matemática computacional, ferramentas de engenharia, etc. Esta ferramenta foi utilizada para o desenvolvimento dos módulos de detecção das regiões faciais (sobrancelhas), extração de características e classificação de expressões. Já o OpenCV é uma biblioteca de funções que implementam alguns dos algoritmos mais usuais no domínio da visão computacional. O algoritmo utilizado no presente trabalho para detectar a face e das regiões faciais (olhos e sobrancelhas) encontra-se disponível na biblioteca do OpenCV. O sistema foi treinado e testado em uma plataforma Intel Core i5 450M, com 4GB de memória e um disco rígido de 500GB.

4.2 BASES DE DADOS

Neste trabalho, foram utilizadas duas base de dados públicas: MUG *Facial Expression* (AIFANTI et al., 2010) e *Face and Gesture Recognition Research Network* (FG-NET) (WALLHOFF, 2006), as quais foram escolhidas devido a apresentarem *expressões básicas universais* já rotuladas, além de exibir imagens coloridas, pois uma das etapas do sistema proposto depende dos canais de cores para extração das características faciais. A base MUG *Facial Expression* foi utilizada para validar as etapas de detecção da face e das regiões faciais e da extração de características. Para avaliar a etapa de classificação foram utilizadas a MUG *Facial Expression* e FG-NET. Cada uma das bases de dados são descritas a seguir:

MUG Facial Expression

A base de dados MUG Facial *Expression* é composto por sequências de imagens de 86 indivíduos que executam diferentes expressões faciais. As imagens possuem uma resolução de 896x896 *pixels* e iluminação uniforme. Dos 86 indivíduos, 35 são mulheres e 51 homens, todos de origem caucasiana entre 20 e 35 anos de idade. Alguns homens apresentam barbas e não existem oclusões (exceto uma parte do cabelo sobre a face). Porém dos 86 indivíduos, apenas imagens de 52 indivíduos são acessíveis para usuários autorizados, sendo que imagens de 25 indivíduos estão disponíveis mediante solicitação e 9 estão disponíveis apenas ao laboratório MUG. Esta base de dados consiste de duas partes. Na primeira parte, os participantes foram orientados a realizar as *expressões básicas universais* que são: felicidade, raiva, tristeza, surpresa, desgosto, medo; incluindo a expressão neutra. A segunda parte é composta emoções laboratoriais induzidas.

Na primeira parte, os participantes imitaram corretamente as *expressões básicas universais*, sendo que eles foram informados sobre como as expressões faciais são realizadas, conforme o manual do FACS. Para todas as expressões foram capturadas uma sequência de imagens, contendo entre 50-160 imagens. Além disso, estão publicamente disponíveis 80 *landmarks* faciais de 401 imagens de 26 indivíduos que foram anotados

manualmente. Estão disponíveis também para pesquisadores 80 *landmarks* faciais para 821 imagens de 28 indivíduos que foram anotadas utilizando o *Active Appearance Models*.

Na segunda parte da base de dados, as expressões dos indivíduos foram capturadas quando estavam assistindo um vídeo indutor de emoções. Os participantes estavam cientes de que estavam sendo filmados e exibiam várias emoções e atitudes emocionais. Estas incluem (além das emoções básicas), mas não estão limitados a: aborrecimento, diversão, entretenimento, amizade, simpatia, interesse e entusiasmo. Entretanto, como esta segunda parte da base de dados ainda não foi rotulada, neste trabalho foi utilizada apenas a primeira parte desta base de dados para treinar e testar o sistema proposto. Exemplos de imagens de faces encontradas na primeira parte da base são ilustradas na Figura 4.1.

Figura 4.1: Exemplos de imagens da base de dados MUG *Facial Expression* da esquerda para direita: felicidade, raiva, tristeza, surpresa, desgosto, medo e neutra.

Face and Gesture Recognition Research Network (FG-NET)

A base de dados FG-NET com expressões faciais e emoções da Universidade *Technical Munich* é uma base de dados contendo imagens de face que mostram 18 diferentes indivíduos (9 do sexo feminino e 9 do sexo masculino) com idades entre 23 e 38 anos, desempenhando as sete *expressões básicas universais* definidas por Ekman e Friesen (1971), incluindo a face neutra. Um dos paradigmas desta base de dados é que os participantes manifestassem as expressões da maneira mais natural possível. Assim, os participantes foram submetidos a vídeos de clipes de curta duração ou imagens (momento em que as expressões foram capturadas). Estas expressões incluem: felicidade, raiva, tristeza, surpresa, desgosto, medo e neutra. Cada indivíduo realizou todas as expressões desejadas três vezes, totalizando uma quantidade de 399 sequências. As sequências de vídeos de curta duração foram filmadas com resolução de 640x480 *pixels*. Os vídeos foram convertidos para imagens com tamanho de 320x240 *pixels*. Algumas imagens de faces encontradas nesta base são mostradas na Figura 4.2.

Figura 4.2: Exemplos de imagens da base de dados FG-NET da esquerda para direita: felicidade, raiva, tristeza, surpresa, desgosto, medo e neutra.

4.3 DETECÇÃO DA FACE E DAS REGIÕES FACIAIS

Para avaliar os módulos de detecção da face e das regiões faciais utilizou-se um subconjunto de 401 imagens de 26 indivíduos da base de dados MUG Facial *Expression*. As taxas de detecção foram obtidas manualmente, através da contagem de todos os acertos e erros apresentados pelo algoritmo. Em seguida, foram calculadas as porcentagens de acertos para cada uma das 401 imagens da face e das regiões faciais (os olhos, as sobrancelhas e a boca) avaliadas. O algoritmo detectou 100% das faces apresentadas. Na Tabela 4.1 é apresentada a taxa de detecção para cada uma das regiões faciais detectadas.

Tabela 4.1: Taxa de detecção das regiões faciais.

Regiões Faciais	**MUG Facial Expression**
Boca	100%
Olho Direito	98%
Olho Esquerdo	98%
Sobrancelha Direita	98%
Sobrancelha Esquerda	98%

4.4 EXTRAÇÃO DE CARACTERÍSTICAS

Para avaliar a precisão do algoritmo desenvolvido na etapa de extração de características, foram realizados testes quanto à similaridade das formas das regiões faciais utilizando as anotações da base de dados MUG *Facial Expression*. Assim como nos módulos de detecção da face e das regiões faciais, foi utilizado um subconjunto de 401 imagens de 26 indivíduos da base MUG para validar esta etapa do sistema. Para analisar a similaridade de um conjunto de formas detectadas neste trabalho, foi calculada uma distância de *Procrustes* entre as anotações da imagem *(ground truth)* e os *landmarks* detectados pelo *Active Shape Model* (ASM) implementado por Milborrow e Nicolls (2008) e o método de extração de características apresentado no presente trabalho. Na Tabela 4.2, é possível visualizar os resultados de similaridade entre as formas detectados pelo método proposto e pela técnica ASM a partir da base de dados MUG Facial *Expresssion*. O valor de similaridade varia entre 0 e 1. Quando mais próximo de 0, mais similar é a forma

comparada. A taxa de similaridade apresentada na Tabela 4.2, é dada por $T_p = 1 - D_p$, sendo D_p a distância de *Procrustes* encontrada.

Tabela 4.2: Taxa de similaridade das regiões faciais.

Regiões Faciais	MUG Facial Expression	ASM
Boca	95%	98%
Olho Direito	95%	98%
Olho Esquerdo	94%	98%
Sobrancelha Direita	95%	98%
Sobrancelha Esquerda	90%	98%

Para avaliar graficamente o grau de similaridade entre todas as formas das regiões facias, foi utilizada a distribuição acumulativa de similaridade entre as formas detectadas. Para isto, utilizou-se a Equação (4.1):

$$P(x \leq d|\mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} \sum_{x<d} e^{-\frac{1}{2}\frac{(d-\mu)^2}{\sigma^2}} \tag{4.1}$$

sendo d o grau de similaridade entre as formas detectadas, μ a média e σ o desvio padrão do grau de similaridade do conjunto das formas detectadas.

Nas Figuras 4.3 - 4.7 é possível visualizar a distribuição acumulativa de similaridade. Apesar do método ASM ser considerado bastante preciso, o método de extração de características apresentado neste trabalho alcançou taxas de similaridades próximas ao obtido pelo método ASM.

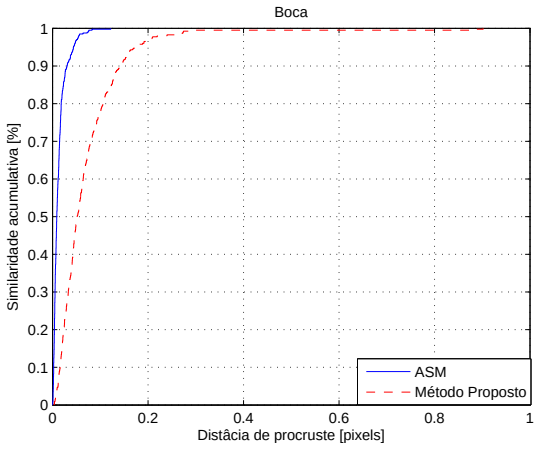


Figura 4.3: Distribuição acumulativa de similaridade da forma da boca.

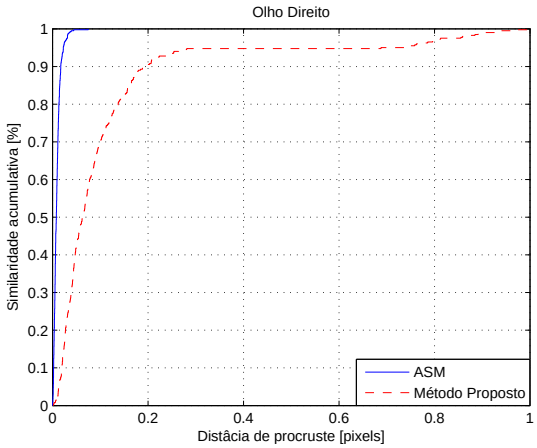


Figura 4.4: Distribuição acumulativa de similaridade da forma do olho direito.

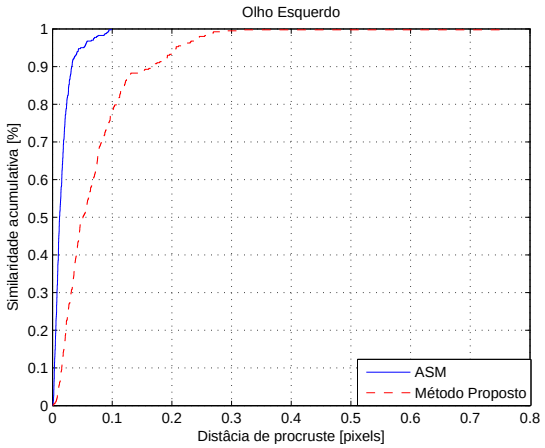


Figura 4.5: Distribuição acumulativa de similaridade da forma do olho esquerdo.

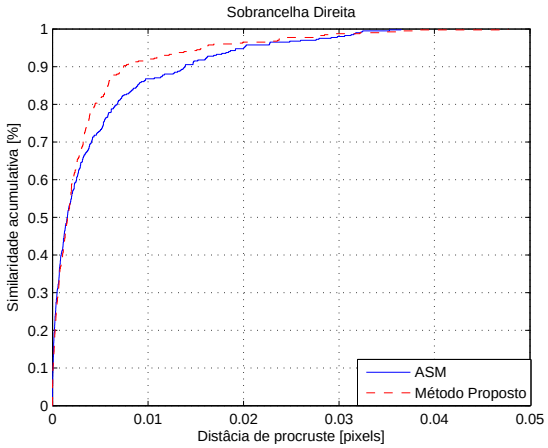


Figura 4.6: Distribuição acumulativa de similaridade da forma da sobrancelha direita.

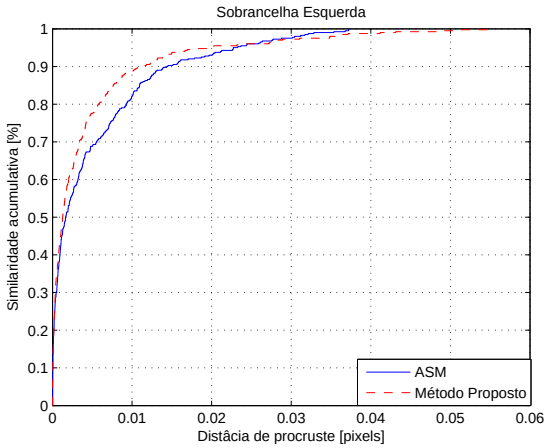


Figura 4.7: Distribuição acumulativa de similaridade da forma da sobrancelha esquerda.

4.5 CLASSIFICAÇÃO DE EXPRESSÕES

Como visto no Capítulo 3 (Seção 3.5), foram utilizadas duas abordagens diferentes na etapa de classificação das expressões. Na primeira abordagem as expressões faciais são classificadas através de redes neurais artificiais e a segunda foi baseada em correspondência entre modelos. Duas bases de dados foram utilizadas para validar a etapa de classificação a MUG *Facial Expression* e FG-NET.

O sistema desenvolvido foi testado e treinado com cada um dos bancos de dados separadamente. A base FG-NET apresenta o mesmo indivíduo executando a mesma expressão 3 vezes, porém apesar de ser a mesma expressão, esta apresenta variações. Já para a base de dados MUG Facial *Expression* existem várias imagens para a mesma expressão de um mesmo individuo, porém estas não variam. Sendo assim, da base de dados MUG Facial *Expression* foi escolhida uma amostra por expressão para cada indivíduo, enquanto que para a base de dados FG-NET foram escolhidas três amostras para cada expressão de um mesmo indivíduo. Da base MUG Facial *Expression* foram utilizadas 210 imagens de 30 (11 do sexo feminino e 19 do sexo masculino) indivíduos. As imagens utilizadas foram apenas as rotulada, ou seja, aquelas em que os participantes imitaram corretamente as *expressões básicas universais* (ver seção 4.2). Da base de dados FG-NET foram selecionadas 378 imagens de 18 indivíduos (9 do sexo feminino e 9 do sexo masculino, sendo três

amostras para cada indviduo). Nas duas bases de dados foram selecionadas as expressões de felicidade, raiva, tristeza, surpresa, desgosto, medo e neutra.

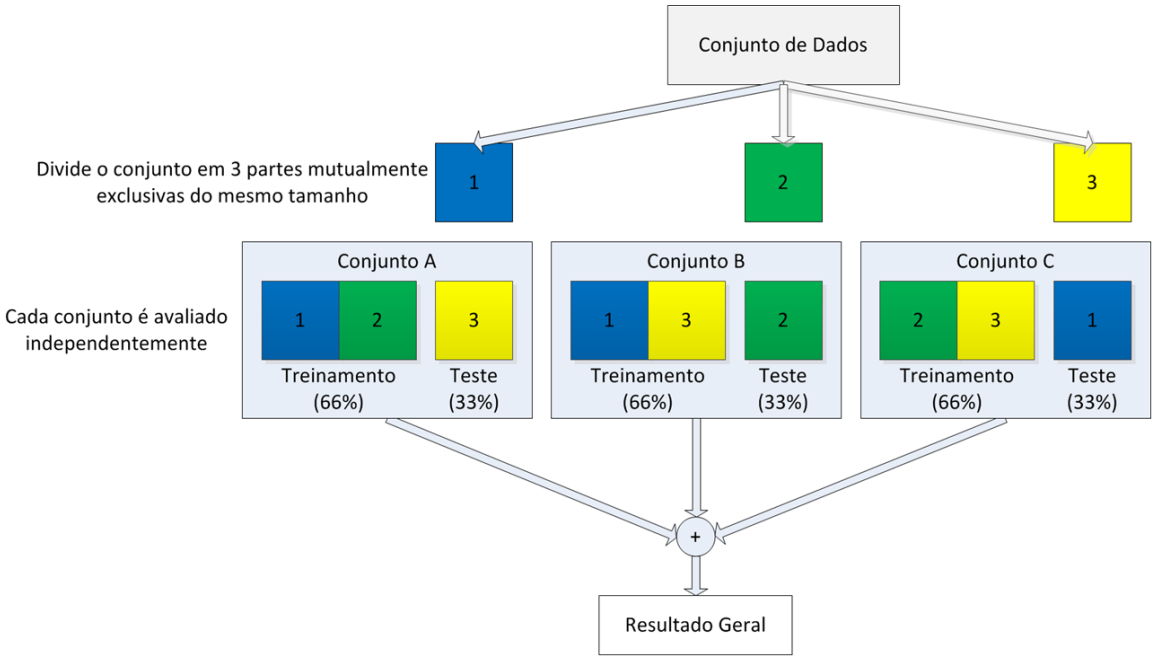


Figura 4.8: Metodologia para avaliar a etapa de classificação.

A metodologia utilizada para avaliar a etapa de classificação consiste em inicialmente formar três conjuntos de imagens. Sendo que cada conjunto foi composto por um total de 30 imagens para cada uma das sete classes de expressões (felicidade, raiva, tristeza, surpresa, desgosto, medo e neutra) para a base *MUG Facial Expression* e 54 imagens para as sete expressões para a FG-NET. Na sequencia, dois terços das imagens foram selecionadas para compor o conjunto de treinamento e um terço para compor o conjunto de teste. Esta metodologia é chamada de validação cruzada. A Figura Figura 4.8 ilustra a metodologia para avaliar a etapa de classificação do presente trabalho.

Os três conjuntos foram representados pelas letras A, B e C e foram utilizados para treinar e avaliar o desempenho das abordagens baseadas em redes neurais artificiais e em correspondência entre modelos utilizadas neste trabalho para classificar sete diferentes expressões. Nas próximas subseções são apresentados os resultados obtidos para as duas abordagens utilizando a metodologia descrita anteriormente.

4.5.1 Classificação Baseada em Redes Neurais Artificiais

Para a abordagem baseada em redes neurais artificiais, foi realizada uma avaliação quanto à melhor configuração em relação à quantidade de neurônios na camada oculta. Após várias inicializações aleatórias dos pesos da rede neural artificial, as configurações que apresentaram melhor desempenho foram com dez e treze neurônios na camada oculta utilizando funções de ativação do tipo sigmóide e com o algoritmo de aprendizado CGP (*Conjugate gradient backpropagation with Polak-Ribiére updates*). Analisando os resultados das Tabelas 4.3 e 4.4, percebe-se que a taxa de acerto variou de forma expressiva quando a pequenas variações na quantidade de neurônio da camada oculta. Esta ocorrência pode ser explicada pela existência de mínimos locais durante o treinamento da rede. Também se percebe que o desempenho variou de acordo com as bases de dados utilizadas. Através da Tabela 4.3 pode-se observar que para a base de dados MUG *Facial Expression* o sistema apresentou melhor precisão quando foram utilizados dez neurônios na camada oculta, enquanto que para a base FG-NET treze neurônios foram necessários, conforme pode ser visto na Tabela 4.4. Sendo assim, estas configurações foram utilizadas para classificar as expressões neste trabalho.

Os resultados do reconhecimento foram apresentados através da matriz de confusão, a qual mostra a taxa de acerto para cada uma das expressões reconhecidas. As taxas de acertos para cada uma das expressões faciais por conjunto foram obtidas através do seguinte calculo: (quantidade de exemplos da base de dados classificadas corretamente / quantidade total de exemplos)*100. Observa-se que ao utilizar a base de dados MUG *Expression Facial* para o conjunto A, a melhor taxa de acerto apresentada pelo sistema foi de 100% (felicidade, neutra, tristeza, raiva, desgosto e medo), e a pior foi de 90% (surpresa). Para o conjunto B, a melhor taxa foi de 100% (felicidade, neutra, tristeza, raiva e medo) e a pior 90% para a expressão surpresa e desgosto. No caso do conjunto C, a melhor taxa foi 100% (felicidade, neutra, surpresa, raiva, desgosto e medo) e a pior 80% para a expressão de tristeza. A taxa média de acerto foi 97,62%.

Tabela 4.3: Avaliação de desempenho da rede neural artificial utilizando a base MUG. A Tabela abaixo apresenta a taxa de acerto (%) para cada conjunto de teste e sua respectiva média.

	Quantidade de Neurônios da Camada Oculta						
Conjunto	**10**	**11**	**12**	**13**	**14**	**15**	**16**
A	98,57	28,57	97,14	84,29	64,29	98,57	28,57
B	97,14	52,86	97,14	81,43	14,29	95,71	41,43
C	97,14	67,14	95,71	80,00	97,14	94,29	94,29
Média	97,62	49,52	96,67	81,90	58,57	96,19	54,76

Predições

Expressões	Neutra	Felicidade	Surpresa	Tristeza	Raiva	Desgosto	Medo
Neutra (10)	10	0	0	0	0	0	0
Felicidade (10)	0	10	0	0	0	0	0
Surpresa (10)	0	0	9	1	0	0	0
Tristeza (10)	0	0	0	10	0	0	0
Raiva (10)	0	0	0	0	10	0	0
Desgosto (10)	0	0	0	0	0	10	0
Medo (10)	0	0	0	0	0	0	10

Conjunto de Teste (A) - Taxa de acerto (98,57%)

Predições

Expressões	Neutra	Felicidade	Surpresa	Tristeza	Raiva	Desgosto	Medo
Neutra (10)	10	0	0	0	0	0	0
Felicidade (10)	0	10	0	0	0	0	0
Surpresa (10)	0	0	9	1	0	0	0
Tristeza (10)	0	0	0	10	0	0	0
Raiva (10)	0	0	0	0	10	0	0
Desgosto (10)	0	0	0	1	0	9	0
Medo (10)	0	0	0	0	0	0	10

Conjunto de Teste (B) - Taxa de acerto (97,14%)

Predições

Expressões	Neutra	Felicidade	Surpresa	Tristeza	Raiva	Desgosto	Medo
Neutra (10)	10	0	0	0	0	0	0
Felicidade (10)	0	10	0	0	0	0	0
Surpresa (10)	0	0	10	0	0	0	0
Tristeza (10)	0	0	2	8	0	0	0
Raiva (10)	0	0	0	0	10	0	0
Desgosto (10)	0	0	0	0	0	10	0
Medo (10)	0	0	0	0	0	0	10

Conjunto de Teste (C) - Taxa de acerto (97,14%)

Predições

Expressões	Neutra	Felicidade	Surpresa	Tristeza	Raiva	Desgosto	Medo
Felicidade (30)	0	30	0	0	0	0	0
Surpresa (30)	0	0	28	2	0	0	0
Tristeza (30)	0	0	2	28	0	0	0
Raiva (30)	0	0	0	0	30	0	0
Desgosto (30)	0	0	0	1	0	29	0
Medo (30)	0	0	0	0	0	0	30

Conjunto Geral de Teste - Taxa de acerto (97,62%)

Figura 4.9: Matrizes de confusão do classificador baseado em redes neurais artificiais utilizando a base MUG.

Para a base de dados FG-NET, no conjunto de teste A, a melhor taxa de reconhecimento foi de 94,44% para neutra, e a pior foi de 72,22% para a expressão de desgosto. No conjunto de teste B, a melhor taxa foi 100% (neutra e surpresa), e a pior 55,55% para a expressão de felicidade. No conjunto C a melhor taxa de reconhecimento foi de 100% para a expressão de tristeza e a pior foi de 66,66% (felicidade e medo). A taxa média de acerto foi 86,50%.

Tabela 4.4: Avaliação de desempenho da rede neural artificial utilizando a base FG-NET. A Tabela abaixo apresenta a taxa de acerto (%) para cada conjunto de teste e sua respectiva média.

	Quantidade de Neurônios da Camada Oculta						
Conjunto	**10**	**11**	**12**	**13**	**14**	**15**	**16**
A	72,22	74,60	34,13	85,71	14,29	82,54	61,90
B	76,98	69,05	84,92	86,51	58,73	84,13	85,71
C	76,98	53,97	74,60	87,30	42,06	69,05	84,92
Média	75,40	65,87	64,55	86,51	38,36	78,57	77,51

Predições

Expressões	Neutra	Felicidade	Surpresa	Tristeza	Raiva	Desgosto	Medo
Neutra (18)	17	0	0	0	1	0	0
Felicidade (18)	0	15	0	0	0	2	1
Surpresa (18)	0	1	16	0	1	0	0
Tristeza (18)	2	0	0	16	0	0	0
Raiva (18)	0	1	0	0	17	0	0
Desgosto (18)	0	1	0	0	0	13	4
Medo (18)	0	3	0	0	0	1	14

Conjunto de Teste (A) - Taxa de acerto (85,71%)

Predições

Expressões	Neutra	Felicidade	Surpresa	Tristeza	Raiva	Desgosto	Medo
Neutra (18)	18	0	0	0	0	0	0
Felicidade (18)	0	10	0	0	0	0	8
Surpresa (18)	0	0	18	0	0	0	0
Tristeza (18)	1	0	0	17	0	0	0
Raiva (18)	1	0	0	0	17	0	0
Desgosto (18)	0	0	0	0	0	17	1
Medo (18)	0	4	0	0	0	2	12

Conjunto de Teste (B) - Taxa de acerto (86,50%)

Predições

Expressões	Neutra	Felicidade	Surpresa	Tristeza	Raiva	Desgosto	Medo
Neutra (18)	17	0	0	1	0	0	0
Felicidade (18)	0	12	0	0	0	1	5
Surpresa (18)	0	1	17	0	0	0	0
Tristeza (18)	0	0	0	18	0	0	0
Raiva (18)	0	0	1	0	17	0	0
Desgosto (18)	0	1	0	0	0	17	0
Medo (18)	0	4	0	0	0	2	12

Conjunto de Teste (C) - Taxa de acerto (87,30%)

Predições

Expressões	Neutra	Felicidade	Surpresa	Tristeza	Raiva	Desgosto	Medo
Neutra (54)	52	0	0	1	1	0	0
Felicidade (54)	0	37	0	0	0	3	14
Surpresa (54)	0	2	51	0	1	0	0
Tristeza (54)	3	0	0	51	0	0	0
Raiva (54)	1	1	1	0	51	0	0
Desgosto (54)	0	2	0	0	0	47	5
Medo (54)	0	11	0	0	0	5	38

Conjunto Geral de Teste - Taxa de acerto (86,50%)

Figura 4.10: Matrizes de confusão do classificador baseado em redes neurais artificiais utilizando a base FG-NET.

4.5.2 Classificação Baseada em Correspondência entre Modelos

Para a classificação baseada em correspondência entre modelos, foram realizados testes em relação à taxa média de acerto entre 3 a 7 expressões, e através dos resultados experimentais, foi possível observar que a taxa de acerto deste método diminuiu à medida que aumentou-se a quantidade de expressões. Por exemplo, para as expressões: felicidade, neutra e surpresa o sistema atingiu uma taxa média de acerto de 97,78%. Para

4 expressões (felicidade, neutra, surpresa e tristeza) a média foi de 73,33% de acertos, enquanto que para 5 expressões (felicidade, neutra, surpresa e tristeza e raiva) a média foi de 69,33%, e para 6 expressões (felicidade, neutra, surpresa e tristeza, raiva e desgosto) uma média de 66,67% de acertos. Entretanto, quando uma sétima expressão é adicionada, no caso o medo, o sistema alcançou uma taxa média de reconhecimento de 52,38%. Para a classificação baseada em correspondência entre modelos, a base de dados FG-NET também foi utilizada para avaliar este método, porém o sistema não conseguiu alcançar uma taxa de reconhecimento maior do que à apresentada para a base MUG *Facial Expression*.

Tabela 4.5: Resultados da classificação baseada em correspondência entre modelos utilizando a base MUG *Expression Facial*

		Expressões						
# de Expressões	**% de Acerto**	**Neutra**	**Felicidade**	**Surpresa**	**Tristeza**	**Raiva**	**Desgosto**	**Medo**
3	97,78	96,67	100,00	96,67	-	-	-	-
4	73,33	83,33	100,00	96,67	13,33	-	-	-
5	69,33	73,33	100,00	96,67	13,33	63,33	-	-
6	66,67	73,33	93,33	96,67	13,33	56,67	66,67	-
7	52,38	70,00	93,33	26,67	13,33	56,67	66,67	40,00

No próximo Capítulo são apresentadas as avaliações dos resultados presentes neste Capítulo e as considerações finais do presente trabalho.

CAPÍTULO 5

AVALIAÇÃO DOS RESULTADOS E CONSIDERAÇÕES FINAIS

5.1 ANÁLISE DOS RESULTADOS

Nesta seção é realizada a análise dos resultados experimentais apresentados no Capítulo 4. Foram realizados experimentos para todos os módulos do sistema desenvolvidos neste trabalho. Para o módulo de detecção da face, o sistema detectou 100% das faces (ver Capítulo 4, seção 4.2). No módulo de detecção das regiões faciais (olhos, sobrancelhas e boca), o sistema atingiu uma taxa média de detecção de 98% (Capítulo 4, seção 4.2). Estes resultados foram obtidos quando o sistema foi treinado e testado utilizando a base de dados MUG *Facial Expression*. Através dos experimentos, pode-se perceber que o método (*Haar-like-features* como extrator de características e *AdaBoost* como classificador) adotado neste trabalho para as etapas de detecção da face e detecção das regiões da face (os olhos e a boca) apresentou resultados favoráveis, comprovando a eficiência do método apresentado pelos autores (VIOLA; JONES, 2001). Em relação ao módulo de detecção das regiões faciais, a área da sobrancelha foi encontrada, a partir das regiões dos olhos. A estratégia utilizada para localizar a área da sobrancelha, apesar de ser relativamente simples apresentou satisfatórios.

No módulo de extração de características, foi realizado um pré-processamento das imagens com o objetivo de remover eventuais ruídos e melhorar a qualidade de iluminação. Em seguida, são aplicados métodos de segmentação, e por fim *landmarks* faciais são extraídos da face. Após a extração dos *landmarks*, o método GPA foi aplicado aos dados com intuito de eliminar efeitos de rotação, translação e escala. Ao avaliar esta etapa do sistema, pode-se perceber que mesmo utilizando técnicas consideradas básicas de processamento de imagem as características extraídas (*landmarks*) apresentaram uma taxa de

precisão acima de 90% (Capítulo 4, seção 4.3). Além disso, ao comparar a precisão do algoritmo desenvolvido para extração de características neste trabalho com o *Active Shape Model* (ASM) (ver Tabela 4.2), percebe-se que apesar do ASM ser considerado bastante preciso, o método apresentado alcançou taxas de similaridades próximas ao obtido pelo método do ASM. Uma alternativa, para melhorar ainda mais a precisão dos *landmarks* extraídos, seria adicionar mais um módulo de correção de formas similar à estratégia utilizada por Beumer et al. (2006) em que as posições dos *landmarks* detectados incorretamente são corrigidos.

Na etapa de classificação as características extraídas, após aplicação da técnica GPA, são utilizadas para classificar as expressões faciais utilizando duas diferentes abordagens: as baseadas em redes neurais artificiais e em correspondência entre modelos. Para a classificação baseada em redes neurais artificiais, os resultados experimentais demonstraram que, ao classificar sete diferentes expressões, o sistema apresentou uma taxa de reconhecimento de 97,62% (Capítulo 4, subseção 4.2.1) para a base de dados MUG *Expression Facial* e 86,50% para a base FG-NET. A diferença entre os resultados pode ser explicada devido a base de dados MUG Facial *Expression* ser formada por imagens de indivíduos que imitaram corretamente as *expressões básicas universais* – os indivíduos foram informados sobre como as expressões faciais são realizadas conforme o manual do FACS. Enquanto que nas imagens da base de dados FG-NET, os indivíduos manifestaram as expressões de forma mais natural, sendo submetidos a vídeos de clipes de curta duração ou imagens (momento em que as expressões foram capturadas).

Para a classificação baseada em correspondência entre modelos, foram realizados testes em relação à taxa média de acerto entre 3 a 7 expressões, e através dos resultados experimentais, foi possível observar que a taxa de acerto deste método diminuiu à medida que aumentou-se a quantidade de expressões. Por exemplo, para as expressões: felicidade, neutra e surpresa o sistema atingiu uma taxa média de acerto de 97,78% (Capítulo 4, subseção 4.2.2). Para 4 expressões (felicidade, neutra, surpresa e tristeza) a média foi de 73,33% de acertos, enquanto que para 5 expressões (felicidade, neutra, surpresa e tristeza e raiva) a média foi de 69,33%, e para 6 expressões (felicidade, neutra, surpresa

e tristeza, raiva e desgosto) uma média de 66,67% de acertos. Entretanto, quando uma sétima expressão é adicionada, no caso o medo, o sistema alcançou uma taxa média de reconhecimento de 52,38%. Para a classificação baseada em correspondência entre modelos, algumas bases de dados foram utilizadas para avaliar este método, porém o sistema não conseguiu alcançar uma taxa de reconhecimento maior do que a apresentada para a base MUG *Facial Expression*.

A diferença dos resultados para os dois métodos de classificação aplicados neste trabalho pode ser explicada devido à aprendizagem de redes neurais artificiais apresentar robustez a ruídos nos dados de treinamento. Além disto, o processo de estimação do modelo médio pode suavizar alguns detalhes faciais eliminando algumas características importantes de cada expressão. Sendo assim, para melhorar a taxa de detecção do método baseado em modelo, é preciso melhorar a precisão dos *landmarks* detectados.

O sistema desenvolvido no presente trabalho foi comparado com outros sistemas propostos na literatura. Nesta comparação, apenas alguns sistemas que utilizam a mesma base de dados e reconhecem sete classes de expressões foram mencionados. Sendo que em cada um dos sistemas apresentados foram utilizadas metodologias diferentes de avaliação. Analisando a Tabela 5.1, observa-se que o método proposto neste trabalho apresentou uma taxa de reconhecimento relativamente maior que a maioria dos outros trabalhos apresentados.

5.2 CONCLUSÃO

Neste trabalho é apresentado um sistema automático de reconhecimento de expressões faciais que reconhece sete diferentes classes de expressões (felicidade, raiva, tristeza, surpresa, desgosto, medo e neutra). O sistema foi treinado e testado utilizando as bases de dados MUG *Facial Expression* e FG-NET em que as imagens apresentam iluminação uniforme, planos de fundo não uniforme e neutro e os indivíduos apresentam as seguintes diferenças individuais ou artefatos: barba, bigode e óculos. As imagens utilizadas pelo sistema estão restritas a ambientes fechados.

Tabela 5.1: Comparação de taxas de reconhecimento de abordagens que utilizaram a base de dados FG-NET

Autores	Características	Classificador	(%) de Acertos
Martins et al. (2008)	Formas das regiões faciais	Distância euclidiana	33,0
	Formas das regiões faciais	Redes neurais artificiais	53,7
	Formas das regiões faciais	SVM	80,1
	Formas das regiões faciais e textura	Redes neurais artificiais	72,0
	Formas das regiões faciais e textura	SVM	92,0
Hupont et al. (2008)	Distâncias entre *landmarks*	Baseado em regras	71,0
	Distâncias entre *landmarks* e *Gabor*	Baseado em regras	85,0
	Distâncias entre *landmarks* e formas da boca	Baseado em regras	91,0
Samad e Sawada (2011)	Gabor	SVM	81,7
Shan et al. (2009)	LBP	SVM	82,0
Sistema Proposto	*Landmarks* faciais	Redes neurais artificiais	86,5

No presente trabalho, dois métodos de classificação foram utilizados para reconhecer as expressões: os métodos baseados em redes neurais artificiais e em correspondência entre modelos. Os resultados experimentais demonstram que as melhores taxas de reconhecimento apresentada pelo sistema foi obtida com a utilização da rede neural artificial, alcançando uma taxa de reconhecimento de 97,62% para a base dados MUG *Facial Expression*, e 86,50% para a base FG-NET.

5.3 TRABALHOS FUTUROS

Algumas possíveis melhorias podem ser realizadas no sistema de reconhecimento de expressões desenvolvido neste trabalho, tais como:

• Adicionar ao sistema desenvolvido um novo módulo de correção de formas semelhante à estratégia utilizada por Beumer et al. (2006) em que as posições dos *landmarks* detectados incorretamente são corrigidos. Espera-se que, com esta estratégia, melhore a precisão dos *landmarks* detectados.

• Extrair outros tipos de características da face, como por exemplo utilizando o filtro de *Gabor* nos *landmarks* extraídos similar ao realizado no trabalho de Guo e Dyer (2005), fazendo como que o sistema funcione mesmo em situações nas quais existam variações de iluminação na imagem, rotação da cabeça, entre outros fatores que possam estar presentes em uma imagem.

APÊNDICE A

DETECÇÃO ROBUSTA DE OBJETO EM TEMPO REAL

A.1 BOOSTING

O conceito fundamental de *boosting* é a combinação de classificadores fracos $h_j(X)$ para construir um classificador forte $H(X)$, de modo que:

$$H(X) = \sum_{j=1}^{n} w_j h_j(X) \tag{1.1}$$

A seleção dos classificadores fracos $h_j(X)$ bem como a estimativa dos pesos w_j são aprendidas pelo procedimento de *boosting*. Cada classificador fraco $h_j(X)$ tem como objetivo minimizar o erro da classificação em uma distribuição especifica das amostras de treinamento. A cada iteração (isto é, para cada classificador fraco), o procedimento de *boosting* atualiza o peso de cada amostra de maneira que as classificadas erroneamente obtém maior peso na próxima iteração. O *boosting* concentra-se nas amostras que são difíceis de classificar. Existem diversas variações do *boosting* que se diferem principalmente pelo processo iterativo de como os pesos são recalculados no treinamento das amostras. O método desenvolvido por Viola e Jones (2001) é baseado no algoritmo de *boosting*, o *AdaBoost* (FREUND; SCHAPIRE, 1995)

A.2 CARACTERÍSTICAS DO TIPO HAAR-LIKE

Para representar a informação da imagem Viola e Jones (2001) definiu um conjunto de amostras de características. Estas características são derivadas de *Haar wavelets*, e o valor de uma determinada característica é calculado pelo somatório dos pixels na região branca subtraindo pelos da região preta. As características do tipo *Haar-like* podem ser

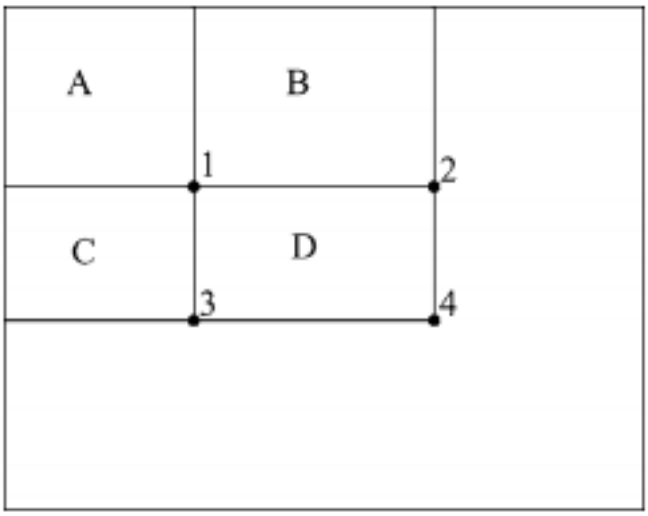


Figura A.1: A soma dos pixels dentro do retângulo D pode ser calculada com quatro referências de matriz. Fonte: (VIOLA; JONES, 2001)

computadas de maneira eficiente utilizando a representação da imagem integral. Dado um ponto de localização $i(x, y)$ em uma imagem, o valor da imagem integrante $ii(x, y)$ até este ponto, é a soma dos pixels acima e à esquerda de $i(x, y)$, de forma que:

$$ii(x, y) = \sum_{x' \leq x, y' \leq y} i(x', y') \tag{1.2}$$

onde $ii(x, y)$ é a imagem integrante até o ponto $i(x, y)$ da imagem original. Sendo $s(x, y)$ a soma cumulativa da linha, com $s(x, -1) = 0$ e $s(-1, y) = 0$, a imagem integral pode ser calculada sobre uma imagem original, conforme:

$$s(x, y) = s(x, y - 1) + i(x, y) \tag{1.3}$$

$$ii(x, y) = ii(x - 1, y) + s(x, y) \tag{1.4}$$

Ao calcular a imagem integral, é possível calcular o valor de uma determinada área retangular utilizando unicamente quatro pontos de vértices da área desejada. Considerando que se tem o valor total da área de origem da imagem até cada um dos quatro vértices, para encontrar a região que se deseja é preciso basicamente realizar subtração de retângulos. Por exemplo, na Figura A.1 para localizar a área da região D, é necessário

fazer o seguinte calculo: $4+1-(2+3)$. Assim, as características do tipo *Haar-like* podem ser calculadas normalmente de maneira rápida em qualquer escala e posição.

O conjunto de características é obtido através da variação do tamanho e posição de cada tipo de característica *Haar-like*. Para selecionar os classificadores fracos $h_j(X)$ da Equação 1.1, o processo de aprendizagem funciona da seguinte maneira: Cada característica candidata f_i é calculada em um conjunto de treinamento de amostras positivas e negativas. O classificador fraco em seguida, determina o limiar ótimo θ_j que minimiza o erro de classificação. A tarefa do processo de aprendizagem é obter a características f, tal que o número mínimo de amostra sejam classificadas erroneamente. Assim um classificador fraco $h_j(X)$ consiste de f_j características do tipo *Haar like*, um limiar θ_j e uma paridade p_j que indica a direção da desigualdade:

$$h(x, f, p, \theta) = \begin{cases} 1 & \text{se } p_f(x) < p_\theta \\ 0 & \text{caso contrário} \end{cases} \tag{1.5}$$

A.3 ARQUITETURA EM CASCATA

Considerando um conjunto de imagens, a taxa de detecção de um detector de face, por exemplo, é definida como sendo o número de faces detectadas corretamente em relação ao número total de faces no conjunto de teste. O falso positivo é calculado cada vez que o sistema classifica erroneamente uma região de plano de fundo como uma face. A maior taxa de detecção (com menos falsos positivos) pode ser conseguida através de um número maior de n de classificadores fracos $h_j(X)$. No entanto, aumentar a quantidade de n irá também aumentar a complexidade do classificador forte e, consequentemente, o tempo de computação. Para não comprometer nem o desempenho nem o tempo de computação, Viola e Jones (2001) propôs a estrutura em cascata de classificadores fortes.

A cascata de classificadores funciona da seguinte maneira: Dada uma entrada, esta é passada pelo primeiro classificador, que decide entre verdadeiro ou falso (objeto encontrado ou não encontrado). Uma determinação de falso interrompe a computação posterior e faz com que o detector retorne falso. Uma determinação verdadeira passa a entrada

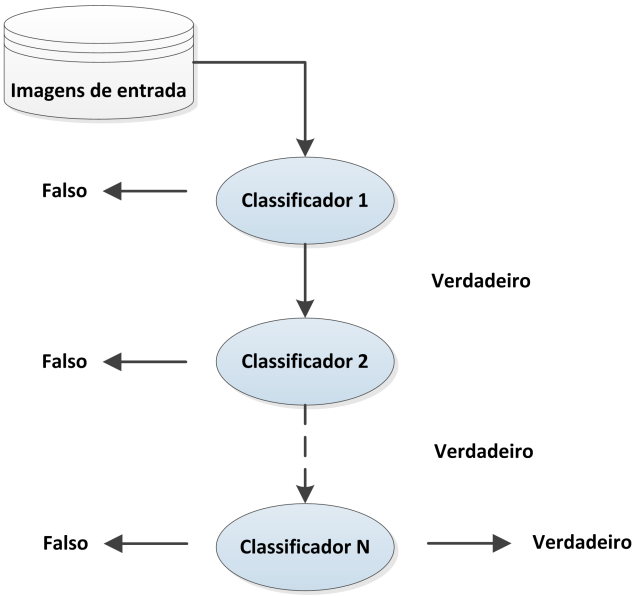


Figura A.2: Modelo em cascata do algoritmo de Viola e Jones (2001).

para o próximo classificador na cascata. Se todos os classificadores votarem em verdadeiro, a entrada é classificada como um exemplo verdadeiro. Consequentemente, vários ciclos computacionais são economizados, já que, se um dado de entrada é rejeitado logo no início, toda a computação posterior é evitada. A Figura A.2 mostra o esquema dos classificadores fortes em cascata. Em seus experimentos, Viola e Jones (2001) conseguiram detectar até 95% de faces frontais em imagens, com baixa taxa de falsos positivos.

REFERÊNCIAS BIBLIOGRÁFICAS

AIFANTI, N.; PAPACHRISTOU, C.; DELOPOULOS, A. The mug facial expression database. In: *11th International Workshop on Image Analysis for Multimedia Interactive Services*. [S.l.: s.n.], 2010. p. 1 –4.

ANDERSON, K.; MCOWAN, P. A real-time automated system for the recognition of human facial expressions. *IEEE Transactions on Systems, Man and Cybernetics*, v. 36, n. 1, p. 96–105, 2006.

BARTLETT, M. et al. *Automatic analysis of spontaneous facial behavior: A final project report*. Technical Report MPLab-TR2001.08, Univ. of California at San Diego, Dec 2001.

BARTLETT, M.; HAGER, J.; EKMAN, P.; SEJNOWSKI, T. Measuring facial expressions by computer image analysis. *Psychophysiology*, v. 36, n. 2, p. 253–63, 1999.

BARTLETT, M. S.; LITTLEWORT, G.; FASEL, I.; MOVELLAN, J. R. Real time face detection and facial expression recognition: Development and applications to human computer interaction. In: *In CVPR Workshop on CVPR for HCI*. [S.l.: s.n.], 2003.

BARTLETT, M. S. et al. Automatic recognition of facial actions in spontaneous expressions. *Journal of Multimedia*, v. 1, n. 6, p. 22–35, 2006.

BARTLETT, M. S. et al. Classifying facial action. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, v. 21, p. 974–989, 1996.

BASSILI, J. N. Emotion recognition: The role of facial movement and the relative importance of upper and lower areas of the face. *Journal of Personality and Social Psychology*, v. 37, p. 2049–2058, 1979.

BELHUMEUR, P. N.; HESPANHA, J. a. P.; KRIEGMAN, D. J. Eigenfaces vs. fisherfaces: Recognition using class specific linear projection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, IEEE Computer Society, Washington, DC, USA, v. 19, n. 7, p. 711–720, 1997.

BERGE, J. T. Orthogonal procrustes rotation for two or more matrices. *Psychometrika*, Springer New York, v. 42, p. 267–276, 1977. ISSN 0033-3123.

BEUMER, G. M.; TAO, Q.; BAZEN, A. M.; VELDHUIS, R. N. J. A landmark paper in face recognition. In: *Proceedings of the 7th International Conference on Automatic Face and Gesture Recognition*. Washington, DC, USA: IEEE Computer Society, 2006. (FGR '06), p. 73–78.

BLACK, M. J.; YACOOB, Y. Recognizing facial expressions in image sequences using local parameterized models of image motion. *International Journal of Computer Vision*, v. 25, p. 23–48, 1997.

BOOKSTEIN, F. L. Landmark Methods for Forms Without Landmarks: Localizing Group Differences in Outline Shape. In: *MMBIA '96: Proceedings of the 1996 Workshop on Mathematical Methods in Biomedical Image Analysis (MMBIA '96)*. Washington, DC, USA: IEEE Computer Society, 1996.

BOUREL, F.; CHIBELUSHI, C. C.; LOW, A. A. Robust facial expression recognition using a state-based model of spatially-localised facial dynamics. In: *5th IEEE International Conference on Automatic Face and Gesture Recognition (FGR 2002)*. [S.l.]: IEEE Computer Society, 2002. p. 113–118.

CANNY, J. A computational approach to edge detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-8, n. 6, p. 679 –698, nov. 1986.

CAO, J.; TONG, C. Facial expression recognition based on lbp-ehmm. In: *Congress on Image and Signal Processing*. [S.l.: s.n.], 2008. v. 2, p. 371 –375.

CHANG, Y.; HU, C.; FERIS, R.; TURK, M. Manifold based analysis of facial expression. *Image Vision Computing*, Butterworth-Heinemann, Newton, MA, USA, v. 24, n. 6, p. 605–614, jun. 2006.

CHUANG, C.-F.; SHIH, F. Y. Recognizing facial action units using independent component analysis and support vector machine. *Pattern Recognition*, v. 39, n. 9, p. 1795–1798, set. 2006.

COHEN, I.; SEBE, N.; CHEN, L.; GARG, A.; HUANG, T. S. Facial expression recognition from video sequences: Temporal and static modelling. In: *Computer Vision and Image Understanding*. [S.l.: s.n.], 2003. p. 160–187.

COHN, J.; ZLOCHOWER, A.; LIEN, J.-J. J.; KANADE, T. Automated face analysis by feature point tracking has high concurrent validity with manual facs coding. *Psychophysiology*, v. 36, p. 35 – 43, 1999.

COHN, J. F.; ZLOCHOWER, A. J.; LIEN, J.; KANADE, T.; ANALYSIS, A. F. *Automated Face Analysis by Feature Point Tracking Has High Concurrent Validity with Manual FACS Coding*. 1999.

COHN, J. F.; ZLOCHOWER, A. J.; LIEN, J. J.; KANADE, T. Feature-point tracking by optical flow discriminates subtle differences in facial expression. In: *Proceedings of the 3rd International Conference on Face & Gesture Recognition*. Washington, DC, USA: IEEE Computer Society, 1998. (FG '98), p. 396.

COOTES, T. F.; EDWARDS, G. J.; TAYLOR, C. *Active Appearance Models*. 1998.

COOTES, T. F.; TAYLOR, C. J.; COOPER, D. H.; GRAHAM, J. Active Shape Models-Their Training and Application. *Computer Vision and Image Understanding*, v. 61, n. 1, p. 38–59, jan. 1995.

DAILEY, M. N.; COTTRELL, G. W.; ADOLPHS, R. A six-unit network is all you need to discover happiness. In: *In TwentySecond Annual Conference of the Cognitive Science Society*. [S.l.]: Erlbaum, 2000. p. 101–106.

DARWIN, C. *The Expression of the Emotions in Man and Animals*. Anniversary. [S.l.]: Harper Perennial, 1872. Paperback.

DAUGMAN, J. Demodulation by complex-valued wavelets for stochastic pattern recognition. *International Journal of Wavelets, Multi-resolution and Information Processing*, v. 1, p. 1–17, 2003.

DONATO, G.; BARTLETT, M.; HAGER, J.; EKMAN, P.; SEJNOWSKI, T. Classifying facial actions. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, v. 21, n. 10, p. 974 –989, oct 1999.

DRYDEN, I. L.; MARDIA, K. *Statistical shape analysis*. Chichester [u.a.]: J. Wiley, 1998. (Wiley series in probability and statistics: Probability and statistics).

EDWARDS, G. J.; COOTES, T. F.; TAYLOR, C. J. Face recognition using active appearance models. In: *Proceedings of the 5th European Conference on Computer Vision*. London, UK, UK: Springer-Verlag, 1998. (ECCV '98).

EKMAN, P.; FRIESEN, W. *Facial Action Coding System: A Technique for the Measurement of Facial Movement*. Palo Alto: Consulting Psychologists Press, 1978.

EKMAN, P.; FRIESEN, W. V. Constants across cultures in the face and emotion. *Journal of Personality and Social Psychology*, v. 17, n. 2, p. 124–129, 1971.

ESSA, I.; PENTLAND, A. Coding, analysis, interpretation, and recognition of facial expressions. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, v. 19, n. 7, p. 757 –763, jul 1997.

FASEL, B.; LUETTIN, J. Automatic facial expression analysis: a survey. *Pattern Recognition*, v. 36, n. 1, p. 259 – 275, 2003.

FASEL, B.; MONAY, F.; GATICA-PEREZ, D. Latent semantic analysis of facial action codes for automatic facial expression recognition. In: *Proceedings of the 6th ACM SIGMM international workshop on Multimedia information retrieval*. New York, NY, USA: ACM, 2004. (MIR '04), p. 181–188.

FEI-FEI, L.; FERGUS, R.; PERONA, P. Learning generative visual models from few training examples: An incremental bayesian approach tested on 101 object categories. *Computer Vision and Pattern Recognition Workshop on Generative Model Based Vision*, Elsevier Science Inc., New York, NY, USA, v. 106, n. 1, p. 59–70, abr. 2007.

FORD, G. *Fully automatic coding of basic expressions from video*. San Diego, 2002.

FREUND, Y.; SCHAPIRE, R. E. A decision-theoretic generalization of on-line learning and an application to boosting. In: *Proceedings of the Second European Conference on Computational Learning Theory*. London, UK, UK: Springer-Verlag, 1995. (EuroCOLT '95), p. 23–37. ISBN 3-540-59119-2.

GAO, Y.; LEUNG, M.; HUI, S. C.; TANANDA, M. Facial expression recognition from line-based caricatures. *IEEE Transactions on Systems, Man and Cybernetics, Part A: Systems and Humans*, v. 33, n. 3, p. 407 – 412, may 2003.

GIRIPUNJE, S.; BAJAJ, P. Recognition of facial expressions for images using neural network. *International Journal of Computer Applications*, v. 40, n. 11, p. 3–7, December 2012.

GONZALEZ, R. C.; WOODS, R. E. *Digital image processing*. Upper Saddle River, N.J.: Prentice Hall, 2008.

GOWER, J. Generalized procrustes analysis. *Psychometrika*, Springer New York, v. 40, p. 33–51, 1975.

GU, H.; JI, Q. Facial event classification with task oriented dynamic bayesian network. In: *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. Washington, DC, USA: IEEE Computer Society, 2004. (CVPR'04), p. 870–875.

GUO, G.; DYER, C. Learning from examples in the small sample case: face expression recognition. *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, v. 35, n. 3, p. 477 –488, june 2005.

HAYKIN, S. *Neural Networks: A Comprehensive Foundation*. 2nd. ed. Upper Saddle River, NJ, USA: Prentice Hall PTR, 1998.

HESSE, N. *Multi-View Facial Expression Classification*. Tese (Doutorado) — Karlsruhe Institute of Technology, March 2011.

HOEY, J.; LITTLE, J. Value directed learning of gestures and facial displays. In: *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. [S.l.: s.n.], 2004. v. 2, p. II–1026 – II–1033 Vol.2.

HUANG, C.-L.; HUANG, Y.-M. Facial Expression Recognition Using Model-Based Feature Extraction and Action Parameters Classification. *Journal of Visual Communication and Image Representation*, v. 8, n. 3, p. 278–290, set. 1997.

HUPONT, I.; BALDASSARRI, S.; HOYO, R. D.; CEREZO, E. Effective emotional classification combining facial classifiers and user assessment. In: LÓPEZ, F. J. P.; FISHER, R. B. (Ed.). *Articulated Motion and Deformable Objects, 5th International Conference, AMDO 2008, Port d Andratx, Mallorca, Spain, July 9-11, 2008, Proceedings*. [S.l.]: Springer, 2008. (Lecture Notes in Computer Science, v. 5098), p. 431–440.

HUPONT, I.; CEREZO, E.; BALDASSARRI, S. Facial emotional classifier for natural interaction. *Electronic Letters on Computer Vision and Image Analysis*, v. 7, n. 4, 2008.

Jiang, B.; Valstar, M. F.; Pantic, M. Action unit detection using sparse appearance descriptors in space-time video volumes. In: *IEEE International Conference on Automatic Face & Gesture Recognition and Workshops, FG 2011*. USA: IEEE Computer Society, 2011. p. 314–321.

JUNG, S.-U.; KIM, D. H.; AN, K. H.; CHUNG, M. J. Efficient rectangle feature extraction for real-time facial expression recognition based on adaboost. In: *IROS*. [S.l.: s.n.], 2005. p. 1941–1946.

KAMBHATLA, N.; LEEN, T. K. Dimension reduction by local principal component analysis. *Neural Computation*, MIT Press, Cambridge, MA, USA, v. 9, n. 7, p. 1493–1516, out. 1997.

KANADE, T.; COHN, J.; TIAN, Y.-L. Comprehensive database for facial expression analysis. In: *Proceedings of the 4th IEEE International Conference on Automatic Face and Gesture Recognition (FG'00)*. [S.l.: s.n.], 2000. p. 46 – 53.

KHANAM, A.; SHAFIQ, M.; AKRAM, M. Fuzzy based facial expression recognition. In: *Congress on Image and Signal Processing*. [S.l.: s.n.], 2008. v. 1, p. 598 –602.

KHANDAIT, S. P.; THOOL, R. C.; KHANDAIT, P. D. Automatic facial feature extraction and expression recognition based on neural network. *Computing Research Repository*, abs/1204.2073, 2012.

KHANUM, A.; MUFTI, M.; JAVED, M. Y.; SHAFIQ, M. Z. Fuzzy case-based reasoning for facial expression recognition. *Fuzzy Sets Systems*, Elsevier North-Holland, Inc., Amsterdam, The Netherlands, The Netherlands, v. 160, n. 2, p. 231–250, jan. 2009.

KOBAYASHI, H.; HARA, F. Facial interaction between animated 3d face robot and human beings. In: *IEEE International Conference on Systems, Man, and Cybernetics*. [S.l.: s.n.], 1997. v. 4, p. 3732 –3737 vol.4.

KOTSIA, I.; NIKOLAIDIS, N.; PITAS, I. Facial expression recognition in videos using a novel multi-class support vector machines variant. In: *IEEE International Conference on Acoustics, Speech and Signal Processing*. [S.l.: s.n.], 2007. p. II–585 –II–588.

LIAO, S.; FAN, W.; CHUNG, A. C. S.; YEUNG, D.-Y. Facial expression recognition using advanced local binary patterns, tsallis entropies and global appearance features. In: . [S.l.]: IEEE International Conference on Image Processing (ICIP), 2006. p. 665–668.

LIEN, C.-C.; CHANG, Y.-K.; TIEN, C.-C. A fast facial expression recognition method at low-resolution images. In: *Proceedings of the 2006 International Conference on Intelligent Information Hiding and Multimedia*. Washington, DC, USA: IEEE Computer Society, 2006. (IIH-MSP '06), p. 419–422.

LIEN, J.-J. J.; KANADE, T.; COHN, J.; LI, C. Detection, tracking, and classification of action units in facial expression. *J. Robotics and Autonomous Systems*, v. 31, p. 131–146, 2000.

LITTLEWORT, G. et al. Towards social robots: Automatic evaluation of human-robot interaction by face detection and expression classification. In: . [S.l.]: In S. Thrun & L. Saul & B. Schoelkopf, 2004. v. 16, p. 1563–1570.

LITTLEWORT, G.; FASEL, I.; BARTLETT, M. S.; MOVELLAN, J. R. *Fully automatic coding of basic expressions from video*. [S.l.], May 2002.

LU, H.-C.; HUANG, Y.-J.; CHEN, Y.-W. Real-time facial expression recognition based on pixel-pattern-based texture feature. *Electronics Letters*, v. 43, n. 17, p. 916 –918, 16 2007.

LUCAS, B. D.; KANADE, T. An iterative image registration technique with an application to stereo vision. In: . [S.l.: s.n.], 1981. p. 674–679.

LUCEY, P. et al. The extended cohn-kanade dataset (ck+): A complete dataset for action unit and emotion-specified expression. In: *EEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*. [S.l.: s.n.], 2010. p. 94 –101.

LYONS, M.; AKAMATSU, S.; KAMACHI, M.; GYOBA, J. Coding facial expressions with gabor wavelets. In: *Third IEEE International Conference on Automatic Face and Gesture Recognition*. [S.l.: s.n.], 1998. p. 200 –205.

LYONS, M. J.; AKAMATSU, S.; KAMACHI, M.; GYOBA, J. Coding facial expressions with gabor wavelets. In: *FG*. [S.l.: s.n.], 1998. p. 200–205.

MALLAT, S. A theory for multiresolution signal decomposition: The wavelet representation. *IEEE Trans. Pattern Analysis and Machine Intelligence*, v. 11, n. 7, p. 674–693, 1989.

MARTIN, C.; WERNER, U.; GROSS, H.-M. A real-time facial expression recognition system based on active appearance models using gray images and edge images. In: *Proceedings of the 8th IEEE International Conference on Automatic Face and Gesture Recognition*. [S.l.: s.n.], 2008. p. 1–6.

MEHRABIAN, A. *Communication without words*. 2. ed. [S.l.: s.n.], 1968. 51-52 p.

MICHEL, P.; KALIOUBY, R. E. Real time facial expression recognition in video using support vector machines. In: *Proceedings of the 5th international conference on Multimodal interfaces*. New York, NY, USA: ACM, 2003. (ICMI '03), p. 258–264.

MILBORROW, S.; MORKEL, J.; NICOLLS, F. The muct landmarked face database. *In Proc. Pattern Recognition Association of South Africa*, 2010.

MILBORROW, S.; NICOLLS, F. Locating facial features with an extended active shape model. In: *Proceedings of the 10th European Conference on Computer Vision*. Berlin, Heidelberg: Springer-Verlag, 2008. (ECCV '08), p. 504–513. ISBN 978-3-540-88692-1.

NAGPAL, A.; GARG, A. Recognition of expressions on human face using ai techniques. *IJCSMS International Journal of Computer Science and Management Studies*, v. 11, Aug 2011.

NIXON, M.; AGUADO, A. S. *Feature Extraction & Image Processing, Second Edition*. 2nd. ed. [S.l.]: Academic Press, 2008.

OJALA, T.; PIETIKäINEN, M.; HARWOOD, D. A comparative study of texture measures with classification based on featured distributions. *Pattern Recognition*, v. 29, n. 1, p. 51–59, jan. 1996.

OJANSIVU, V.; HEIKKILä, J. Blur insensitive texture classification using local phase quantization. In: ELMOATAZ, A.; LEZORAY, O.; NOUBOUD, F.; MAMMASS, D. (Ed.). *3rd International Conference on Image and Signal Processing*. [S.l.]: Springer, 2008. (Lecture Notes in Computer Science, v. 5099), p. 236–243.

OLIVEIRA, F. V. *Facial Expression Classification using RBF and Back-Propagation Neural Networks*. 2000.

OLIVER, N.; PENTLAND, A.; BéRARD, F. Lafter: a real-time face and lips tracker with facial expression recognition. *Pattern Recognition*, p. 1369–1382, 2000.

OTSUKA, T.; OHYA, J. Spotting segments displaying facial expression from image sequences using hmm. In: *Proc. Int'l Conf. Automatic Face and Gesture Recognition*. [S.l.: s.n.], 1998. p. 442–447.

PADGETT, C.; COTTRELL, G. W. Representing face images for emotion classification. In: MOZER, M.; JORDAN, M. I.; PETSCHE, T. (Ed.). *NIPS*. [S.l.]: MIT Press, 1996. p. 894–900.

PADGETT, C.; COTTRELL, G. W.; ADOLPHS, R. Categorical perception in facial emotion classification. In: *In Proceedings of the 18th Annual Conference of the Cognitive Science Society*. [S.l.]: Erlbaum, 1996. p. 249–253.

PANTIC, M.; MEMBER, S.; ROTHKRANTZ, L. J. M. Automatic analysis of facial expressions: The state of the art. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, v. 22, p. 1424–1445, 2000.

PANTIC, M.; PATRAS, I. Dynamics of facial expression: recognition of facial actions and their temporal segments from face profile image sequences. *IEEE Transactions on Systems, Man, and Cybernetics*, v. 36, n. 2, p. 433–449, mar. 2006.

PANTIC, M.; ROTHKRANTZ, L. J. Facial action recognition for facial expression analysis from static face images. *Trans. Sys. Man Cyber. Part B*, IEEE Press, Piscataway, NJ, USA, v. 34, n. 3, p. 1449–1461, jun. 2004.

PANTIC, M.; ROTHKRANTZ, L. J. M. Expert system for automatic analysis of facial expressions. *Image Vision Computing*, v. 18, n. 11, p. 881–905, 2000.

PAPAGEORGIOU, C.; POGGIO, T. A trainable system for object detection. *International Journal of Computer Vision*, v. 38, n. 1, p. 15–33, 2000.

PARDAS, M.; BONAFONTE, A.; LANDABASO, J. L. Emotion recognition based on mpeg4 facial animation parameters. In: *IEEE International Conference on Acoustics, Speech, and Signal Processing ICASSP 2002*. Orlando, Florida, USA: [s.n.], 2002.

PATEL, I.; RAO, Y. S. Speech recognition using hidden markov model with mfcc-subband technique. In: *Proceedings of the 2010 International Conference on Recent Trends in Information, Telecommunication and Computing*. Washington, DC, USA: IEEE Computer Society, 2010. (ITC '10), p. 168–172.

POLAK, E.; RIBIERE, G. Note sur la convergence de méthodes de directions conjuguées. *ESAIM: Mathematical Modelling and Numerical Analysis - Modélisation Mathématique et Analyse Numérique*, EDP Sciences, v. 3, n. R1, p. 35–43, 1969.

RAO, K. S.; SAROJ, V. K.; MAITY, S.; KOOLAGUDI, S. G. Recognition of emotions from video using neural network models. *Expert Systems with Applications*, Pergamon Press, Inc., Tarrytown, NY, USA, v. 38, n. 10, p. 13181–13185, set. 2011. ISSN 0957-4174.

ROSS, A. *Procrustes Analysis*. [S.l.], 2005. Technical Report, Department of Computer Science and Engineering, University of South Carolina.

RYAN, A. et al. Automated facial expression recognition system. In: *International Carnahan Conference on Security Technology*. [S.l.: s.n.], 2009. p. 172 –177.

SAKET, K.; NARENDER, R.; S, H. Facial expression (mood) detection from facial images using committee neural networks. *Biomedical Engineering Online, V8, 16*, 2009.

SAKO, H.; SMITH, A. V. W. Real-time facial expression recognition based on features' positions on dimensions. In: *Proceedings of the International Conference on Pattern Recognition (ICPR'96)*. Washington, DC, USA: IEEE Computer Society, 1996. (ICPR '96), p. 643.

SAMAD, R.; SAWADA, H. Extraction of the minimum number of gabor wavelet parameters for the recognition of natural facial expressions. *Artificial Life and Robotics*, Springer-Verlag New York, Inc., v. 16, n. 1, p. 21–31, jun. 2011. ISSN 1433-5298.

SAYEED, A.; SOHAIL, M.; BHATTACHARYA, P. *Detection of Facial Feature Points Using Anthropometric Face Model*. 2006.

SEBE, N. et al. Authentic facial expression analysis. *Image and Vision Computing*, v. 25, n. 12, p. 1856 – 1863, 2007.

SEYEDARABI, H.; A., A.; S., K.; E., K. Analysis and synthesis of facial expressions by feature-points tracking and deformable model. *Iranian Journal of Electrical and Electronic Engineering*, 2007.

SHAN, C. *Inferring Facial and Body Language*. [S.l.], February 2008. Submitted for the degree of Doctor of Philosophy of the University of London.

SHAN, C.; BRASPENNING, R. Recognizing facial expressions automatically from video. In: NAKASHIMA, H.; AGHAJAN, H.; AUGUSTO, J. (Ed.). *Handbook of Ambient Intelligence and Smart Environments*. [S.l.]: Springer US, 2010. p. 479–509.

SHAN, C.; GONG, S.; MCOWAN, P. W. *Conditional Mutual Information Based Boosting for Facial Expression Recognition*. 2005.

SHAN, C.; GONG, S.; MCOWAN, P. W. Facial expression recognition based on local binary patterns: A comprehensive study. *Image Vision Computing*, v. 27, n. 6, p. 803–816, 2009.

SHAN, C.; GRITTI, T. Learning discriminative lbp-histogram bins for facial expression recognition. In: EVERINGHAM, M.; NEEDHAM, C. J.; FRAILE, R. (Ed.). *BMVC*. [S.l.: s.n.], 2008.

SHIH, F. Y.; CHUANG, C.-F. Automatic extraction of head and face boundaries and facial features. *Informatics and Computer Science*, Elsevier Science Inc., New York, NY, USA, v. 158, n. 1, p. 117–130, jan. 2004.

SIM, T.; BAKER, S.; BSAT, M. The cmu pose, illumination, and expression (pie) database. In: *Proceedings of the IEEE International Conference on Automatic Face and Gesture Recognition*. [S.l.: s.n.], 2002.

SMITH, E.; BARTLETT, M. S.; MOVELLAN, J. Computer recognition of facial actions: A study of co-articulation effects. In: *Proceedings of the 8th Annual Joint Symposium on Neural Computation*. [S.l.: s.n.], 2001.

SOBOTTKA, K.; PITAS, I. A fully automatic approach to facial feature detection and tracking. In: *Audio and Video-based Biometric Person Authentication, LNCS*. [S.l.]: Springer Verlag, 1997. p. 77–84.

STATHOPOULOU, I.-O.; TSIHRINTZIS, G. An improved neural-network-based face detection and facial expression classification system. In: *IEEE International Conference on Systems, Man and Cybernetics*. [S.l.: s.n.], 2004. v. 1, p. 666 –671 vol.1.

SUN, Y.; LI, Z.; TANG, C.; ZHOU, W.; JIANG, R. An evolving neural network for authentic emotion classification. In: *Proceedings of the 2009 Fifth International Conference on Natural Computation*. Washington, DC, USA: IEEE Computer Society, 2009. (ICNC '09), p. 109–113.

SUWA, M.; SUGIE, N.; K.FUJIMORA. A preliminary note on pattern recognition of human emotional expression. *International Joint Conference on Pattern Recognition*, p. 408–410, 1978.

TIAN, Y.; CHEN, S. Understanding effects of image resolution for facial expression analysis. *Journal of Computer Vision and Image Processing*, 2012.

TIAN, Y.-I.; KANADE, T.; COHN, J. Recognizing action units for facial expression analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, v. 23, n. 2, p. 97 –115, feb 2001.

TIAN, Y.-L.; KANADE, T.; COHN, J. Facial Expression Analysis. In: . [S.l.: s.n.], 2005. p. 247–275.

TIAN, Y.-l.; KANADE, T.; COHN, J. F. Eye-state action unit detection by gabor wavelets. In: *Proceedings of the Third International Conference on Advances in Multimodal Interfaces*. London, UK, UK: Springer-Verlag, 2000. (ICMI '00), p. 143–150.

TIAN, Y.-l.; KANADE, T.; COHN, J. F. Evaluation of gabor-wavelet-based facial action unit recognition in image sequences of increasing complexity. In: *Proceedings of the Fifth IEEE International Conference on Automatic Face and Gesture Recognition*. Washington, DC, USA: IEEE Computer Society, 2002. (FGR '02), p. 229.

TIAN, Y. li et al. Real world real-time automatic recognition of facial expressions. In: . [S.l.: s.n.], 2003.

TONG, Y.; LIAO, W.; JI, Q. Inferring facial action units with causal relations. In: *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. Washington, DC, USA: IEEE Computer Society, 2006. (CVPR '06), p. 1623–1630.

TORRE, F. D. la; COHN, J. F. Facial expression analysis. In: MOESLUND, T. B.; HILTON, A.; KRüGER, V.; SIGAL, L. (Ed.). *Visual Analysis of Humans*. [S.l.]: Springer, 2011. p. 377–409.

TSAPATSOULIS, N.; KARPOUZIS, K.; STAMOU, G.; PIAT, F.; KOLLIAS, S. A fuzzy system for emotion classification based on the mpeg-4 facial definition parameter set. In: . [S.l.]: European Signal Processing Conference (EUSIPCO'00), 2000.

TURK, M.; PENTLAND, A. Eigenfaces for recognition. *J. Cognitive Neuroscience*, Cambridge, MA, USA, v. 3, n. 1, p. 71–86, jan. 1991. ISSN 0898-929X.

VALSTAR, M.; PANTIC, M. Fully automatic facial action unit detection and temporal analysis. In: *Proceedings of the 2006 Conference on Computer Vision and Pattern Recognition Workshop*. Washington, DC, USA: IEEE Computer Society, 2006. (CVPRW '06), p. 149–157.

VIOLA, P.; JONES, M. Robust real-time object detection. In: *International Journal of Computer Vision*. [S.l.: s.n.], 2001. v. 57, p. 137–154.

VISUTSAK, P. Emotion classification using adaptive svm's. *International Journal of Computer and Information Engineering*, 2012.

VUKADINOVIC, D.; PANTIC, M. Fully automatic facial feature point detection using gabor feature based boosted classifiers. In: *Proceedings of IEEE Int. Conf. Systems, Man and Cybernetics (SMC'05)*. Waikoloa, Hawaii: [s.n.], 2005. p. 1692–1698.

WALLHOFF, F. *Database with Facial Expressions and Emotions from Technical University of Munich (FEEDTUM)*. 2006.

WANG, M.; IWAI, Y.; YACHIDA, M. Expression recognition from time-sequential facial images by use of expression change model. In: *Third IEEE International Conference on Automatic Face and Gesture Recognition*. [S.l.: s.n.], 1998. p. 324 –329.

WANG, Y.; AI, H.; WU, B.; HUANG, C. Real time facial expression recognition with adaboost. In: *17th International Conference on Pattern Recognition (ICPR'04)*. Washington, DC, USA: IEEE Computer Society, 2004. (ICPR '04), p. 926–929.

WHITEHILL, J.; OMLIN, C. Haar features for facs au recognition. In: *7th International Conference on Automatic Face and Gesture Recognition*. [S.l.: s.n.], 2006. p. 5 pp. –101.

XIAO, J.; KANADE, T.; COHN, J. Robust full motion recovery of head by dynamic templates and re-registration techniques. In: *Proceedings of the Fifth IEEE International Conference on Automatic Face and Gesture Recognition (FG'02)*. [S.l.: s.n.], 2002. p. 156 – 162.

YEASIN, M.; BULLOT, B.; SHARMA, R. From facial expression to level of interest: a spatio-temporal approach. In: *Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. Washington, DC, USA: IEEE Computer Society, 2004. (CVPR'04), p. 922–927.

YONEYAMA, M.; IWANO, Y.; OHTAKE, A.; SHIRAI, K. Facial expressions recognition using discrete hopfield neural network. In: *ICIP (3)*. [S.l.: s.n.], 1997. p. 117–120.

YOUSSIF, A. A. A.; ASKER, W. A. A. Automatic facial expression recognition system based on geometric and appearance features. *Computer and Information Science*, p. 115–124, 2011.

ZENG, Z. et al. Spontaneous emotional facial expression detection. *Journal of Multimedia*, v. 1, p. 1–8, 2006.

ZHAN, C.; LI, W.; OGUNBONA, P.; SAFAEI, F. Facial expression recognition for multiplayer online games. In: *Proceedings of the 3rd Australasian conference on Interactive entertainment*. Murdoch University, Australia, Australia: Murdoch University, 2006. (IE '06), p. 52–58.

ZHAN, C.; LI, W.; OGUNBONA, P.; SAFAEI, F. Facial expression recognition for multiplayer online games. In: *Proceedings of the 3rd Australasian conference on Interactive entertainment*. Murdoch University, Australia, Australia: Murdoch University, 2006. (IE '06), p. 52–58.

ZHANG, Y.; JI, Q. Facial expression understanding in image sequences using dynamic and active visual information fusion. In: *Proceedings of the Ninth IEEE International Conference on Computer Vision*. Washington, DC, USA: IEEE Computer Society, 2003. (ICCV '03), p. 1297. ISBN 0-7695-1950-4.

ZHANG, Y.; JI, Q. Active and dynamic information fusion for facial expression understanding from image sequences. *IEEE on Pattern Analysis and Machine Intelligence*, IEEE Computer Society, Washington, DC, USA, v. 27, n. 5, p. 699–714, maio 2005. ISSN 0162-8828.

ZHANG, Z.; LYONS, M.; SCHUSTER, M.; AKAMATSU, S. Comparison between geometry-based and gabor-wavelets-based facial expression recognition using multi-layer perceptron. In: . [S.l.: s.n.], 1998. p. 454–459.

ZHANG, Z.; ZHANG, Z. Feature-based facial expression recognition: Sensitivity analysis and experiments with a multilayer perceptron. *International Journal of Pattern Recognition and Artificial Intelligence*, v. 13, p. 893–911, 1999.

ZHAO, G.; PIETIKAINEN, M. Dynamic texture recognition using local binary patterns with an application to facial expressions. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, IEEE Computer Society, Washington, DC, USA, v. 29, n. 6, p. 915–928, jun. 2007.

ZHAO, J.; KEARNEY, G. Classifying facial emotions by backpropagation neural networks with fuzzy inputs. *Proceedings of Conference on Neural Information Processing*, v. 1, p. 454–457, 1996.

ZLOCHOWER, A. J.; COHN, J. F.; LIEN, J. J.-J.; KANADE, T. Automated face coding: A computer-vision based method of facial expression analysis in parent-infant interaction. *Infant Behavior e Development*, v. 21, p. 16–16, 1998.