UNIVERSIDADE DO VALE DO RIO DOS SINOS
UNIDADE ACADÊMICA DE PESQUISA E PÓS-GRADUAÇÃO
PROGRAMA DE PÓS-GRADUAÇÃO EM COMPUTAÇÃO APLICADA
NÍVEL MESTRADO

EDUARDO DE OLIVEIRA

UM SISTEMA DE INFERÊNCIA DE EXPRESSÕES FACIAIS EMOCIONAIS ORIENTADO NO MODELO DE EMOÇÕES BÁSICAS

SÃO LEOPOLDO
2011

Eduardo de Oliveira

Um Sistema de Inferência de Expressões Faciais Emocionais Orientado no Modelo de Emoções Básicas

Dissertação apresentada como requisito parcial para a obtenção do título de Mestre, pelo Programa de Pós-Graduação em Computação Aplicada da Universidade do Vale do Rio dos Sinos.

Orientadora:
Profa. Dra. Patrícia Augustin Jaques Maillard

São Leopoldo
2011

AGRADECIMENTOS

Agradeço, em primeiro lugar, a minha amada Daniela. Ela me apoiou, incentivou, auxiliou e participou em meu projeto durante os intensos 24 meses de mestrado. Nos momentos de intranquilidade e de questionamentos, sempre pude contar com seus preciosos e inspiradores conselhos, que foram essenciais para o desenvolvimento de meu trabalho. Todos agradecimentos são poucos em comparação a sua importância não só sobre essa dissertação, mas sobre minha vida.

À professora Patrícia Maillard, pelo seu apoio, paciência, dedicação e pronto auxílio com minhas demandas. A orientação a esta dissertação soma-se a outra, ocorrida em meu trabalho de conclusão de curso de graduação.

A meus professores do PIPCA, pelos válidos ensinamentos. Participaram ou contribuíram mais diretamente com minhas pesquisas os professores Cláudio Jung, Luiz Paulo Luna de Oliveira e João Valiati.

Ao professor Wilson Gavião Neto, por suas sugestões em Visão Computacional e a Henrique Seffrin, pelo auxílio com as redes neurais.

A Unisinos, pelo suporte oferecido, incluindo nisso o auxílio oferecido por Sandra Rodrigues, da secretaria do PIPCA.

Ao Banco Santander, pela bolsa.

RESUMO

Este trabalho apresenta um sistema que realiza automaticamente a inferência das emoções básicas (alegria, tristeza, raiva, medo, repulsa e surpresa) pelas expressões da face de um usuário de computador, através de imagens capturadas por uma *webcam*. A aplicação desenvolvida baseia-se no sistema de codificação facial FACS, que classifica as ações faciais em códigos específicos, conhecidos como AUs (*Action Units*). A abordagem utilizada consiste em coletar dados de movimentações da boca, olhos e sobrancelhas para classificar, via redes neurais, os códigos AUs presentes nas expressões faciais executadas. Por meio de árvore de decisão, conjunto de regras ou rede neural, as emoções dos AUs, anteriormente classificados, são inferidas. O sistema construído foi avaliado sobre três cenários diferentes: (1) utilizando amostras de bases de faces para avaliação de reconhecimento de AUs e emoções; (2) com amostras de bases de faces para avaliação de reconhecimento de emoções por rede neural (abordagem alternativa); (3) utilizando uma amostra composta por imagens capturadas por *webcam* para avaliação de emoção, por árvore de decisão e rede neural. Como resultados, foi obtida uma taxa de reconhecimento sobre AUs de 53,83%, implicando em 28,57% de reconhecimento de emoções pelo inferidor da árvore de decisão - Cenário 1. Já, a inferência de emoção pela rede neural obteve como melhor resultado 63,33% de taxa de reconhecimento - Cenários 2 e 3. O trabalho desenvolvido pode ser utilizado para ajustar o comportamento do computador ao estado afetivo do usuário ou fornecer dados para outros *softwares*, como sistemas tutores inteligentes.

Palavras-chave: Visão Computacional. Interação Humano-Computador. Computação Afetiva.

ABSTRACT

This work presents a system that automatically performs the inference of basic emotions (happiness, sadness, anger, fear, disgust and surprise) through facial expressions from a computer user, using images captured by a webcam. The developed application is based on facial coding system FACS, that classifies specific facial actions, known as AUs (Action Units). The proposed approach consists in collecting moviment data of mouth, eyes and eyebrows to classify, by neural networks, AUs codes presents in performed facial expressions. With decision tree, ruleset or neural network, the emotions of AUs, previously classified, are inferred. The designed system was evaluated in three different scenarios: (1) using samples of faces bases to evaluate the recognition of AUs and emotions; (2) with samples of face bases to evaluate emotion recognition by neural network (alternative approach); (3) using a sample of images captured by webcam for evaluation of emotion in decision tree and neural network. As results, was obtained 53.83% of recognition rate over AUs, which implicating 28.57% of emotions recognition with decision tree - Scenario 1. The emotion inference by neural network achieve, 63.33% of recognition rate as the best result - Scenarios 2 and 3. The developed paper can be used to adjust computer's behavior to address user's affective state, or provides data to other softwares, such as intelligent tutoring systems.

Keywords: Computer Vision. Human-Computer Interaction. Affective Computing.

SUMÁRIO

LISTA DE FIGURAS

LISTA DE TABELAS

LISTA DE ABREVIATURAS E SIGLAS

AAM	*Active Appearance Models*
AFA	*Automatic Face Analisys*
AU	*Action Unit*
CK+	*The Extended Cohn-Kanade Dataset*
DO	Distância entre Cantos Internos dos Olhos
DO_n	Distância entre Cantos Internos dos Olhos - face neutra
DO_e	Distância entre Cantos Internos dos Olhos - face com expressão
E/S	Entrada/Saída
EM	*Expectation-Maximization*
FA	*Factor Analysis*
FACS	*Facial Action Code System*
FaD	*Face Detection*
FAP	*Facial Animation Parameters*
FCP	*Facial Characteristic Point*
FeD	*Feature Detection*
FN	Falso Negativo
FP	Falso Positivo
GF	*Gabor Filter*
HMM	*Hidden Markov Model*
IEC	*International Electrotechnical Commission*
IHC	Interação Humano Computador
ISO	*International Organization for Standardization*
JAFFE	*The Japanese Female Facial Expression Database*
KLT	Kanade-Lucas-Tomasi
k-NN	*k-Nearest Neighbor*
LDA	*Linear Discriminant Analysis*

LF	Lógica Fuzzy
MSE	*Mean Square Error*
MLP	*Multi-Layer Perceptron*
MPEG	*Moving Picture Experts Group*
MPI-FVD	*Max Planck Institute Face Video Database*
N/D	Não Disponível
OpenCV	*Open Source Computer Vision Library*
PCA	*Principal Component Analysis*
RBC	Raciocínio Baseado em Casos
RBD	Redes Bayesianas Dinâmicas
RNA	Rede Neural Artificial
RNDA	*Recursive Nonparametric Discriminant Analysis*
SNoW	*Sparse Network of Winnows*
SVM	*Support Vector Machine*
TREC	Taxa de reconhecimento
VN	Verdadeiro Negativo
VP	Verdadeiro Positivo
VECF	Vetor de Estados das Características Faciais
$VECF_n$	Vetor de Estados das Características Faciais - face neutra
$VECF_e$	Vetor de Estados das Características Faciais - face com expressão
VC	Vetor de Características
VC_N	Vetor de Características - normalizado

LISTA DE SÍMBOLOS

$\widehat{A}$	Reflexão do conjunto A
$(A)_x$	Translação do conjunto A ao ponto x
$\oplus$	Operação morfológica de dilatação
$\ominus$	Operação morfológica de erosão
$\circ$	Operação morfológica de abertura
$\bullet$	Operação morfológica de fechamento

1 INTRODUÇÃO

Hoje o computador é uma ferramenta de uso cotidiano e indispensável em várias atividades. Sua presença é crescente e convergente aos mais diversos dispositivos, porém, a interface básica de entrada existente entre o homem e o computador é praticamente a mesma nos últimos 30 anos, ou seja, *mouse* e teclado. A área de pesquisa Interação Humano-Computador (IHC) é uma das que procuram melhorar e evoluir essa interação, tornando-a mais amigável, ágil e clara (BOOTH, 1995). Seus estudos vão desde melhorias em interfaces gráficas até sistemas que utilizam as emoções expressas por usuários como parâmetros de entrada. Pesquisas em Computação Afetiva, área de intersecção entre IHC, ciências cognitivas e psicologia, buscam levar em consideração os estados afetivos de seus usuários em sua interação. Para Computação Afetiva, o computador deve ter a capacidade para inferir as emoções humanas, além de expressar afeto e até possuir suas próprias emoções (PICARD, 1995).

As emoções humanas podem ser manifestadas de diversas formas, como pela voz, expressões faciais e pelos sinais fisiológicos do corpo (respiração, ritmo cardíaco etc). Da mesma forma, uma emoção pode ser inferida utilizando uma ou mais fontes combinadas. A inferência de emoções pelas expressões faciais é a mais próxima das formas mais primitivas utilizadas pelo homem (EKMAN, 1999). No entanto, esta é uma tarefa complexa e desafiadora para aplicações computacionais, pois necessita combinar diferentes técnicas para classificação das emoções pelas expressões faciais.

Visando contribuir para IHC, considerando emoções nas relações entre homem e máquina, objetiva-se realizar a inferência de emoção expressa por um usuário a frente do computador utilizando imagens de sua face captadas por uma *webcam*. Embora a maioria dos trabalhos existentes realizem a inferência de emoção através da abordagem conexionista, neste trabalho foi empregada uma abordagem simbólica, com base em um modelo de classificação de movimentos de músculos faciais, o FACS (introduzido na Seção 2.3.2), e por antropometria facial. Trata-se de uma abordagem simbólica, pois através da utilização de FACS é possível obter representações simbólicas de emoções construídas sobre um modelo psicológico de emoções

básicas. Além disso, será utilizada uma árvore de decisão, que utiliza estes símbolos em seu aprendizado.

FACS fornece códigos, chamados de AUs (*Action Units*), para todas as movimentações de músculos faciais (podem representar um ou mais músculos no mesmo código), como por exemplo, comprimir os lábios, representado pelo AU 24. Desta forma, pode-se representar qualquer expressão facial utilizando combinações de AUs. Já, estudos sobre antropometria facial são utilizados com dois objetivos: localizar regiões de características faciais; e utilizar áreas sobre características faciais relevantes para identificação de expressões faciais. Estes estudos antropométricos fornecem as proporções faciais e as relações entre regiões da face.

A obtenção de emoção pela face é a mais comumente utilizada nas relações homem - homem (PANTIC; BARTLETT, 2007), pois oferece elementos mais facilmente detectáveis. Sabendo a emoção expressa por um usuário, o sistema poderia, por exemplo, tentar acalmar uma pessoa que aparenta raiva, tornando o seu ambiente mais agradável e evitando situações conhecidamente irritantes para este usuário. Outras vantagens da obtenção da emoção pela *webcam* são a valorização deste dispositivo como interface de entrada adicional ao teclado e *mouse*, e a disponibilidade de dados sobre afetividade, que podem ser empregados como entrada em sistemas tutores inteligentes, por exemplo. A *webcam* também tem a vantagem de ser um meio de detecção não intrusivo, pois não necessita de contato físico com o usuário, diferentemente de equipamentos para detectar emoções através da pressão arterial, respiração ou condutividade elétrica da pele.

Mais especificamente, o objetivo deste trabalho é inferir as emoções demonstradas por uma pessoa a frente do computador, através das imagens provenientes de uma *webcam*. Para chegar a este objetivo principal, foram traçados os seguintes objetivos específicos:

- Pesquisar métodos para detecção da face e de características faciais, para realizar classificações das emoções presentes nas expressões faciais;
- Verificar uma abordagem simbólica/probabilística para tratar esse problema;
- Implementar uma aplicação computacional para efetuar automaticamente a identificação da emoção expressa em uma face.

Esta proposta de trabalho utilizou como ponto de partida estudos anteriores (OLIVEIRA, 2008; OLIVEIRA; JAQUES, 2008), onde foi realizada a inferência de emoções básicas utilizando um sistema semi-automático. Essa base anteriormente construída sofreu aperfeiçoamentos nos métodos, tornando o sistema automático e eficiente. Neste sentido, foram adotados

métodos de Processamento de Imagens que possibilitam condições mais favoráveis à aplicação das detecções sobre características faciais. Também houve inclusão de métodos de classificação mais adequados aos problemas de classificação de expressões faciais e inferência de emoções, ou seja, redes neurais, árvores de decisão e conjunto de regras.

Na organização deste trabalho, foram revisados, no capítulo de Referencial Teórico, os conceitos necessários ao entendimento e implementação dos objetivos traçados, com ênfase nos estudos de Visão Computacional e de Computação Afetiva. Alguns trabalhos relacionados são apresentados no Capítulo 3 e o trabalho proposto no capítulo seguinte (Capítulo 4). Em seguida, foram realizadas avaliações do sistema no Capítulo 5 e, encerrando este trabalho, o capítulo Considerações Finais apresenta resultados, desafios e perspectivas futuras.

2 REFERENCIAL TEÓRICO

Este capítulo aborda os termos necessários para compreensão e execução desse trabalho, que envolve as áreas de pesquisa de Processamento de Imagens, Visão Computacional e Computação Afetiva. Em Processamento de Imagens são abordadas diversas técnicas relevantes que favorecem a captura de expressões faciais. Na seção de Visão Computacional, serão apresentadas as técnicas de detecção de face e de características faciais que são usadas para extração das partes relevantes da imagem para a inferência da emoção. Destaca-se nesta seção o método de Viola-Jones, que é utilizado no trabalho. Em seguida, na seção de Computação Afetiva, é apresentada uma breve explanação sobre a pesquisa atual na área, assim como definições encontradas na literatura da psicologia de emoções. Também é apresentado o sistema de codificação facial FACS, que será empregado no trabalho proposto para inferência de emoções, assim como as abordagens que vêm sendo utilizadas pelos pesquisadores para reconhecimento computacional de emoções por face.

2.1 Processamento de Imagens

Uma imagem pode ser representada como sendo uma função de duas dimensões, $f(x,y)$, composta por um domínio discreto e finito, em que os valores de cada coordenada (equivalente a um pixel[1]) representam a intensidade ou nível de cinza sobre um ponto da imagem (GONZALEZ; WOODS, 2001). Porém, para que o computador possa representar e manipular uma imagem, é comumente utilizada a forma matricial, onde $f(x,y)$ equivale a uma matriz $M \times N$.

Segundo Gonzalez e Woods (2001), a área de Processamento de Imagens digitais apresenta métodos computacionais para duas principais aplicações: melhoria da representação de imagens e processamento de dados da imagem para armazenamento, transmissão e representação. A primeira área de aplicação foca em ajustes de imagens que viabilizem a interpretação humana, principalmente na restauração e aperfeiçoamento de imagens em áreas astronômicas,

[1] Pixel é derivado de *picture element*.

biológicas, médicas e industriais, por exemplo. Já a segunda, tem como objetivo a utilização de métodos para percepção de máquina em aplicações de reconhecimento de caracteres, visão de máquina na indústria (inspeção e montagem), reconhecimento de impressões digitais, dentre outras.

Como definição geral, Processamento de Imagens possui como entrada e como saída uma imagem, porém existem alguns relacionamentos entre Processamento de Imagens e outras áreas que pesquisam imagens. Nessas relações, existem métodos que realizam processos desde os níveis mais baixos (onde são aplicados operações como filtros), passando por um nível intermediário (onde já há descrição e reconhecimento de formas e objetos), até o nível mais alto (onde há análises em nível cognitivo sobre imagens). Se as áreas de pesquisa relacionadas fossem organizadas em ordem de complexidade, encontraria-se, em sequência, métodos de Processamento de Imagens (realizado nos níveis de processo baixo e intermediário), Análise de Imagens e Visão Computacional (ambas realizadas no nível alto de processo). Embora exista esta ordem, não existe uma fronteira de abrangência bem definida entre os métodos de cada área (GONZALEZ; WOODS, 2001). Nas seções a seguir, baseadas em Gonzalez e Woods (2001), serão apresentados alguns métodos existentes em Processamento de Imagens que são de interesse deste trabalho.

2.1.1 Operações morfológicas

Operações morfológicas são técnicas que realizam transformações sobre imagens que podem resultar em filtragens, afinamento, junção ou separação de elementos (regiões que têm formas, apresentadas nas imagens). Sua teoria vem da morfologia matemática, que realiza estudos sobre formas de objetos utilizando a teoria de conjuntos para isso. As operações morfológicas foram originalmente desenvolvidas para imagens binárias[2], mas também existem técnicas para operações sobre imagens em tons de cinza[3] e coloridas[4].

Operações de Dilatação e Erosão

Existem diversas operações morfológicas, sendo base para muitas delas dilatação e erosão. Em uma definição matemática formal, considerando A e B como conjuntos no espaço bidimensional $\mathbb{Z}^2$ (espaço que representa imagens binárias) e $\varnothing$ como conjunto vazio, a **dilata-**

[2]Uma imagem binária é aquela que utiliza duas cores em sua representação, ou seja, 1 bit.

[3]Uma imagem em tons cinza é, normalmente, representada por 256 tons (níveis) de cinza, isto é, 8 bits.

[4]Uma imagem colorida pode ser representada por vários espaços de cores diferentes, sendo o RGB o mais comum. Ele é composto por três canais, um para a cor vermelha (*Red*), um para verde (*Green*) e um para azul (*Blue*), que contém a intensidade dessas cores. É normalmente representado por 24 bits (8 bits por canal).

ção é definida pela Equação (2.1) e a **erosão**, pela Equação (2.2).

$$A \oplus B = \left\{ x | (\widehat{B})_x \cap A \neq \varnothing \right\} \tag{2.1}$$

$$A \ominus B = \{ x | (B)_x \subseteq A \} \tag{2.2}$$

Normalmente, o conjunto B é chamado de elemento estruturante, que tem um ponto de origem (também chamado de ponto âncora) e pode possuir qualquer forma. Nos exemplos apresentados a seguir pode-se ver a ação de um elemento estruturante quadrado sobre outro maior, em uma operação de dilatação (Figura 2.1) e de erosão (Figura 2.2).

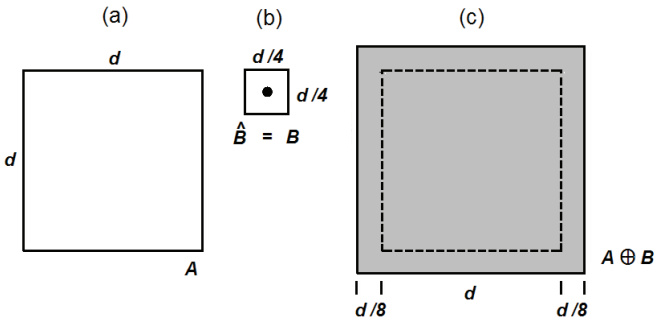


Figura 2.1: (a) Conjunto original; (b) elemento estruturante; (c) dilatação de A por B. (Adaptado de Gonzalez e Woods (2001)).

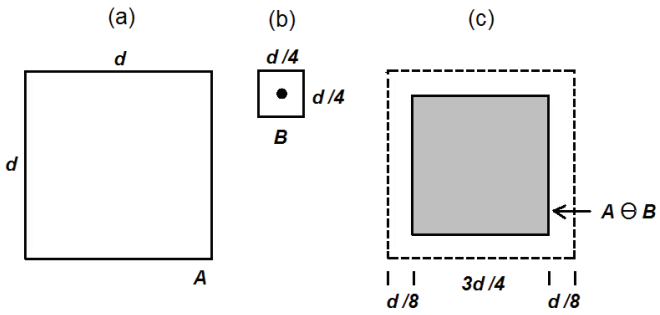


Figura 2.2: (a) Conjunto original; (b) elemento estruturante; (c) erosão de A por B. (Adaptado de Gonzalez e Woods (2001)).

Conforme dito anteriormente, as operações morfológicas foram construídas para utilização sobre imagens binárias, porém, para que elas sejam aplicadas em imagens em tons de cinza (ou coloridas), outras técnicas são necessárias. Os conjuntos passam a ser representados como funções, realizando-se convolução[5] entre uma imagem (anteriormente chamado de

[5]Convolução, que em Processamento de Imagens é normalmente aplicável em processos de filtragem e obtenção de bordas, trata-se de uma transformação que ocorre entre funções, onde uma das funções é a imagem e a outra um *kernel*. Este *kernel* percorre toda a imagem e atribui nela, pelo seu ponto âncora (normalmente no centro do *kernel*), o resultado, que geralmente é a soma de produtos de sua vizinhança com a imagem.

conjunto A) e o elemento estruturante (anteriormente chamado de conjunto B). No caso da dilatação, substitui-se o valor do pixel do ponto âncora pelo valor máximo de sua vizinhança dentro do elemento estruturante, resultando em uma imagem mais clara. O mesmo método é aplicado na erosão, porém busca-se o valor mínimo, além do resultado ser o oposto, obtém-se uma imagem mais escura. A Figura 2.3 exemplifica a aplicação de erosão e dilatação.

Figura 2.3: (a) Imagem original; (b) resultado da dilatação; (c) resultado da erosão. (Adaptado de Bradski e Kaehler (2008)).

Operações de Abertura e Fechamento

Em imagens binárias, a operação de abertura do conjunto A por um elemento estruturante B é definida pela Equação (2.3).

$$A \circ B = (A \ominus B) \oplus B. \tag{2.3}$$

Isto é, aplica-se erosão e após dilatação. Como resultado, obtém-se a suavização de contornos, a formação de ilhas e eliminação de pontos pequenos.

Já na operação de fechamento em imagens binárias do conjunto A por um elemento estruturante B, utiliza-se como definição a Equação (2.4).

$$A \bullet B = (A \oplus B) \ominus B. \tag{2.4}$$

Neste caso, aplica-se dilatação e em seguida erosão. Com isso, é possível conectar pequenas regiões, suavizar contornos e eliminar buracos. Pode-se ver na Figura 2.4 o passo a passo dos operadores de abertura e fechamento.

Sobre imagens em tons de cinza, as operações de abertura e fechamento realizam a mesma sequência de dilatação e erosão realizadas sobre imagens binárias. Observa-se que

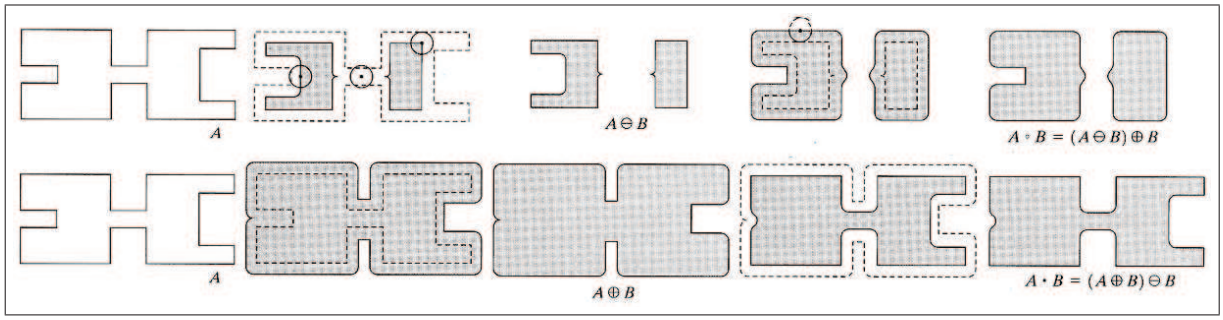


Figura 2.4: Passo a passo da abertura (linha superior) e fechamento (linha inferior). (De Gonzalez e Woods (2001)).

com abertura, detalhes claros são eliminados, mantendo-se os níveis de cinza, enquanto que o fechamento realiza o contrário, removendo detalhes escuros e mantendo os elementos claros. A Figura 2.5 mostra um exemplo dessas operações sobre imagem em tons de cinza.

Figura 2.5: (a) Imagem original; (b) resultado da abertura; (c) resultado do fechamento. (De Bradski e Kaehler (2008)).

2.1.2 Processamento de Histograma

Por definição, um histograma de uma imagem em tons de cinza com intervalo $[0, L-1]$ é uma função discreta $p(r_k) = \frac{n_k}{n}$, em que r_k é o k-ésimo nível de cinza, n_k é o número de pixels da imagem com esse tom de cinza, n é o número total de pixels e k = 0, 1, 2, ... , L-1 (GONZALEZ; WOODS, 2001). Através do gráfico da função $p(r_k)$ há indícios da aparência de uma imagem, mas sem descrições sobre seu conteúdo (Figura 2.6). Porém, os indícios fornecidos pelo histograma podem ser suficientes para que este possa ser realçado.

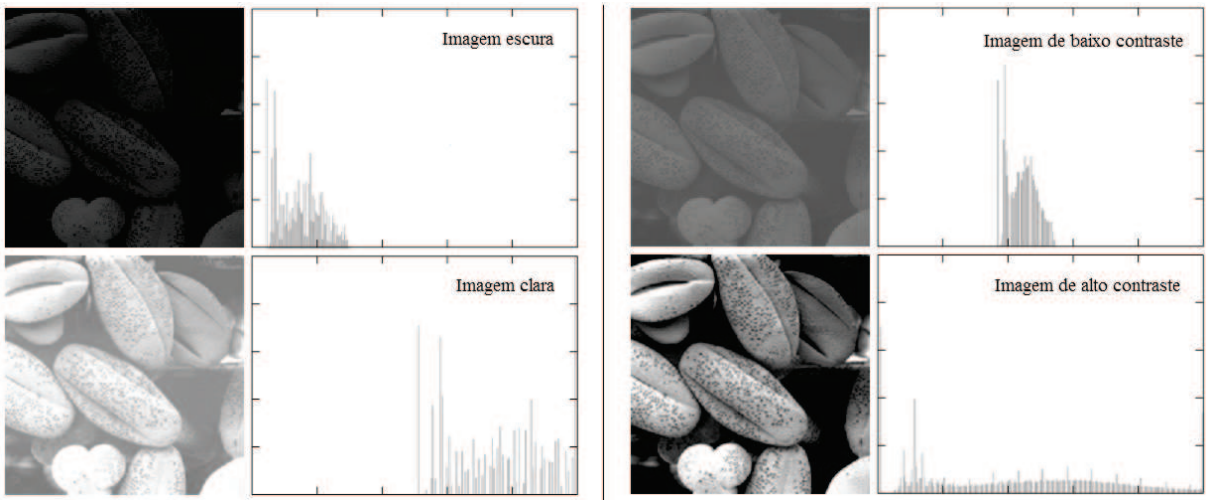


Figura 2.6: Imagem escura/clara, baixo/alto contraste e seu respectivo histograma. (De Gonzalez e Woods (2001)).

Equalização de Histograma

O método equalização de histograma pode ser aplicado sobre imagens que apresentem distorções na distribuição dos tons de cinza. Ele consiste em utilizar uma função de transformação para mapear o histograma igual à distribuição acumulada da imagem. A Figura 2.7 é um exemplo de equalização de histograma.

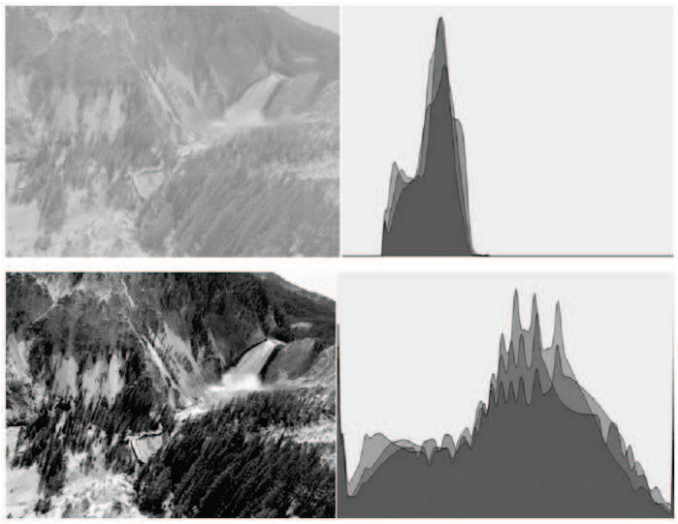


Figura 2.7: Exemplo de equalização de histograma. Linha superior, imagem original; linha inferior, imagem equalizada. (De Bradski e Kaehler (2008)).

Especificação de Histograma

A ideia de especificação de histograma é poder utilizar uma função de distribuição específica para mapear um histograma. A sua vantagem sobre equalização de histograma, que gera uma aproximação de um histograma uniforme, é a liberdade de emprego de uma transformação que obtenha realces específicos.

A Figura 2.8(b) exibe um exemplo do resultado da equalização de histograma aplicado sobre a Figura 2.8(a). Devido a sua natureza, composta por pixels mais escuros em sua maioria, a equalização de histograma não resultou em uma imagem melhor, porém, aplicando a especificação de histograma, os resultados, que são vistos na Figura 2.8(c), foram superiores.

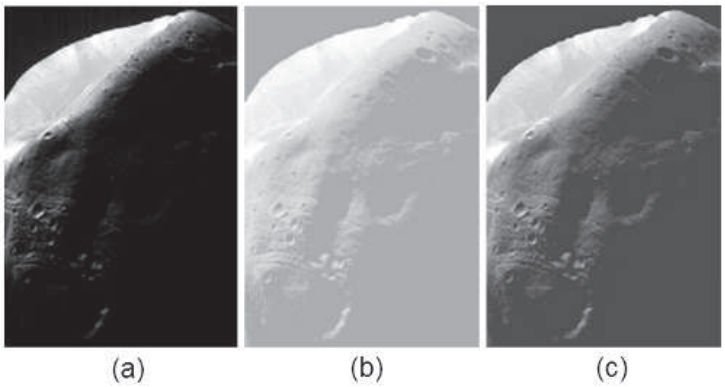


Figura 2.8: (a) Imagem original; (b) resultado da equalização de histograma; (c) resultado da especificação de histograma. (De Gonzalez e Woods (2001)).

2.1.3 Limiarização

As técnicas de limiarização têm como objetivo a segmentação de imagens. Elas utilizam-se da distribuição de intensidade das imagens em suas operações. Através dessas técnicas, imagens binárias em preto e branco (não em tons de cinza) podem ser obtidas.

Limiarização Global Simples

Analisando o histograma (Figura 2.9(b)) da Figura 2.9(a), nota-se claramente que existem dois agrupamentos de intensidades de pixels que referem-se ao fundo e a impressão digital. O agrupamento menor e mais próximo de zero, que representa a intensidade de pixels mais escuros, refere-se a impressão digital e, consequentemente, o outro agrupamento, ao fundo.

Este caso é considerado apropriado à utilização de limiarização global simples. Nessa técnica, a imagem é varrida e cada pixel analisado e, com base em um limiar (pré-definido ou

identificado por algum algoritmo), recebe uma atribuição sobre ser componente do fundo ou do primeiro plano. O resultado da aplicação deste método é exibido na Figura 2.9(c).

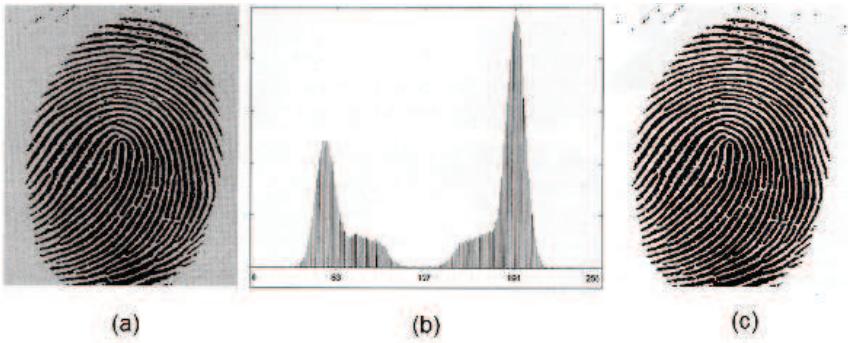


Figura 2.9: (a) Imagem original; (b) histograma desta imagem; (c) Imagem após limiarização global simples. (De Gonzalez e Woods (2001)).

Limiarização Adaptativa

Na Figura 2.10(a), vê-se uma imagem que possui em seu histograma dois agrupamentos de níveis de cinza bem definidos. Porém, adicionando sombra (Figura 2.10(b)) à imagem da Figura 2.10(a), obtém-se a imagem da Figura 2.10(c), que apresenta um histograma sem um limiar explícito.

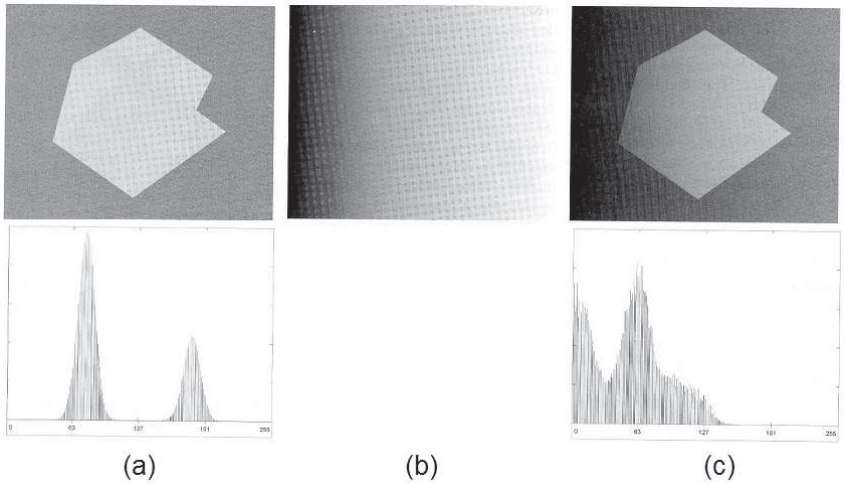


Figura 2.10: (a) Imagem original com histograma; (b) sombra; (c) resultado da soma entre a primeira e segunda imagens e o histograma resultante. (De Gonzalez e Woods (2001)).

Aplicando manualmente a técnica de limiarização global simples, considerando o vale apresentado no histograma da Figura 2.10(c) como limiar, é obtido o resultado da Figura 2.11(b). Este resultado não se mostra satisfatório, mas a aplicação da técnica de limiarização adaptativa fornece um resultado melhor.

A técnica de limiarização adaptativa consiste em aplicar limiarização global simples sobre regiões menores e, desta forma, obter limiares mais precisos. Como pode-se ver na Figura 2.11(c), a imagem foi dividida em quatro quadrantes e cada um novamente dividido em outros quatro quadrantes. O resultado da técnica de limiarização adaptativa aplicada sobre cada região pode ser visto na Figura 2.11(d). Resultados ainda melhores podem ser obtidos no exemplo caso sejam aplicadas novas divisões nas regiões centrais, que não foram bem sucedidos em sua limiarização.

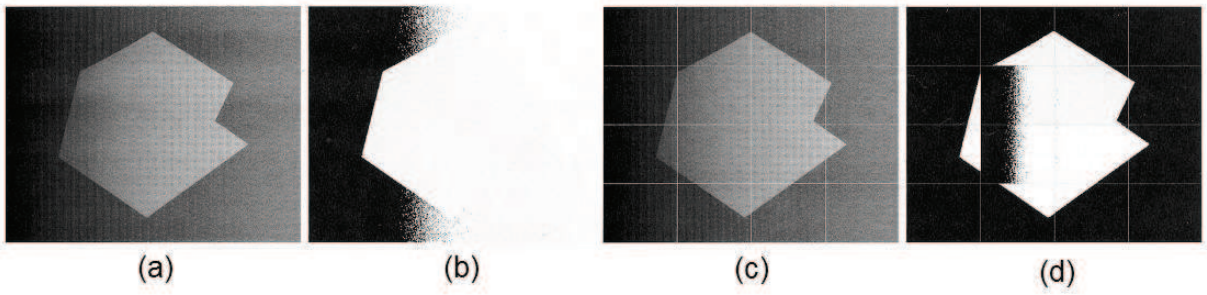


Figura 2.11: (a) Imagem original; (b) resultado da limiarização global simples sobre imagem original; (c) imagem original dividida em quadrantes; (d) resultado da limiarização adaptativa sobre imagem original. (De Gonzalez e Woods (2001)).

2.1.4 Alargamento de contraste

A técnica de alargamento de contraste é utilizada para alterar os níveis de cinza de uma imagem, de forma a aumentar o seu contraste. Para isso, essa técnica realiza uma transformação utilizando-se de uma função definida em trechos, que descreve o comportamento na escala de cinza de uma imagem de acordo com as variáveis da função.

A Figura 2.12(a) ilustra um exemplo de função utilizada em alargamento de contraste, onde os pontos (r_1, s_1) e (r_2, s_2) fazem o controle da função. No exemplo, caso $r_1 = s_1$ e $r_2 = s_2$, a transformação é uma função linear que não altera o contraste; caso $r_1 = r_2$, $s_1 = 0$ e $s_2 = L - 1$, a transformação opera como uma função de limiarização; valores intermediários de (r_1, s_1) e (r_2, s_2) produzem vários níveis de tons de cinza na imagem. A Figura 2.12(b) contém a imagem original, a Figura 2.12(c) a imagem onde foi aplicado um intervalo que considerou os valores de níveis de cinza máximos e mínimos e a Figura 2.12(d) apresenta o resultado da transformação com a função de limiarização.

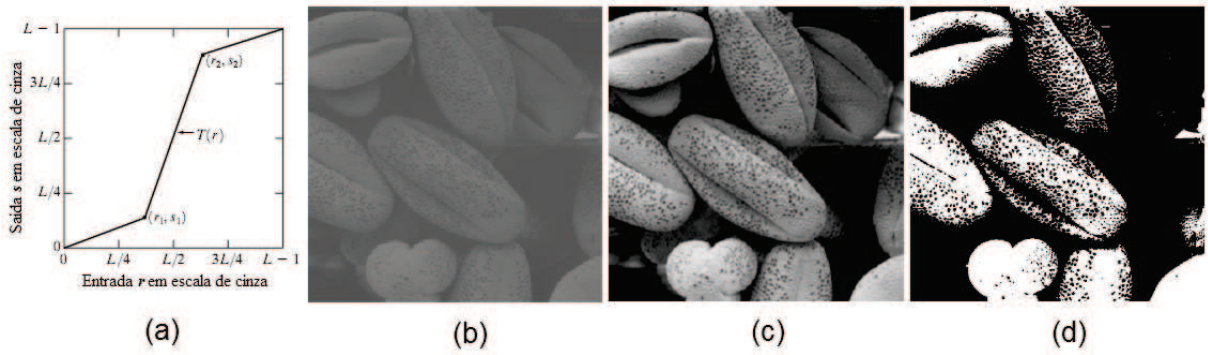


Figura 2.12: (a) Função de transformação; (b) imagem com baixo contraste; (c) imagem com alargamento de contraste; (d) imagem com limiarização. (De Gonzalez e Woods (2001)).

2.2 Visão Computacional

Visão Computacional aplica métodos para obter informações sobre imagens, para realizar inferências posteriormente. Shapiro e Stockman (2001, p. 13) definem os objetivos de Visão Computacional como: *"tomar decisões úteis sobre cenas e objetos físicos reais com base em imagens captadas."*[6]. É considerada por alguns autores como uma subárea de Inteligência Artificial, pois utiliza-se de técnicas dessa área, como aprendizado e reconhecimento de padrões (GONZALEZ; WOODS, 2001).

Para a detecção de emoção de expressões faciais pelo computador, é essencial a aplicação dos conhecimentos de Visão Computacional. Dentre as várias frentes encontradas em Visão Computacional, as mais importantes para este trabalho são Detecção de Face (FaD) e Detecção de Características (FeD). A FaD busca reconhecer em uma imagem a área referente ao rosto de uma pessoa, enquanto FeD tenta localizar determinadas características em imagens. O termo "características", em Visão Computacional, refere-se a um determinado elemento que se busca encontrar ou identificar em uma imagem. As características de um rosto (como olhos, boca, sobrancelhas etc.), que serão buscadas neste trabalho, podem ser rastreadas em uma imagem e terem a sua área delimitada. De posse da área dos olhos, sobrancelhas e boca, por exemplo, pode-se tentar identificar que tipo de emoção é expressa por essas características.

Nas próximas seções serão revisados conceitos e métodos que envolvem FaD e FeD. Na Seção 2.2.1, sobre FaD, serão apresentados alguns métodos relacionados que executam esta tarefa. Sobre FeD (Seção 2.2.2), a abordagem de alguns autores e as principais técnicas terão uma breve apresentação. Há na Seção 2.2.3 a apresentação da biblioteca OpenCV, com enfoque

[6] *"The goal of computer vision is to make useful decisions about real physical objects and scenes based on sensed images"*

em métodos para detecção de objetos. E a última seção (Seção 2.2.4) introduz métodos de rastreamento de objetos.

2.2.1 Detecção de Faces

A FaD é aplicada em muitas pesquisas como uma etapa preliminar de atividades mais complexas. Este é o caso de trabalhos envolvendo Interação Humano-Computador, reconhecimento ou identificação de faces, localização de faces e detecção de características faciais. Conforme definição de Yang, Kriegman e Ahuja (2002, p. 1), a tarefa básica de algoritmos de FaD é: *"Dada uma imagem arbitrária, o objetivo da detecção facial é determinar se existem ou não faces na imagem e, se presentes, retornar a localização e área de cada face."*[7]

Existem alguns desafios relacionados à FaD que aumentam a complexidade de métodos para esta tarefa. A pose da face (frontal, perfil, inclinação, cabeça para baixo), presença ou falta de componentes estruturais (barba, bigode, óculos), expressões faciais, oclusão de face (partes da face ocultas), orientação da imagem (rotação do eixo óptico de câmeras) e condições da imagem (intensidade luminosa e características das câmeras) são obstáculos encontrados na FaD. A Figura 2.13 exemplifica estes desafios: tem-se na subfigura A um rosto com inclinação e oclusão; a B mostra um rosto em perfil e outro inclinado; na C existe uma imagem com baixa qualidade e com inclinação da face; na D um rosto sobre baixa intensidade luminosa aparece rotacionado (nem frontal, nem em perfil); E é um exemplo de rostos em vários graus de inclinação; F mostra uma oclusão em grande área do rosto; em G uma face tem as sobrancelhas cobertas por um chapéu e a boca parcialmente coberta por bigode; H apresenta muita luminosidade; I apresenta oclusão de um dos olhos e a presença de óculos. Estes desafios acabam causando dois tipos de erros nos algoritmos criados para este fim: os **falsos negativos** (FN), quando faces existentes não são identificadas, e os **falsos positivos** (FP), quando regiões são identificadas como faces, mas não contêm uma. O inverso destes erros, que são as identificações com sucesso, são o **verdadeiro positivo** (VP), quando um rosto foi realmente encontrado, e **verdadeiro negativo** (VN), quando não houve nenhuma identificação onde não havia rosto.

Esta seção é composta por quatro subseções. As três primeiras, baseadas nos trabalhos de Yang, Kriegman e Ahuja (2002) e Yang (2004), abordarão algumas técnicas que se destacam na tarefa de localização de rostos em imagens. Mais especificamente, na primeira seção, serão descritos métodos de detecção sobre imagens estáticas[8] que estarão classificados

[7] *"Given an arbitrary image, the goal of face detection is to determine whether or not there are any faces in the image and, if present, return the image location and extent of each face."*

[8] A imagem estática em questão, refere-se ao fato de os métodos realizarem suas buscas individualmente sobre cada imagem, não considerando as buscas realizadas em imagens anteriores ou posteriores.

Figura 2.13: Exemplos de desafios encontrados pelos métodos de FaD. Subfigura A obtida na Internet; B, D, E e G de Rowley, Baluja e Kanade (1998a); C, H e I de Rowley, Baluja e Kanade (1998b); F de Schneiderman e Kanade (1998).

segundo definição dos autores. Na segunda seção, serão apresentados os estudos relacionados a métodos envolvendo FaD sobre imagens coloridas. Técnicas sobre identificação de rostos em vídeo são introduzidas na terceira seção. Por fim, uma quarta seção trata sobre outras categorias estudadas.

Detecção de faces sobre imagens estáticas

Os métodos para detecção de face em imagens estáticas, que serão aqui descritos, foram classificados por Yang, Kriegman e Ahuja (2002) em quatro categorias:

- **Métodos baseados em conhecimento:** funções que contêm propriedades atribuídas a rostos humanos são estudadas por estes métodos;

- **Métodos baseados em características:** são os que tentam localizar rostos pela identificação de características humanas (características faciais, textura, cor da pele, múltiplas características) que mantêm determinadas propriedades independente de sua disposição ou condição ambiental;

- **Métodos de comparação de modelos:** modelos de contorno de faces são comparados com objetos na busca de possíveis rostos;

- **Métodos baseados em aparência:** são métodos que utilizam algoritmos de aprendizado de máquina para que, com base em exemplos de imagens de rostos, consigam realizar sua identificação.

A seguir, é realizada uma descrição sucinta sobre cada categoria, sobre alguns de seus métodos, os prós e contras e a indicação de trabalhos significativos. Os métodos baseados em aparência terão maior atenção em relação aos demais, pois são os que se ajustam mais a proposta deste trabalho.

Métodos baseados em conhecimento Esta é uma abordagem onde estão métodos que aplicam o conhecimento empírico sobre os elementos constituintes de faces humanas e sobre a relação entre suas características como um todo, em forma de regras. Sua pesquisa desenvolveu-se com base no fato de que, em imagens de pessoas, normalmente encontram-se dois olhos, uma boca e sobrancelhas que mantêm uma determinada simetria entre si, que podem ser definidas em regras.

O método baseado em conhecimento implementado por Yang e Huang (1994) (que é uma referência para a categoria), atua de forma *top-down*, pois divide a sua busca em três níveis de complexidade, do mais abrangente ao mais detalhado. No primeiro nível, os candidatos a rosto são selecionados de uma imagem pela aplicação de algoritmos de janelas deslizantes[9] e por regras sobre definição de rostos. No nível seguinte, alguns filtros são aplicados sobre os candidatos selecionados no primeiro nível, com o objetivo de aumentar a qualidade da análise da fase seguinte. E no último nível, outras regras específicas relacionadas às características faciais são aplicadas sobre os candidatos restantes, sendo selecionados aqueles que atenderem as especificações definidas.

São vantagens encontradas nesta abordagem, a facilidade para definição de regras e o bom desempenho obtido na busca por faces em fundos de cenas não complexos[10]. Como

[9]Janela deslizante (*window scanning*): algoritmos que exploram uma imagem deslocando-se pixel a pixel, sobre um conjunto de pixels que compõem a chamada janela (2x2, 100x100 etc). Janela também pode ser chamada por: máscara, filtro, *kernel* ou *template*.

[10]Fundo de cena não complexo pode ser definido como um fundo de um objeto que não possui muitos elementos em sua constituição, como uma parede branca. Um exemplo de fundo de cena complexo pode ser um corredor movimentado de um *shopping center*.

desvantagem, existe a dificuldade em definir regras abrangentes para definição do que é uma face, pois se o nível de detalhamento for muito alto, o índice de rejeição será maior e, se o nível de detalhamento for baixo, aumentam as possibilidades de falsos positivos. Outra dificuldade é a definição de faces em poses variadas.

Métodos baseados em características Os métodos desta abordagem, também conhecida como abordagem sobre características invariantes, baseiam-se na capacidade que os seres humanos têm de identificar objetos nas mais variadas condições luminosas e de posicionamento, a partir de fragmentos. Pesquisas foram direcionadas, neste sentido, sobre as características ou propriedades que invariam nas diversas condições e que permitem ao homem realizar a identificação de uma face.

Uma particularidade dos métodos baseados em características, é que eles buscam encontrar primeiramente as características faciais contidas em um rosto e, posteriormente, modelos estatísticos são aplicados para confirmar se a suspeita é verdadeira. Por executar primeiramente tarefas em nível mais detalhado para somente depois tentar inferir a presença de um rosto, a abordagem caracteriza-se por ser implementação de nível *bottom-up*, inverso ao proposto pelos métodos baseados em conhecimento.

Existem métodos implementados sobre várias características humanas que podem se referir a uma face ou permitir que se chegue até ela. As características utilizadas para esta busca podem ser:

- **Características faciais:** alguns métodos desta abordagem tentam localizar características faciais que mantêm certas propriedades em diferentes faces (caso dos dois olhos, da boca, do nariz) (LEUNG; BURL; PERONA, 1995; YOW; CIPOLLA, 1997);

- **Textura:** a característica da pele e do cabelo presente em faces contém texturas que podem diferenciar faces humanas de outros objetos (DAI; NAKANO, 1996);

- **Cor de pele:** pela identificação da cor de pele, pode-se encontrar a região de uma face (YANG; WAIBEL, 1996; MCKENNA; GONG; RAJA, 1998);

- **Múltiplas características:** a utilização das técnicas acima descritas combinadas já foi implementada em algumas pesquisas (KJELDSEN; KENDER, 1996).

A grande vantagem apresentada pelos métodos desta abordagem é que as características invariantes se mantêm e são localizadas independentemente de pose ou orientação. Porém, existem dificuldades de detecção em fundos de cena complexos e os métodos têm pouca sensibilidade a variações luminosas e a oclusão de características.

Métodos de comparação de modelos Os métodos de comparação de modelos (*template matching*) utilizam-se de um padrão de face (máscara) que é pré-definido ou parametrizado por funções. A correlação entre os contornos da máscara com a imagem de uma possível face (contornos da face, olhos, boca, nariz) é avaliada e pode confirmar que se trata de uma identificação positiva ou não.

Entre os métodos de comparação de modelos, existem os que utilizam modelos pré-definidos e modelos deformáveis:

- **Modelos pré-definidos:** métodos deste tipo normalmente tentam, em uma primeira etapa, extrair as bordas e contrastes de imagens que podem estar relacionadas a contornos e a características de uma face. Os dados extraídos são posteriormente comparados com os modelos de faces que, conforme sua aderência, confirmam ou não se tratar da identificação positiva de uma face (CRAW; TOCK; BENNETT, 1992);

- **Modelos deformáveis:** funções tentam localizar os contornos de um rosto pela aplicação de técnicas de identificação de bordas. Após um rosto candidato ser identificado, suas bordas, picos e vales são submetidos a um modelo, que tenta se ajustar a estes elementos. O atendimento de alguns requisitos definidos por funções na utilização do modelo deformável confirma a existência de uma face em uma imagem (LANITIS; TAYLOR; COOTES, 1995).

Os métodos desta abordagem têm como vantagem a simplicidade de implementação, porém existem dificuldades para definição de modelos devido a variações de pose, de escala e de forma de faces.

Métodos baseados em aparência Nesta categoria, os métodos utilizam classificadores binários para a detecção de faces. Estes classificadores tentam classificar dados de entradas em dois grupos: dos que atendem determinadas propriedades e dos que não atendem estas propriedades. Trazendo esta definição ao presente trabalho, um classificador tem a função de decidir se elementos de uma imagem correspondem ou não a uma face. Para isso, anteriormente é

necessário realizar o treinamento do classificador sobre as propriedades a serem consideradas. Nestes treinos, um conjunto de imagens positivas do objeto a ser localizado (no caso, imagens de faces) e de imagens negativas (que são aquelas que não contêm o objeto alvo presente) devem ser submetido aos algoritmos de classificação. Através de técnicas de análise estatística e de conhecimento de máquina, os métodos desta categoria localizam e demarcam a região correspondente a uma face.

Os métodos baseados em aparência são os que vêm obtendo mais atenção devido a sua maior eficiência e robustez em comparação aos demais. Por este motivo, nesta categoria é encontrada uma maior quantidade de implementações. Mas, além da robustez já destacada, que proporciona uma grande taxa de acerto, a rapidez na busca e FaDs em poses e orientações variadas são outros pontos fortes. Pesa contra este método o grande esforço necessário para a construção de um classificador, que exige grande quantidade de exemplos positivos e negativos de faces para o treinamento dos classificadores, o que consome tempo para esta seleção e posterior treinamento. Também, a necessidade de realização de buscas sobre toda imagem e em escalas variadas é outro ponto fraco.

Alguns dos trabalhos mais representativos (e o método utilizado) na categoria baseada em aparências são: Eigenfaces (TURK; PENTLAND, 1991); *Distribution Based* (SUNG; POGGIO, 1998); Redes Neurais (ROWLEY; BALUJA; KANADE, 1998b); *Support Vector Machines* (SVM) (OSUNA; FREUND; GIROSI, 1997); *Naive Bayes* (SCHNEIDERMAN; KANADE, 1998); *Hidden Markov Model* (HMM) (RAJAGOPALAN et al., 1998); S*parse Network of Winnows* (SNoW) (YANG; ROTH; AHUJA, 1999); *Principal Component Analysis* (PCA) e *Factor Analysis* (FA) (YANG; AHUJA; KRIEGMAN, 2000); Viola-Jones (*Haar-like features*) (VIOLA; JONES, 2001).

Detecção de faces em imagens coloridas

É possível localizar faces realizando buscas em imagens por tom de pele de pessoas, considerando as diversas etnias, utilizando para isto métodos estatísticos sobre vários sistemas de cores (RGB, *normalized* RGB, HSV, HIS, YCrCb, YIQ, UES, CIE XYZ, CIE LIV). Estes métodos conseguem identificar a cor de pele em cerca de 80% dos casos, porém, devido a grande taxa de falsos positivos, é necessária a aplicação de métodos posteriores (YANG, 2004).

Os métodos que utilizam cor de pele para localização de faces são de fácil implementação e não são afetados por variação de pose, rotação ou expressão. Entretanto, pode ocorrer sensibilidade a variação luminosa, além de interferências pela variação de tom de pele e de outras partes do corpo.

Na seção anterior, de detecção de faces em imagens estáticas, na categoria de imagens baseadas em características, o método para localização de faces pela identificação de cor de pele também é abordado. Na seção Outras categorizações é descrita uma explicação para este fato.

Algumas referências aos estudos de detecção de faces em imagens coloridas são: (SENIOR et al., 2002; MENSER; MÜLLER, 1999; TOMAZ, 2010; STÖRRING; ANDERSEN; GRANUM, 1999; STÖRRING, 2004; JONES; REHG, 1998; HSU; ABDEL-MOTTALEB; JAIN, 2002).

Detecção de faces baseada em vídeo

A diferenciação entre *frames*[11] em vídeos denuncia movimentação de objetos em um ambiente. Parte do cenário que se apresenta estático pode, então, ser descartado, reduzindo a área de busca que se restringirá a parte onde há movimento detectado. Desta forma, os métodos baseados em movimento isolam uma área onde pode ser encontrada uma face, que dificilmente se mantém sempre estática.

Estes métodos possibilitam que sejam encontradas faces de forma mais fácil do que em relação a imagens estáticas. É possível, também, utilizar a combinação de movimentação, profundidade e voz para aumentar a redução da área de busca, porém faltam métodos eficientes para processar as entradas combinadas. A existência de outros movimentos além da face, também dificulta a utilização do método, como destaca Frischholz (2010). Mikolajczyk, Choudhury e Schmid (2001), além de Hjelmås, Lerøy e Johansen (1998) e Crowley e Coutaz (1995) são algumas importantes referências nesta abordagem.

Outras categorizações

Não existe um padrão para organizar em categorias os métodos para FaDs, pois é possível que um método situado em uma categoria utilize algumas técnicas de outra categoria. Por exemplo, os métodos baseados em conhecimento e os de comparação de modelos utilizam-se de heurísticas sobre rostos humanos para identificar faces em suas buscas. Este trabalho utiliza o modelo apresentado por Yang, Kriegman e Ahuja (2002), mas outros estudos relacionados podem ser encontrados, como os do site *Face Detection* (FRISCHHOLZ, 2010) e de Hjelmås e Low (2001).

No site *Face Detection*, mantido pelo Dr. Robert Frischholz, encontram-se as técnicas

[11] Um *frame*, em português quadro, é cada uma das imagens estáticas que são exibidas sequencialmente para compor um vídeo. A frequência de *frames* é medida em fps (*frames per second*) e as taxas mais comumente encontradas em filmes, televisão e *video games* são entre 24 e 60 fps.

de FaD divididas em categorias diferentes das definidas por Yang, Kriegman e Ahuja (2002). Estas categorias, apresentadas na sequência, se somam as vistas até o momento.

- **Localização de faces em imagens com fundo controlado:** nesta técnica, a face é extraída de imagens que possuem um fundo uniforme (monocromático, por exemplo). Pela abstração do fundo, encontra-se a região relacionada a uma face;

- **Mistura de técnicas:** a combinação de técnicas é aplicada em alguns trabalhos para obtenção de melhores resultados na busca pela identificação de faces (DARRELL et al., 1998).

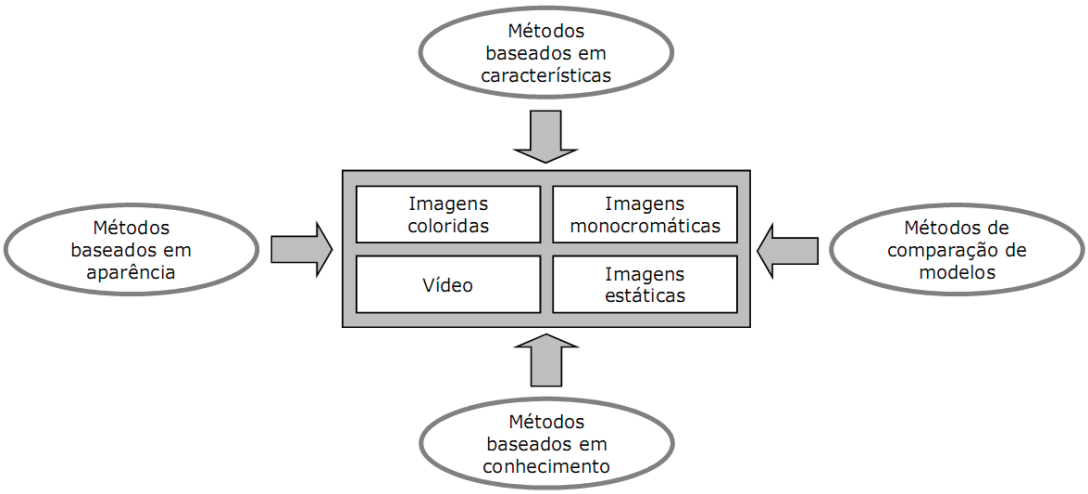


Figura 2.14: Categorias de Yang, Kriegman e Ahuja (2002).

As diferenças encontradas entre as categorizações de métodos para FaD realizadas por Yang, Kriegman e Ahuja (2002) e o site *Face Detection*, são devidas ao fato do primeiro ter buscado agrupar os métodos em torno de atributos genéricos (imagens estáticas, em movimento, a cores), enquanto que o segundo deu maior atenção ao cenário onde a face será extraída (onde as técnicas podem ser mais ou menos complexas em função dele). No modelo de Yang e colegas, as quatro categorias por eles organizadas (métodos baseados em conhecimento, métodos baseados em características, métodos de comparação de modelos e métodos baseados em aparência) são aplicáveis sobre imagens coloridas ou em tons de cinza, estática ou em sequência de vídeo (ver Figura 2.14).

Já, os estudos realizados por Hjelmås e Low (2001) estão concentrados em uma categorização mais hierárquica, mantendo os métodos organizados em função de dois grandes grupos: os que se baseiam na localização de características faciais e os que se orientam sobre a

abordagem baseada em imagens. A Figura 2.15 mostra uma visão de como estão agrupados os métodos organizados por estes autores.

Em geral, as categorizações apresentam similaridades, com exceção da abordagem de localização de faces de fundos controlados, apresentada somente pelo site *Face Detection* (FRISCHHOLZ, 2010).

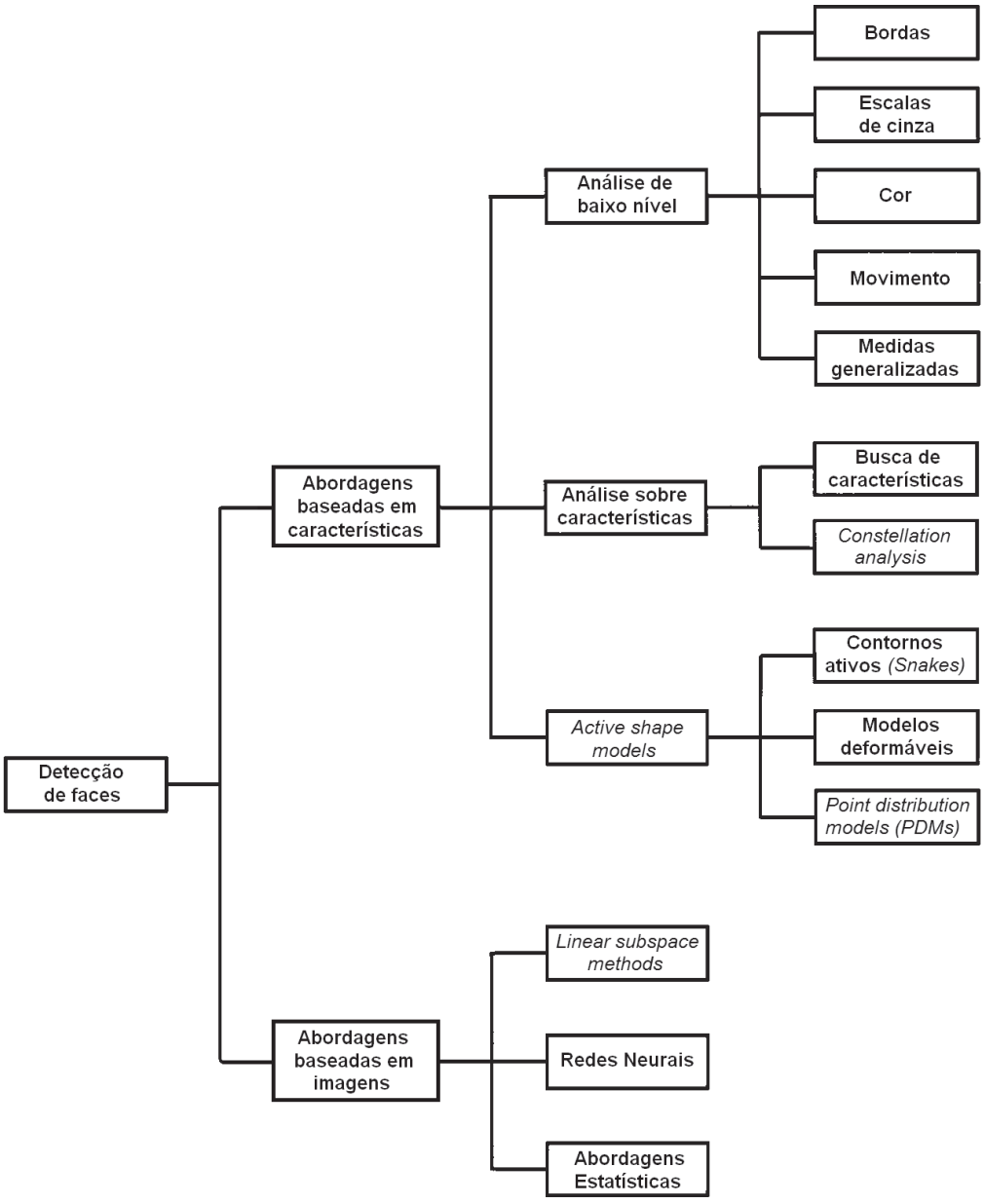


Figura 2.15: Categorização de Hjelmås e Low (2001).

2.2.2 Detecção de Características

A detecção de características faciais (FeD) de rostos humanos é a etapa posterior a detecção da região onde se encontra uma face, sendo essencial para se buscar a identificação

de expressões afetivas. Mesmo existindo técnicas que podem realizar a identificação e extração de características, independente de localização prévia do rosto, realizando a FaD consegue-se otimizar a tarefa de FeD, pois nestes casos, aplicados em vários algoritmos, a região onde se encontram as características reduz consideravelmente (ver exemplo da Figura 2.16). Algumas características faciais são mais relevantes para a tarefa de identificação de emoções, no caso, olhos, boca e sobrancelhas. Para a execução da tarefa de detecção e extração de características, normalmente são utilizadas três categorias de métodos (YANG; KRIEGMAN; AHUJA, 2002; LOPES; FILHO, 2005): (*i*) baseados em conhecimento, (*ii*) baseados em aparência e (*iii*) de comparação de modelos.

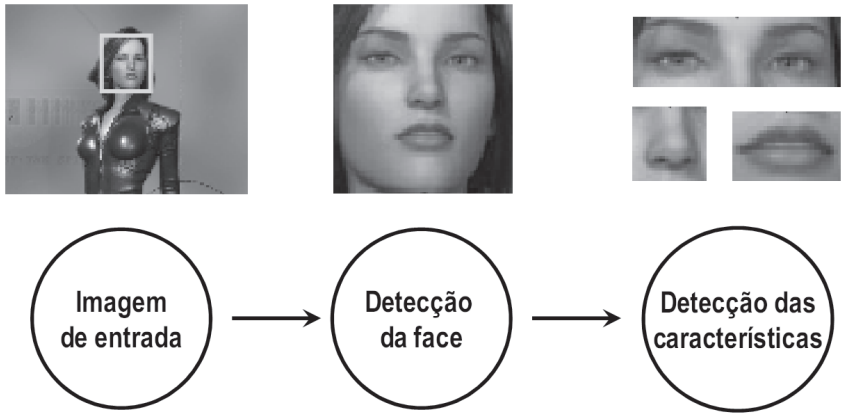


Figura 2.16: Etapas para detecção de características. (Adaptado de Lopes e Filho (2005)).

A tarefa de detecção de características, embora seja comumente realizada após a detecção da face, pode ser realizada independentemente desta etapa. O desempenho de um sistema desenhado para buscar características em uma imagem qualquer, sem ter orientação da existência ou não de um rosto, pode ser muito fraco para alguns métodos, mas para outros, como o caso dos que utilizam comparação de modelos, pode ser bem aceitável. Isto é possível, pois o modelo de um rosto (composto por uma ou mais características faciais para alinhamento do modelo) pode realizar a FeD juntamente com a FaD, já que, quando se encontram as características, se encontra uma face (LOPES; FILHO, 2005).

De modo geral, os métodos que são utilizados para a FaD são os mesmos para a FeD. O que faz com que eles não sejam executados diretamente para FeD é o fato de que as características faciais (que isoladas são compostas por poucos elementos em sua constituição) são difíceis de serem identificadas em imagens de baixa resolução. Isso obriga a utilização combinada de metodologias para facilitar e aumentar a precisão das tarefas de FeD.

2.2.3 Biblioteca OpenCV

A biblioteca aberta para uso acadêmico e comercial de Visão Computacional OpenCV (*Open Source Computer Vision Library*) (BRADSKI; KAEHLER, 2008), lançada pela Intel em 1999, é mantida pela Willog Garage (WILLOW GARAGE, 2010) desde 2008. Esta biblioteca possui mais de 500 funções implementadas em C/C++, que são destinadas à pesquisas em Interação Humano-Computador, identificação de objetos, reconhecimento de faces, reconhecimento de objetos, rastreamento de movimento, dentre outras. Em suas funções, estão implementados diversos algoritmos de Processamento de Imagens e aprendizado de máquina que se encontram divididos em quatro módulos (até a versão 2.1): CV + CVAUX, MLL, HighGUI e CXCORE.

Funções para Processamento de Imagens, como filtros, transformações geométricas, histogramas, detecção de cantos, detecção de bordas, pirâmides, transformações, análise de formas, análise de movimento, detecção de objetos, entre outras, estão implementadas no módulo CV + CVAUX. MLL é o módulo de aprendizado de máquina, onde é possível encontrar algoritmos de classificação estatística, regressão e agrupamento de dados, permitindo a utilização de classificadores SVM, k-NN (*k-Nearest Neighbor*), árvores de decisão, redes neurais (MLP), árvores randômicas, *boosting*, classificador Normal Bayes e algoritmos EM (*Expectation-Maximization*). O módulo de HighGUI implementa funções de Entrada/Saída (E/S) de imagens, E/S de vídeo, eventos de teclado, evento de *mouse* e barras de rolagem. Por fim, em CXCORE estão implementadas funções para cálculo de matrizes, vetores, álgebra linear, operações lógicas, operações aritméticas, funções de desenho e E/S de XML.

OpenCV disponibiliza para detecção de objetos o método de Viola-Jones, empregado neste trabalho para realizar a detecção de faces. A subseção a seguir descreve de forma geral o funcionamento deste método.

Método Viola-Jones

No modelo proposto por Viola e Jones (VIOLA; JONES, 2001), é introduzido um classificador para detecção de objetos, com ênfase em detecção de faces. Este processo ocorre sobre imagens estáticas e em tons de cinza, podendo ser aplicado sobre aplicações em tempo real. Este modelo estrutura-se em três módulos distintos e complementares: (1) a criação da imagem integral, (2) a utilização do algoritmo Adaboost (explicado na Seção 2.2.3) para classificação utilizando *Haar-like features* e (3) a criação de uma estrutura em árvore, chamada de cascata de classificadores.

Nesta abordagem, é apresentado o conceito de imagem integral, que corresponde à

representação da imagem original, onde sobre cada ponto desta representação está contido o somatório da intensidade de pixels de uma imagem original. Nestes cálculos, o valor contido em um determinado ponto, corresponde a soma da intensidade dos pixels (soma das colunas e linhas) de todos os outros pixels acima e a esquerda deste ponto. Utilizando a Figura 2.17 como exemplo, o ponto 1 contém o somatório da área do retângulo A. O ponto 2 corresponde ao somatório dos retângulos A e B, da mesma forma que o ponto 3 é igual a A + C e o ponto 4, A + B + C + D. A soma dos pixels na área do retângulo D é obtido pelo cálculo entre os pontos: 4 + 1 - (2 + 3). Os processos seguintes realizados pelo modelo serão executados com base na imagem integral obtida.

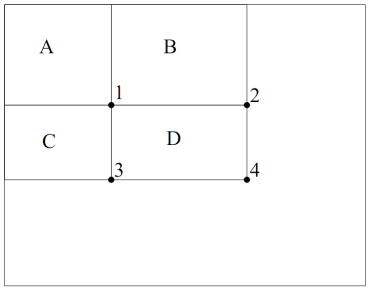


Figura 2.17: Exemplo de cálculos com imagem integral. (De Viola e Jones (2001)).

Características do tipo Haar (*Haar-like features*) são representações retangulares baseadas em *Haar wavelets*. No modelo de Viola-Jones, as características são representadas por retângulos que contêm regiões, sobre as quais é realizada a soma entre as regiões claras, que são subtraídas pelas regiões escuras. Este resultado representará o valor encontrado pela característica para determinada região (Figura 2.18).

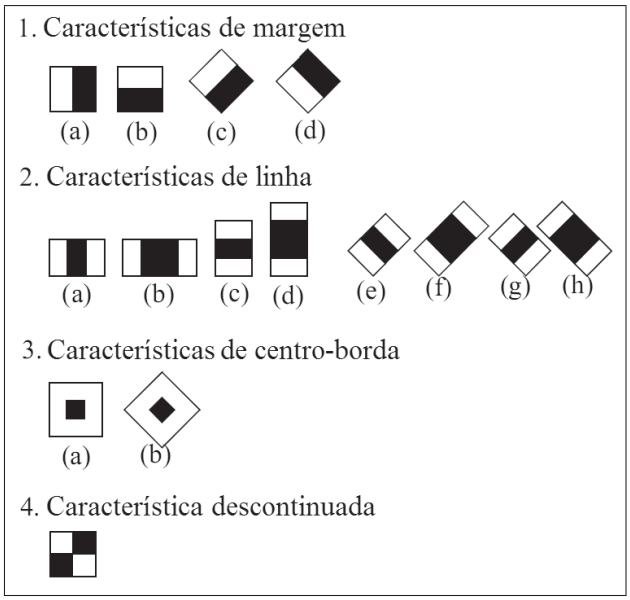


Figura 2.18: Características do tipo Haar (*Haar-like features*). (De Lienhart e Maydt (2002)).

Em seu trabalho, Viola-Jones definiram quatro tipos de características para uso (Figura 2.18, subfiguras 1(a), 1(b), 2(a) e 4). Seguindo propostas sugeridas por Lienhart e Maydt (2002), estas características foram ampliadas, sendo acrescentadas mais algumas (Figura 2.18, subfiguras 1(c), 1(d), 2(b), 2(c), 2(d), 2(e), 2(f), 2(g), 2(h), 3(a) e 3(b)) e retirada uma (Figura 2.18, subfigura 4, que foi substituída pela combinação das subfiguras 2(g) e 2(e)). Nota-se que algumas características aparecem inclinadas à 45 graus, posição adotada para aumentar o desempenho da proposta original. Para se considerar estas características inclinadas no processo, passou-se a calcular, além da imagem integral normal, uma imagem integral inclinada.

O modelo de Viola-Jones utiliza um classificador fraco[12] onde são considerados, para um conjunto de características, um limiar e uma paridade. Este classificador busca encontrar a característica que obtenha o melhor limiar que separa as imagens definidas como positivas e negativas. As imagens que forem classificadas abaixo do limiar sobre os valores de paridade têm hipótese verdadeira atribuída.

Este classificador é submetido a treinamento utilizando o algoritmo Adaboost. Devem ser submetidos à classificação um conjunto de casos positivos e outro conjunto de casos negativos, ambos com a mesma escala para todos os exemplos. Quanto maior o número de exemplos (na ordem de milhares), melhor será o desempenho do classificador.

Obtido o classificador, este sofrerá um processo de otimização para tornar a tarefa de classificação mais rápida. Para isso é realizado um segundo processo de classificação que resulta na construção de uma árvore degenerativa de decisão, chamada de cascata. Nesta cascata, as classificações estão arranjadas em estágios de complexidade crescente. Nos primeiros estágios são utilizados classificadores mais simples (mais genéricos) e não tão precisos, que são sucedidos por classificadores mais específicos e criteriosos nos estágios seguintes. Os casos que são classificados como corretos são submetidos ao próximo estágio, até que o classificador do último estágio tenha feito a classificação corretamente para o caso. Esta estrutura possui a intenção de evitar que testes desnecessários sejam realizados para atestar se um caso realmente é negativo, pois se um classificador mais fraco e genérico não o considera positivo, um classificador mais específico também não o considerará.

A última etapa do processo de criação do classificador no modelo Viola-Jones é o treinamento da cascata de classificadores. Neste treinamento devem ser considerados: taxa de detecção mínima aceitável, a taxa máxima de falsos positivos aceitável de cada estágio, um conjunto de amostras positivas e negativas e os valores de falsos positivos para todos os está-

[12] Um classificador fraco é aquele que, em sua classificação, retorna hipóteses com baixo nível de cobertura (taxa de erro menor que 50%).

gios da cascata (ver Figura 2.19). A cascata, então, é construída com um número de estágios que obtenham, na classificação sobre a amostra, os valores de falsos positivos menores que o definido para a cascata.

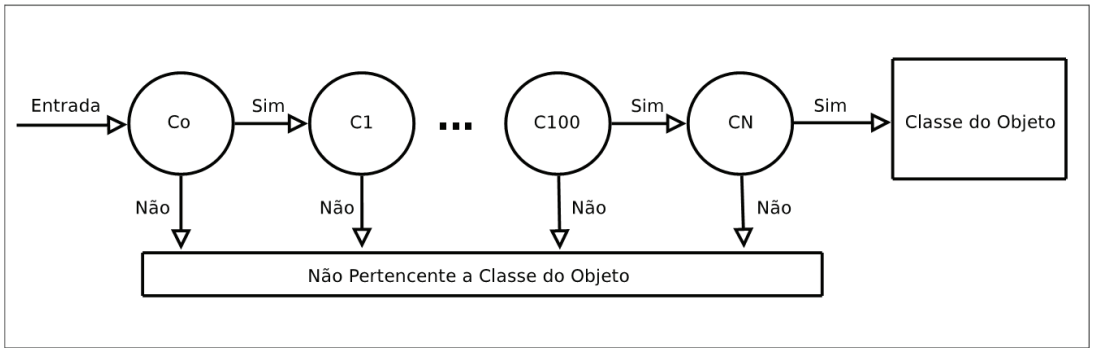


Figura 2.19: Cascata de classificadores com *n* estágios. (De Ma (2007)).

Na utilização da cascata treinada, a classificação é realizada sobre uma subjanela do mesmo tamanho definido no treinamento. Este processo de classificação percorre a imagem a ser explorada e, nesta exploração, são realizados ajustes na escala do classificador para que a face possa ser identificada em imagem de tamanho maior.

A próxima subseção apresenta o algoritmo de *boosting* utilizado por Viola-Jones, o Adaboost.

Adaboost

Adaboost, nome derivado de *Adaptative Boosting*, é um dos mais conceituados algoritmos de *boosting* existentes (FREUND; SCHAPIRE, 1995). *Boosting* é um método de aprendizado de máquina que utiliza a combinação de vários classificadores fracos (*weak learners* - de hipóteses fracas) para obter uma classificação forte (de hipótese forte) (SCHAPIRE, 1990). A intenção de realizar uma classificação com um conjunto de várias classificações fracas, ao invés de realizar uma classificação utilizando um classificador mais forte, é que este conjunto obteria no final uma classificação forte efetiva. Dependendo da quantidade de classificadores fracos combinados, *boosting* pode obter resultados melhores do que se fosse utilizado um único classificador forte.

Durante o aprendizado (treinamento) com o Adaboost, são realizadas várias iterações onde as classificações de um classificador fraco são ponderadas. Cada classificação que é corretamente realizada sobre os exemplos de teste recebe um peso menor, ao passo que as classificações incorretas recebem um peso maior. A cada classificação, o processo atualiza os índices de distribuição de erro. O termo *adaptative*, contido em seu nome, refere-se à distribuição de

pesos que é realizada sobre o desempenho nos testes de aprendizado.

Abaixo são descritas as etapas do algoritmo de classificação Adaboost (para classificação binária e discreta), que é exibido na Figura 2.20).

Dado: S= $\{(x_1, y_1), ..., (x_m, y_m); x_i \in X, y_i \in \{-1, +1\}\}$

Algoritmo *Adaboost*:

Inicialize $D_1(i) := \frac{1}{m}, \forall\ (x_i, y_i) \in S$

For $t = 1, ... T,$ faça

Treine o classificador base usando a distribuição D_t

Obtenha a hipótese fraca $h_t: X \rightarrow \{-1, +1\}$

Calcule α_t: $\alpha_t = \frac{1}{2} ln \frac{1 - e_t}{e_t}$, onde e_t é a taxa de erro do classificador h_t.

Atualize a distribuição:

$$D_{t+1}(i) = \frac{D_t(i)}{Z_t} = \begin{cases} e^{-\alpha_t}, & \text{se } h_t(x_i) = y_i \\ e^{\alpha_t}, & \text{se } h_t(x_i) \neq y_i \end{cases}$$

$$= \frac{D_t(i) \exp(-\alpha_t y_i h_t(x_i))}{Z_t}$$

Onde Z_t é o fator de normalização

End For

Saída: Hipótese Final: $H(x) = sign\left(\sum_{1...T} \alpha_t * h_t(x)\right)$

Figura 2.20: Algoritmo Adaboost. (Adaptado de Souza (2006)).

1. Adaboost recebe um conjunto de teste pré determinado, onde cada exemplo possui uma classificação (caso positivo ou caso negativo). O algoritmo faz *n* iterações, em cada uma utilizando o classificador fraco;

2. A cada iteração, o classificador obtém uma classificação (hipótese) sobre exemplos do teste. O termo que define a regra do classificador para o Adaboost, como padrão, realiza uma classificação binária (caso positivo, caso negativo);

3. Os erros na classificação (erro sobre a hipótese) são calculados com base no conjunto de teste. Ele corresponde ao peso dos falsos positivos e falsos negativos do conjunto de teste;

4. São atribuídos pesos sobre as classificações dos exemplos de teste, sendo que as classificações corretas têm seu peso reduzido (ou, em outro sentido, aumentado quando é mal classificado). Inicialmente, todos os exemplos de treino têm o mesmo peso (1/número de exemplos), mas no fim os casos de classificação mais difícil têm peso maior;

5. Dentro de cada iteração os pesos são normalizados, de forma que a soma de todos os pesos seja igual a 1;

6. No fim, é montada uma combinação de classificadores fracos que apresentaram classificações positivas, sendo as classificações com menores erros as mais destacadas nesta montagem.

2.2.4 Rastreamento de Objetos

Rastreamento de objetos ocorre sobre sequência de imagens e possui entre suas áreas de aplicação o reconhecimento de pessoas, vigilância, IHC, controle de tráfico, entre outros. Yilmaz, Javed e Shah (2006) definem rastreamento como o problema de estimar a trajetória de um objeto que se movimenta, aplicando registros sobre o rastro do objeto ao longo dos *frames* onde ele foi capturado.

Os diversos algoritmos existentes para rastreamento de objetos possuem similaridades e particularidades, o que torna-os mais aplicáveis a problemas específicos. Entre estas especificidades está a forma de representação do objeto a ser rastreado (com pontos, formas geométricas, contornos, silhueta etc) e as características de imagem utilizadas no rastreamento (cor do objeto, suas bordas, o fluxo óptico, textura etc). O objeto a ser rastreado, normalmente necessita de algum método de detecção (pontos, subtração de fundo, segmentação de imagem, classificadores ou regressores etc), o que pode ocorrer no primeiro ou em cada *frame*.

O rastreador de objeto em si, registra a posição e trajetória de um objeto entre os *frames* (YILMAZ; JAVED; SHAH, 2006). Estes rastreadores normalmente se encaixam em três categorias: rastreamento de pontos, que se subdivide em métodos estatísticos e probabilísticos; rastreamento de *kernel*, que tem as subcategorias modelos de aparência baseados em *template* e densidade e modelos de aparência *multi-view* e, por fim, rastreamento de silhueta, com os métodos de evolução de contorno e comparação de formas.

Entre os rastreadores de objetos, algumas implementações encontram-se disponíveis para *download* na Internet. Existem *toolboxes* no Matlab para os métodos estatísticos de rastreamento de pontos Filtro de Kalman[13] e Filtro de Partículas[14], assim como encontram-se disponíveis códigos-fonte para os métodos de rastreamento de *kernel*, tais como o Mean-Shift[15] e o KLT.[16]

[13]Disponível em: `http://www.cs.ubc.ca/~murphyk/Software/index.html`

[14]Disponível em: `http://www-sigproc.eng.cam.ac.uk/smc/software.html`

[15]Disponível em: `http://coewww.rutgers.edu/riul/research/code.html` e na biblioteca OpenCV (WILLOW GARAGE, 2010), denominada como CAMSHIFT.

[16]Disponível em: `http://www.ces.clemson.edu/~stb/klt/`

2.3 Computação Afetiva

A busca por uma relação onde o computador entenda e manifeste emoções é o foco das pesquisas em Computação Afetiva (PICARD, 1995). Este campo de pesquisa da Inteligência Artificial busca fazer com que a emoção existente na comunicação entre pessoas, também esteja presente na relação entre homem e computador. Para a Computação Afetiva, o computador pode ser capaz de interagir com humanos reconhecendo e expressando afeto.

A emoção encontra-se presente em várias ocasiões e manifestações na vida do homem: em tomadas de decisões, nas interações sociais, na inteligência, na criatividade e em outros eventos. Fazer com que um computador tenha habilidades emocionais, trará a ele a possibilidade de ter maior sucesso em tomadas de decisão, capacidade de percepção dos estados emocionais de seu usuário e maior nível de ajuste de seu comportamento. Com isso, o computador poderá se adaptar às pessoas e não o contrário (PICARD, 1997).

Existem algumas propostas de aplicativos que utilizam Computação Afetiva, que são sugeridos e estão sendo desenvolvidos pelo grupo liderado pela pesquisadora Rosalind Picard (M.I.T. Media Labs, 2010), um dos pioneiros em pesquisas nesta área. Entre os exemplos de projetos está o *Affective Mirror*, que se trata de um agente capaz de servir de espelho para uma pessoa, captando suas manifestações e retornando para ela como lhe pareceu o seu comportamento. *Beyond Emoticon* é outro aplicativo que evitaria as comuns más interpretações emocionais expressas em um *e-mail*, pois ele capturaria as expressões emocionais da pessoa e as transmitiria ao destinatário através de animação de um agente e/ou por expressões faciais do remetente junto com o conteúdo da mensagem. No exemplo de *Agents that Learn your Preference* o computador poderá ser capaz de se ajustar conforme as preferências de seus usuários. Como exemplo, o computador exibiria preferencialmente as páginas de esporte, que foram consideradas as favoritas de seu usuário pela análise de sua expressão facial e comportamento observável.

Conforme define Picard (1995), de maneira resumida, pode-se encontrar em Computação Afetiva quatro categorias de reconhecimento e expressão de afeto:

- **Computador não expressa e não reconhece afeto:** caso da maioria dos computadores atuais, que não tem nenhuma capacidade afetiva;

- **Computador que não reconhece, mas expressa afeto:** ocorre em alguns computadores que expressam afeto por emissão de voz e/ou por animações faciais, como acontece em computadores Macintosh que exibem um ***smile*** quando um disquete é inserido;

- **Computador tem capacidade de reconhecer afeto, mas não expressa afeto:** este tipo de computador tem a capacidade de, pela percepção de estados afetivos dos seus usuários, ajustar o seu desempenho para proporcionar um melhor aproveitamento nas interações. Um exemplo desta categoria é a implementação de um professor de piano interativo, que ajusta sua aula conforme as expressões afetivas do aluno, propondo novos e atraentes desafios quando este encontra-se frustrado (por não conseguir executar um exercício, por exemplo) (M.I.T. Media Labs, 2010);

- **Computador tem capacidade de reconhecer e expressar afeto:** nesta categoria, um computador pode alcançar um nível de interação amigável avançada, ajustando-se completamente aos estados expressos pelo usuário. Um exemplo é Pat, um agente pedagógico animado que infere as emoções de um aluno pelo seu comportamento observável utilizando-se, para tal, de um modelo psicológico cognitivo de emoções. Com base nestes dados, Pat adapta o sistema a possíveis dificuldades nas tarefas dos alunos, evitando sua desmotivação e posterior abandono de suas tarefas escolares. O agente expressa suas emoções ao aluno falando e movimentando-se na tela do computador (JAQUES; VICCARI, 2005b).

2.3.1 Emoções

A utilização do termo **emoção** acontece muitas vezes de forma desmedida, mas, conceitualmente, emoção é considerada como um elemento do conjunto genérico de estados afetivos, no qual também se encontra o humor, entre outros (JAQUES; VICCARI, 2005a). Ao contrário do humor, que costuma ter uma duração mais longa (horas, dias) e não tem uma causa bem definida, a emoção é normalmente breve (minutos) e ocorre em função de um estímulo interno ou externo (JAQUES; VICCARI, 2005a apud SCHERER, 2000a), (PICARD, 1997). Neste contexto, uma expressão emocional é aquilo que é demonstrado a outras pessoas, voluntária ou involuntariamente (PICARD, 1997).

Embora não exista uma consolidação quanto à definição de emoções, estas podem ocupar uma lista de mais de vinte tipos (PICARD, 1997). Existem diversas teorias de emoções (SCHERER, 2000b), como os modelos dimensionais que se baseiam em duas principais categorias, *arousal* (calmo/excitado) e valência (negativo/positivo), para diferenciar as emoções. Uma outra teoria de emoções bastante difundida é o modelo de emoções básicas, que recebem este nome por terem as mesmas manifestações corporais em diferentes culturas. Ao todo,

são seis expressões faciais emocionais básicas que foram constatadas por Ekman (1999) como estando presentes desde a infância em crianças de qualquer parte do mundo (Figura 2.21).

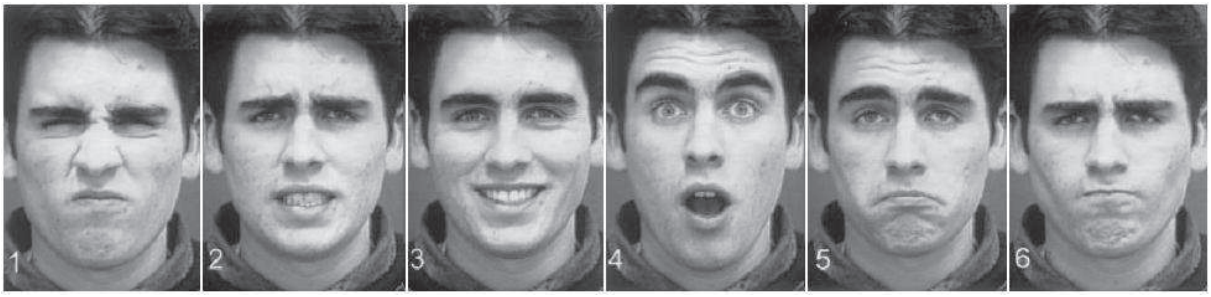


Figura 2.21: Seis expressões faciais emocionais básicas: (1) repulsa, (2) medo, (3) alegria, (4) surpresa, (5) tristeza e (6) raiva. (De Schmidt e Cohn (2001)).

Atualmente, o modelo de componentes (*componential model*) tem recebido considerável atenção dos pesquisadores em emoções. Segundo este modelo, as emoções em humanos são caracterizadas pela presença de quatro componentes principais (CLORE; ORTONY, 1999): (*i*) componente motivacional-comportamental: diz respeito às inclinações de um indivíduo para agir de acordo com estas interpretações; (*ii*) componente subjetivo: responsável pela parte de "sentimento subjetivo" e é mais elaborado em seres humanos que estão habituados a rotular as emoções que sentem; (*iii*) componente somático: envolve a ativação dos sistemas nervosos central e automático e sua manifestação corporal; e (*iv*) componente cognitivo: correspondente aos processos cognitivos que avaliam as situações e disparam as emoções.

O modelo de emoções básicas, originado das expressões comportamentais humanas, apresenta uma grande quantidade de trabalhos relacionados (os estudos sobre FACS foram baseados neste modelo - ver Seção 2.3.2), principalmente no que se refere ao reconhecimento computacional de emoções através de expressões faciais. O modelo de componentes, principalmente os modelos cognitivos que se interessam no componente cognitivo de emoções (ORTONY; CLORE; COLLINS, 1988), tem atraído crescente atenção nos últimos anos, porém sendo mais utilizado para inferência de emoções de usuários através de suas ações na interface do sistema computacional.

2.3.2 Sistema de codificação facial FACS

Estudos de Paul Ekman e Wallace V. Friesen sobre o comportamento facial resultaram na construção do sistema *Facial Action Coding System* - FACS (EKMAN; FRIESEN; HAGER, 2002a). Este sistema categoriza todas as ações faciais causadas por contrações musculares (um

ou mais músculos) em *Action Units* (AU), que, com ou sem combinações, representam todas as expressões faciais possíveis, incluindo sua intensidade, duração e simetria.

FACS é composto por 44 *action units* responsáveis pela descrição de ações faciais, que se dividem em duas regiões faciais: superior, onde são considerados os olhos (pálpebras), sobrancelhas e testa; e inferior, onde são consideradas as bochechas, queixo, nariz e boca (lábios). Destes 44 AUs, 30 estão relacionados a ação de músculos específicos e 14 AUs não têm as suas ações musculares especificadas, ou seja, não têm descrição exata sobre seu comportamento (caso do AU 19 - colocar a língua para fora da boca).

Além dos 44 AUs utilizados nas descrições de ações musculares faciais, existem outros códigos que auxiliam e complementam estas ações. Estes códigos estão agrupados em relação a posições de cabeça, posições de olhos, movimento de olhos, visibilidade das características faciais, comportamentos primitivos[17] e movimentos de cabeça. A lista completa de todos AUs e códigos auxiliares está disponível no Apêndice A, que contém a descrição com exemplos (imagens) para os AUs.

Os estudos sobre FACS, iniciados no final da década de 70, foram expandidos na década de 80 para a criação de outro método, o *Emotion* FACS - EMFACS (FRIESEN; EKMAN, 1983), que mapeia e seleciona os AUs utilizados na manifestação de emoções, descartando os demais. Algumas combinações de AUs estão presentes em expressões faciais que ilustram as emoções básicas (EKMAN; FRIESEN; HAGER, 2002b). Conhecidos os AUs contidos em uma expressão facial, é possível obter a emoção por eles representados. Por exemplo, a emoção de alegria é caracterizada pela presença dos AUs 6 (levantar maçãs do rosto) e 12 (levantar cantos dos lábios), conforme apresentado na Tabela 2.1 e na Figura 2.22.

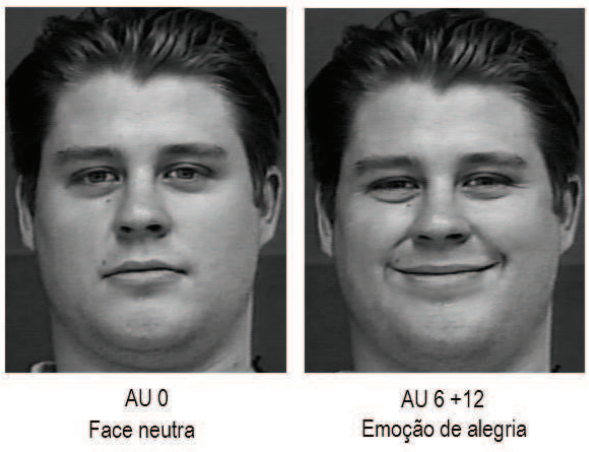


Figura 2.22: Exemplo de composição de AUs na representação da emoção de alegria.

[17]Do inglês *gross behaviors*, como, por exemplo, cheirar, falar, engolir, mastigar, dar os ombros, balançar a cabeça afirmativamente.

Tabela 2.1: Relação entre emoções e AUs. (Adaptado de Ekman, Friesen e Hager (2002b)).

Emoção	Protótipos	Maiores variantes
Surpresa	1+2+5B+(26,27)	1+2+5B 1+2+(26,27) 5B+(26,27)
Medo	1+2+4+5*+20*+(25,26,27) 1+2+4+5*+(25,26,27)	1+2+4+5*+(L20*,R20*)+(25,26,27) 1+2+4+5* 1+2+5Z+[25,26,27] 5*+20*+[25,26,27]
Alegria	6+12* 12C,12D	
Tristeza	1+4+11+15B+[54+64]+[25,26] 1+4+15*+[54+64]+[25,26] 6+15*+[54+64]+[25,26]	1+4+11+[54+64]+[25,26] 1+4+15B+[54+64]+[25,26] 1+4+15B+17+[54+64]+[25,26] 11+15B+[54+64]+[25,26] 11+17+[25,26]
Repulsa	9 9+15+(16,26) 9+17 10* 10*+16+(25,26) 10+17	
Raiva	4+5*+7+10*+22+23+(25,26) 4+5*+7+10*+23+(25,26) 4+5*+7+23+(25,26) 4+5*+7+17+(23,24) 4+5*+7+(23,24)	Quaisquer protótipos sem um dos seguintes AUs: 4, 5, 7 ou 10.

Nota: "*" significa que código pode ocorrer com qualquer nível de intensidade.

A Tabela 2.1 é constituída pelas colunas emoção, protótipos e maiores variantes. Emoções, são as seis básicas, protótipos são considerados as principais formas de combinação de AUs para cada emoção, já maiores variantes são formas alternativas de combinação de AUs para expressar uma emoção que costumam estar relacionadas a referências simbólicas (descritiva) de emoção e não necessariamente a uma emoção sentida. Também existem letras que precedem (prefixos) e/ou sucedem (sufixos) os códigos, que tratam-se de referências adicionais opcionais do sistema FACS que auxiliam e complementam as informações sobre uma ação facial. Os sufixos (A, B, C, D, E, X, Y, e Z) indicam a intensidade dos movimentos e os prefixos (L e R), a simetria do movimento. O simbolo "+" tem a função do operador lógico AND, indicando que ambos códigos são necessários para que seja considerada existente uma emoção. Já o símbolo "," tem a função do operador lógico OR e os colchetes contêm códigos opcionais, ou seja, que podem ou não estarem presentes na combinação.

Segundo a definição dos autores, é preciso salientar que esta tabela foi criada utilizando elementos que evidenciam que um conjunto de AUs representa uma emoção, mas estas evidências não são completas. São muitas as considerações que devem ser tomadas na avaliação da expressão de uma emoção, como questões culturais, ambiente, experiências e comportamento

da pessoa. As evidências encontradas são mais frágeis em se tratando da diferenciação entre a emoção de surpresa e medo e, em escala menor, em relação as definições da metade inferior da face em relação a emoção de tristeza.

MPEG-4

Moving Picture Experts Group (MPEG) é o grupo de trabalho da ISO/IEC, constituído em 1988 e formado por pesquisadores da academia e da indústria. A função deste grupo é desenvolver para indústria padrões de representação de dados digitais (compressão, descompressão, processamento e codificação) sobre áudio, vídeo e assuntos relacionados (MPEG, 2002).

O MPEG iniciou em 1993 o seu terceiro projeto, o padrão ISO/IEC MPEG-4 (os anteriores eram o MPEG-1[18] e MPEG-2[19]). O título do projeto naquele momento era *Very low bit rate audio visual coding* e tinha foco em comunicações em rede. Essa primeira versão foi concluída em 1998, porém, foram incluídas no escopo informações sintéticas áudio-visuais, alterando o título do projeto para *Coding of audio visual objects*, que tratou da segunda versão, aprovada em 1999. MPEG-4 é considerado como um padrão multimídia web móvel e fixa, mas devido a sua abrangência, pode ter aplicações em animações e em IHC. Atualmente, o padrão MPEG-4 encontra-se em desenvolvimento e é composto por 27 partes, em que cada uma trata de um assunto específico. A animação facial, que é de interesse específico nesse trabalho, possui especificações tanto na parte 1 (*Visual*), quanto na parte 2 (*Systems*) (PANDZIC; FORCHHEIMER, 2003).

MPEG-4 permite a codificação de expressões faciais através de parâmetros chamados de *Facial Animation Parameters* (FAPs). Existem 68 FAPs em MPEG-4 que representam as expressões faciais, como o FAP 19 (close_t_l_eyelid), que representa o deslocamento vertical da pálpebra superior esquerda. Por sua vez, estes FAPs são representados por 84 pontos no modelo facial, conhecidos por *Features Points* (FPs) (TEKALP, 2000). O FP 3.1 é o que representa os deslocamentos da pálpebra superior esquerda, que é o FAP 19.

Uma possibilidade existente em MPEG-4 é a realização de ações análogas aos AUs em animações faciais (RAOUZAIOU et al., 2002). Por exemplo, o AU 1 equivale à manifestação de FAP 31 e FAP 32. Desta forma, utilizando-se de um mapeamento entre AUs e FAPs, ou simplesmente os FAPs, é possível realizar a classificação de expressões faciais (IOANNOU et al., 2007).

Analogamente à tarefa de ajustar os códigos dos FAPs para a obtenção de AUs e expres-

[18]Padrões para armazenamento de filmes e áudio, onde Vídeo CD e MP3 são incluídos (concluído em 1992).
[19]Padrões para TV Digital, incluindo DVD (concluído em 1994).

sões emocionais, é possível aplicar o modelo de animação facial de MPEG-4 sobre uma face para obtenção de FPs, FAPs, AUs e expressões emocionais. Para isso, é necessário posicionar os pontos do modelo MPEG-4 sobre imagens de face e converter os estados das características faciais para a codificação dos FAPs. O benefício dessa abordagem é poder obter a expressão facial e também realizar a animação correspondente da face, porém o objetivo do MPEG-4 não é inferência de emoções, mas fornecer métodos para comunicação multimídia (entre elas animação facial).

2.3.3 Reconhecimento computacional de emoções em expressões faciais

São várias as formas de uma pessoa manifestar suas emoções e mais variados ainda são os métodos para captar e reconhecer a emoção transmitida. A voz, as ações do usuário na interface com o sistema, as expressões faciais e os sinais fisiológicos são considerados os principais modos de reconhecimento de emoções (JAQUES; VICCARI, 2005a), conforme esboçado na Figura 2.23.

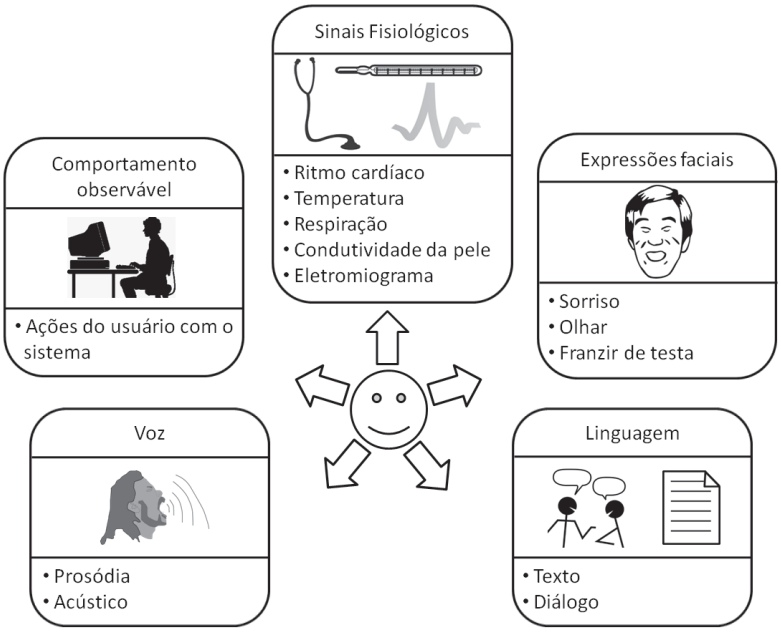


Figura 2.23: Mecanismos de reconhecimento de emoções. (Adaptado de Jaques e Viccari (2005a)).

Os métodos atuais de reconhecimento de emoções de uma pessoa pelo computador se aproximam e, em alguns casos, superam o reconhecimento humano. Enquanto o reconhecimento de expressões faciais por humanos é de aproximadamente 87%, alguns algoritmos computacionais, em ambiente controlado, obtêm sucesso entre 74% e 98% (SEBE et al., 2005). No reconhecimento vocal existe equilíbrio entre humanos e computadores, por volta de 65%,

entretanto alguns algoritmos alcançaram o nível de quase 80% de acerto (SEBE et al., 2005). Um problema que pode ocorrer na identificação de emoção pela voz é a interferência de sons externos (ruídos) na captação sonora, caso se trate de ambiente não controlado. Sincronizando a movimentação labial com fala (mecanismo utilizado na percepção humana), é possível obter uma redução dos efeitos de ruído. Resultados ainda melhores de reconhecimento de expressões emocionais podem ser obtidos utilizando a combinação de mecanismos de reconhecimento, como facial e vocal, que são considerados como principais aspectos utilizados por uma pessoa para reconhecer emoções.

Ocorre em Visão Computacional uma confusão frequente entre reconhecimento de expressões faciais e reconhecimento de emoções humanas (FASEL; LUETTIN, 2003). Para o reconhecimento de expressões faciais, são necessários dados sobre ações de características faciais, que são extraídos basicamente de imagens. Já, para o reconhecimento de emoções, é preciso considerar várias condições, como variações de voz, de pose, gestos, direções de olhar e expressões faciais. Uma análise apenas da expressão labial, por exemplo, não tem como concluir se um sorriso refere-se realmente a uma emoção de alegria ou é apenas uma pose, mas fornece subsídios que podem aumentar essa possibilidade. Uma pessoa pode tentar expressar e convencer uma emoção que não sente, mas alguns músculos faciais acionados de determinado modo, somente quando algum tipo verdadeiro de emoção é manifestado, podem desmentir essa tentativa (EKMAN, 1993). Conforme ilustra a Figura 2.24, uma expressão facial é composta por vários elementos, sendo um destes o sentimento da emoção.

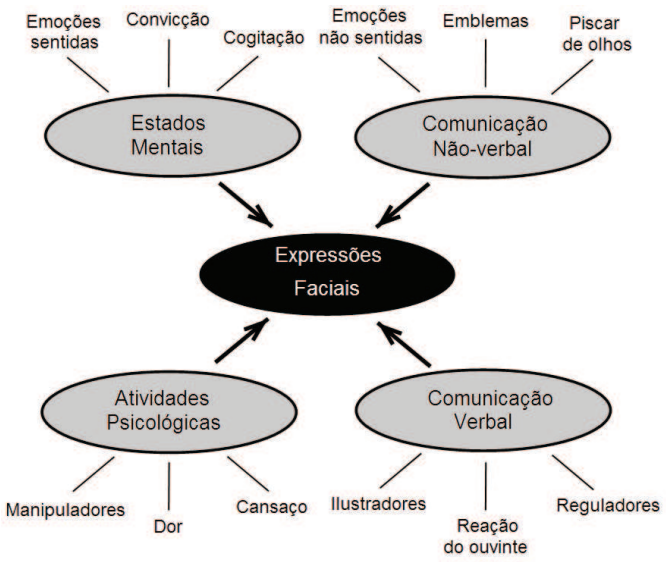


Figura 2.24: Elementos que compõem expressões faciais. (Adaptado de Fasel e Luettin (2003)).

Algumas metodologias são utilizadas pelos pesquisadores para que a identificação de expressões faciais e a posterior classificação sejam realizadas pelo computador. Inicialmente, é

necessário encontrar a face humana em uma imagem e esta tarefa pode se tornar não trivial devido a ocorrência de alguns fatores negativos, conforme visto na Figura 2.13. Após a obtenção da região onde se encontra um rosto, o desafio seguinte é localizar as características faciais relevantes numa expressão facial (como boca, olhos etc), que, neste caso, é realizado por técnicas de FeD similares ou iguais a FaD. Após isto, outros desafios são apresentados, por exemplo: "como classificar o que uma expressão facial demonstra?". Neste caso, muitos pesquisadores utilizam direta (MARTIN et al., 2005) ou indiretamente (BATISTA; GOMES; CARVALHO, 2006) as codificações definidas em FACS para a classificação de emoções (ver Capítulo 3).

3 TRABALHOS RELACIONADOS

Foram encontrados durante o levantamento bibliográfico, vários trabalhos que tratam sobre a identificação de expressões faciais emocionais pelo computador. Em sua maioria, tentam inferir se a expressão facial realizada por uma pessoa se ajusta entre uma das seis expressões básicas (alegria, repulsa, raiva, tristeza, medo e surpresa). São adotadas, para isto, duas abordagens principais: classificação direta de emoções sobre face e detecção de ações faciais, normalmente utilizando o sistema FACS, para posterior classificação da emoção. Este capítulo visa apresentar estes trabalhos relacionados.

3.1 Reconhecimento de expressões faciais básicas por redes neurais

Kobayashi e Hara (1991) desenvolveram um sistema (a ser utilizado por um robô) que objetiva o reconhecimento de emoções humanas pela categorização, por redes neurais, de expressões faciais. As emoções são retiradas de pontos (chamados de FCP - *Facial Characteristic Point*) situados em três características faciais: sobrancelhas, olhos e boca, conforme apresentado pela Figura 3.1.

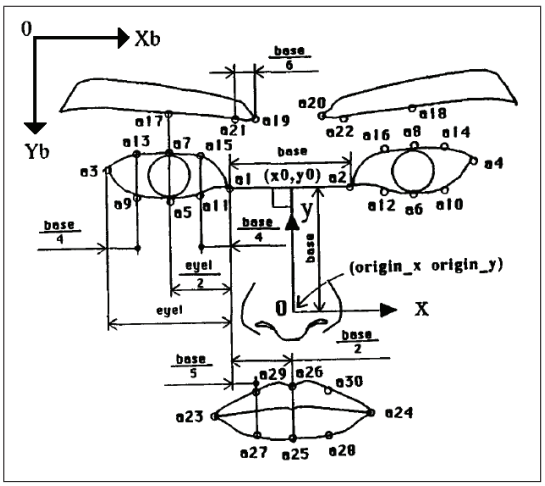


Figura 3.1: Modelo de FCP. (De Kobayashi e Hara (1991)).

Os autores utilizaram as seis expressões básicas de emoção como categorias de expressões emocionais, que foram extraídas levando em consideração as alterações apresentadas na disposição de 30 pontos (marcados manualmente) agrupados sobre as três características. Os 30 pontos são agrupados por funções em 21 formas de expressões de dados, utilizados para explicar as características faciais, que servirão de entrada para o reconhecimento da expressão. Uma rede neural (Figura 3.2), empregando o algoritmo *back propagation*, foi treinada utilizando majoritariamente imagens coletadas de usuários que tiveram suas expressões faciais filmadas para realizar a identificação das seis emoções básicas. Ela realiza a categorização tanto sobre os 30 pontos, quanto sobre 21 dados de informação facial. Ambos modos de execução obtêm as seguintes taxas de reconhecimento: 91,2% sobre os 21 dados e 87,5% sobre os 30 pontos.

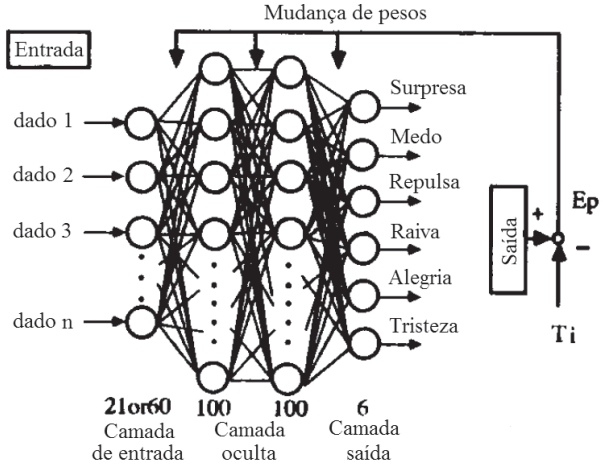


Figura 3.2: Modelo de rede neural utilizada por Kobayashi e Hara (1991).

3.2 AFA

No trabalho realizado por Tian, Kanade e Cohn (2001), foi desenvolvido o sistema AFA (*Automatic Face Analisys*), capaz de detectar os AUs considerados neste sistema com sucesso de até 97%. Para isto, são analisadas as alterações sobre dois tipos de características faciais: permanentes (sobrancelhas, olhos e boca) e transientes (sulcos e rugas causados pelas expressões).

AFA utiliza duas redes neurais que classificam até 16, dos 30 AUs contidos em ações faciais, a partir de dados extraídos das características faciais. Uma rede neural trabalha sobre dados da região da face superior e outra sobre dados da face inferior. Os resultados dessas classificações são combinados para obtenção do possível conjunto de AUs presente na imagem.

A face é detectada automaticamente e, após, as características faciais permanentes são

rastreadas por cor, forma e movimento. Quando localizadas, *templates*, que fornecem os estados dessas características, são ajustados manualmente, no primeiro *frame*, sobre sua região. As características faciais transientes são localizadas utilizando como parâmetro as posições das características faciais permanentes. Após sua localização, estas são analisadas utilizando algoritmos de detecção de bordas. A Figura 3.3 exibe as etapas de execução do sistema.

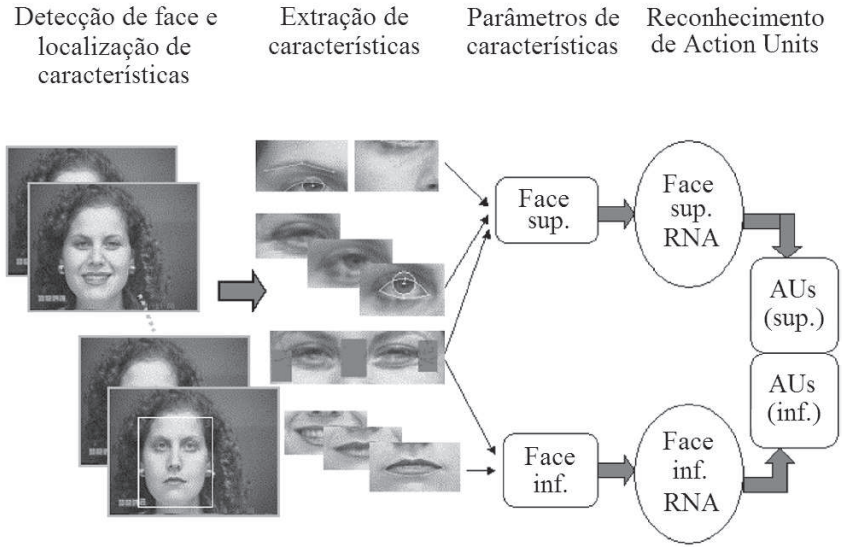


Figura 3.3: Modelo do sistema AFA - *Automatic Face Analysis*. (De Tian, Kanade e Cohn (2001)).

3.3 Reconhecimento de ações faciais sobre imagens estáticas

Com base em 19 pontos referenciais, que contornam componentes de faces em posição frontal, e/ou utilizando 10 pontos, da mesma face em perfil, o sistema desenvolvido por Pantic e Rothkrantz (2004) obtém até 86% de acerto na identificação de AUs. Sobre imagens em perfil, são detectados até 24 AUs e, em imagens frontais, até 22 AUs, totalizando 32 AUs distintos.

Na primeira etapa de execução do sistema, a face (tanto em perfil, quanto frontal) é localizada realizando busca pela cor de pele. Identificada a face, o processo para a localização dos componentes faciais é executado independentemente sobre a face em perfil e frontal. Para a face em perfil, são extraídos 10 pontos da imagem utilizando funções que calculam as extremidades encontradas no contorno da face (picos e vales). Já para a imagem frontal, os 19 pontos referenciais correspondem a vértices no contorno dos componentes faciais localizados (boca, olhos, sobrancelhas, narina e queixo). Estes componentes são previamente identificados utilizando vários processos e algoritmos combinados (*templates*, detecção de bordas e contornos, redes neurais, classificadores baseados em regras).

Obtidos os pontos referenciais, a próxima etapa consiste na obtenção de parâmetros intermediários, que são resultado da diferença entre os pontos referenciais da face neutra com a face com expressão. Na última etapa, são utilizadas duas tabelas que definem os AUs com base em regras, uma para a face em perfil e outra para a face frontal. Nestas regras (estabelecidas com base no sistema FACS) são testadas determinadas condições sobre os parâmetros intermediários, que poderão determinar a existência de um AU.

A Figura 3.4 ilustra como estão estruturados os processos do sistema.

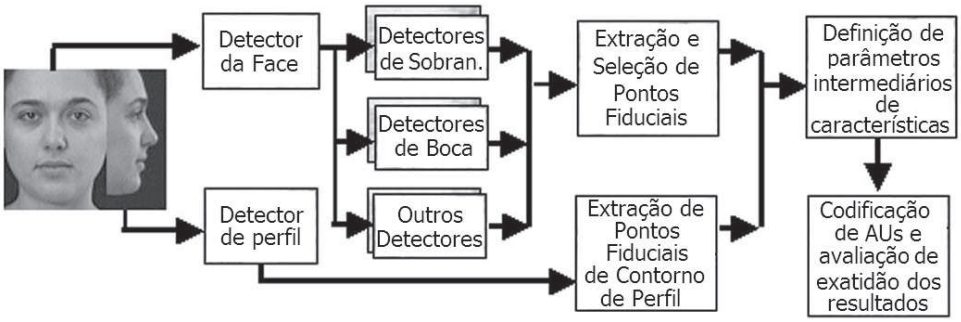


Figura 3.4: Modelo do sistema proposto por Pantic e Rothkrantz (2004).

3.4 Espelho caricato multimodal

Foi construído por Martin et al. (2005) um sistema que exibe por um avatar as expressões faciais e a voz extraídas do usuário que interage com o computador. Este trabalho foi dividido em três etapas: na primeira, foram realizadas as extrações e manipulações de dados visuais; na segunda, foram extraídos dados vocais; e, por último, foi realizada a integração entre as etapas anteriores.

Na primeira etapa, após a detecção da face, realizada por técnicas da biblioteca OpenCV sobre o primeiro *frame*, é efetuada a detecção das características faciais. Funções de intensidade de brilho indicam a região onde se encontram as características, que têm seus cantos marcados pela identificação do primeiro pixel mais escuro das áreas mais extremas (ainda sobre a imagem do primeiro *frame*). De posse dos pontos referentes aos cantos das características faciais, o algoritmo de Kanade-Lucas-Tomasi (KLT) rastreia os deslocamentos das características. Estes pontos referentes às características rastreados anteriormente são depois adaptados ao modelo de face deformável Candide (AHLBERG, 2001) - o avatar.

Tomando como base coordenadas geométricas correspondentes à expressões faciais fornecidas através de Candide, a classificação de seis expressões faciais básicas é realizada por SVM. Na sequência, o avatar pode ser utilizado de duas formas: exibindo as expressões faciais

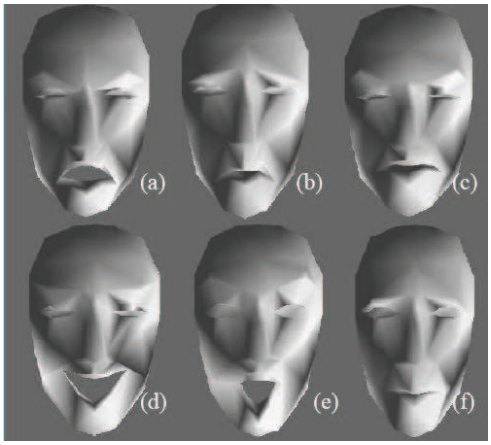


Figura 3.5: Expressões de raiva (a), tristeza (b), repulsa (c), alegria (d), surpresa (e) e medo (f) realizadas por Candide3. (De Martin et al. (2005)).

captadas do usuário ou exibindo apenas as emoções básicas quando estas forem captadas (Figura 3.5). Em uma última etapa, a voz do usuário é extraída, sintetizada e sincronizada com os movimentos labiais do avatar.

3.5 Discriminação de expressões faciais fotogênicas

No trabalho de Batista, Gomes e Carvalho (2006), foi criado um sistema para discriminar expressões faciais divididas em duas classes: fotogênicas e não fotogênicas. Foram consideradas imagens fotogênicas, as que apresentavam expressões faciais neutras ou de alegria (Figura 3.6).

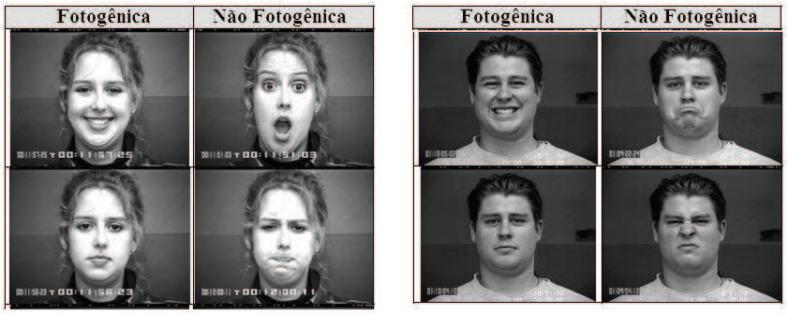


Figura 3.6: Imagens fotogênicas e não fotogênicas. (De Batista, Gomes e Carvalho (2006)).

O modelo criado neste trabalho é dividido em seis blocos, apresentados na Figura 3.7. Foram testadas diversas combinações de técnicas para extração de características e reconhecimento de fotogenia com base nestas características. Primeiramente, as características foram extraídas com filtros de Gabor (*Gabor filters* - GF) e a classificação foi realizada por SVM e por k-NN. Em um segundo experimento, as características foram extraídas por PCA e classificadas

por SVM e por *Multi-Layer Perceptron* (MLP). As comparações de desempenho demonstraram que as melhores combinações entre extratores de características e classificadores de fotogenia ocorreram com as técnicas de PCA com MLP e GF com SVM, tendo a primeira, maior aproveitamento entre as duas.

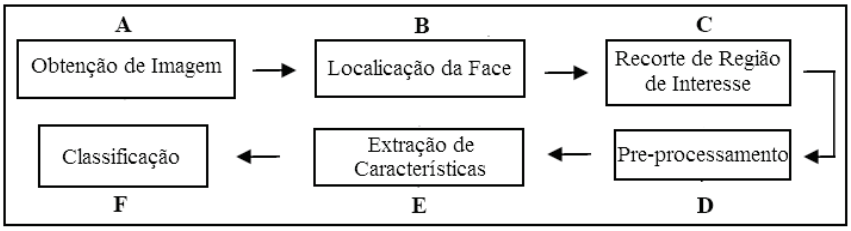


Figura 3.7: *Framework* para discriminação de imagens. (De Batista, Gomes e Carvalho (2006)).

3.6 Sistema automático de detecção de AUs sobre tempo

Valstar e Pantic (2006) propõem um sistema totalmente automático que realize a detecção da manifestação de 15 AUs do modelo FACS. Diferentemente de outros trabalhos, que consideram a manifestação de AUs como um evento estático, neste são consideradas as variações que ocorrem sobre o tempo em sua manifestação. A ativação de um AU ocorre em 4 fases: arranque, onde os músculos começam a contrair; ápice, momento onde o pico da manifestação de um AU ocorre; recuo, quando os músculos começam a relaxar; neutra, onde não há manifestação de ação facial. Normalmente, estas fases ocorrem na sequência neutra-arranque-ápice-recuo-neutra, mas existe a possibilidade de variação nesta sequência. Os autores consideram que a detecção de AUs considerando estas fases é mais confiável para análises de expressões faciais, pois fornece a dinâmica da manifestação de expressões faciais.

O sistema é constituído pelas etapas macro de (*i*) extração de características faciais e (*ii*) análise dos AUs. Na primeira grande etapa é realizada, inicialmente, a detecção da face utilizando um detector *Haar-like features*. Em seguida, os centros dos olhos e da boca são localizados por projeções integrais e verticais da intensidade dos pixels na imagem da face. Os centros dos olhos e boca são a base para a localização de 20 regiões de interesse. Filtros de Gabor e classificador GentleBoost são treinados e aplicados sobre cada uma das 20 regiões de interesse para detectar a posição correta dos pontos sobre as características faciais (ver Figura 3.8). Estes 20 pontos são posicionados sobre o primeiro *frame* da sequência de imagens, que assume-se não possuir nenhuma expressão facial. Os 20 pontos são, então, rastreados (*traking*) utilizando o algoritmo *Particle Filtering with Factorized Likelihoods* e sobre cada *frame* são calculadas características que são baseadas nas diferenças entre a face sem expressão (primeiro

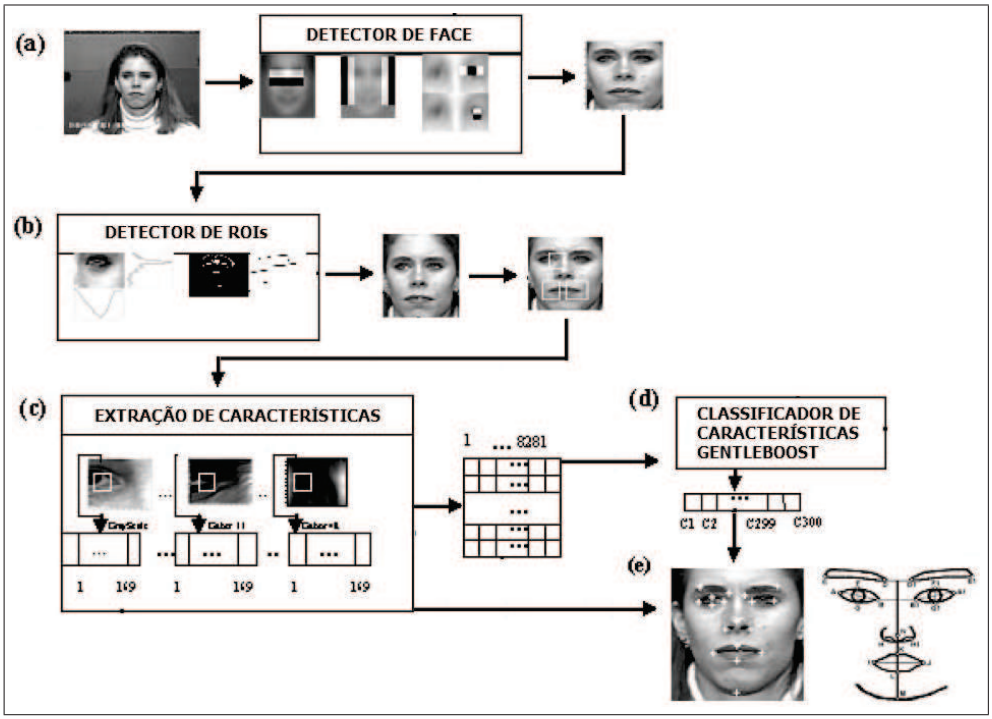


Figura 3.8: Método para detecção dos pontos sobre a face. (a) Detecção da face (*Haar-like features*); (b) extração de regiões de interesse; (c) extração de características (filtros de Gabor); (d) seleção e classificação de características (GentleBoost); (e) face com os pontos detectados ao lado do modelo. (De Valstar e Pantic (2006)).

frame) e os demais, totalizando 840 características por *frame*. Já, na segunda grande etapa, estas 840 características são submetidas a um classificador GentleBoost que classifica em ordem de relevância cada característica, que é submetida a SVMs construídos para cada um dos 15 AUs a serem detectados. Após, outro SVM multi-classe, composto por 6 subclassificadores, realiza a classificação temporal dos AUs (neutra-arranque-ápice-recuo).

O sistema construído, que apresenta taxa de reconhecimento médio de 90,2% sobre os 14 AUs, obtém dados sobre a ativação temporal dos AUs considerando as variações nos 20 pontos. Segundo os autores, por considerar apenas estes pontos, o tempo de duração da ativação dos AUs é maior do que ocorre de fato. Isto seria corrigido caso fosse considerada a variação na aparência das características. Uma outra limitação do sistema é que a operação ocorre somente sobre faces frontais.

3.7 Modelo analítico baseado em pontos para classificação de expressões faciais

É encontrado no trabalho de (SOHAIL; BHATTACHARYA, 2007) um sistema que utiliza variações na distância entre pontos sobre características faciais como dados para a classifi-

cação das seis emoções básicas. Este sistema, que é automático e opera sobre imagens estáticas, obtém a face e o centro dos olhos utilizando o método baseado em *Haar-like features*. Pelo centro dos olhos, um modelo antropométrico é aplicado, fornecendo o centro do nariz, boca e sobrancelhas, que são o ponto de partida para a demarcação da região sobre estas características. A Figura 3.9 ilustra as etapas até aqui descritas.

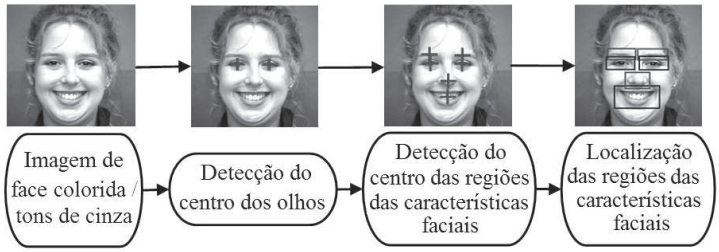


Figura 3.9: Localização da região das características faciais de interesse. (De Sohail e Bhattacharya (2007)).

Nas regiões das características faciais demarcadas são aplicados vários métodos de Processamento de Imagens, até a obtenção das bordas dos olhos, sobrancelhas e boca. Por estas bordas são localizados os pontos de interesse (Figura 3.10(a)). Somente para o nariz um método diferente é aplicado, filtro Laplaciano de Gaussiano, o qual possibilita a localização das narinas.

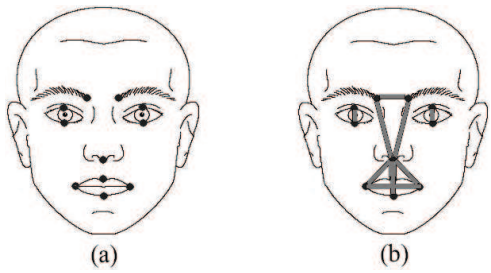


Figura 3.10: (a) Pontos para captura de ações faciais; (b) distâncias consideradas pelos classificadores SVM. (De Sohail e Bhattacharya (2007)).

Após localizados, as distâncias entre estes pontos são submetidas como características para classificação pelos SVMs (Figura 3.10(b)). Foi adotada a estratégia *pair-wise* com 15 SVMs, que foram construídos utilizando como *kernel Radial Basis Function*. Como resultado, esta construção obteve taxas médias de reconhecimento de 89,44% em uma base de faces e 84,86% em outra.

3.8 Detecção automática de AUs e suas relações dinâmicas

Tong, Liao e Ji (2008) propõem um sistema que reconhece AUs automaticamente e em tempo real. O sistema é composto por duas etapas: na primeira é realizado o treinamento do

sistema e na segunda ocorre a execução do processo de reconhecimento de AUs (Ver Figura 3.11).

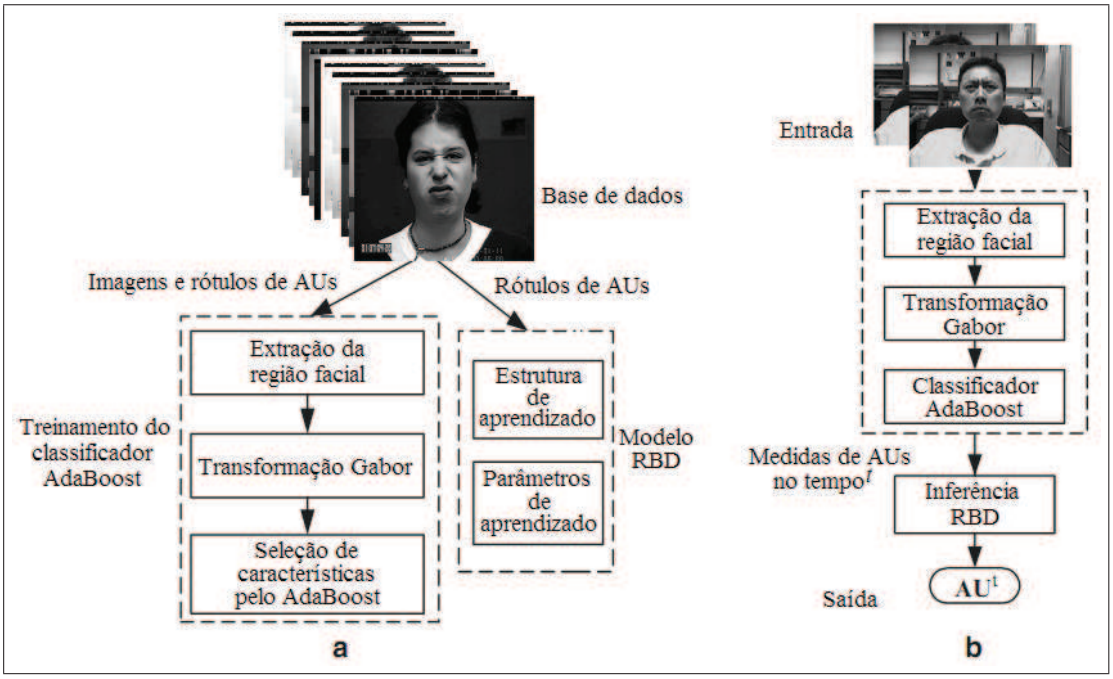


Figura 3.11: (a) Fluxo do processo de treinamento do sistema; (b) fluxo do processo de reconhecimento de AUs.

Durante o treinamento, um conjunto de imagens, que possuem um ou mais AUs expressos, é utilizado como entrada. Sobre as imagens desse conjunto é aplicado um detector de face e olhos baseado no método de Análise Recursiva Discriminante Não-paramétrica (*Recursive Nonparametric Discriminant Analysis* - RNDA), que obtém a face alinhada. A face é dividida em uma metade superior e inferior e sobre essas metades são aplicados filtros de Gabor, que fornecem características Wavelet que são submetidas a classificadores AdaBoost, construídos para realizar a classificação de cada AU. Também durante o treinamento, ocorre a modelagem da Rede Bayesiana Dinâmica (RBD), que se trata de uma rede bayesiana estática (Figura 3.12), modelada ao longo do tempo por um modelo oculto de Markov, que contém AUs e suas relações, além de parâmetros de aprendizado (Figura 3.13).

Já, no fluxo de execução do processo de reconhecimento, são aplicados os mesmos métodos da etapa de treinamento para detecção de face, obtenção das características por filtros de Gabor e classificação dessas características por AdaBoost. O classificador AdaBoost realiza a classificação de cada AU que são submetidos a RBD. A partir dessas entradas, a RBD realiza um ajuste na classificação dos AUs, considerando os relacionamentos entre eles, gerando um resultado mais apurado.

São obtidas pelo sistema proposto uma taxa média de reconhecimento de 91,2% uti-

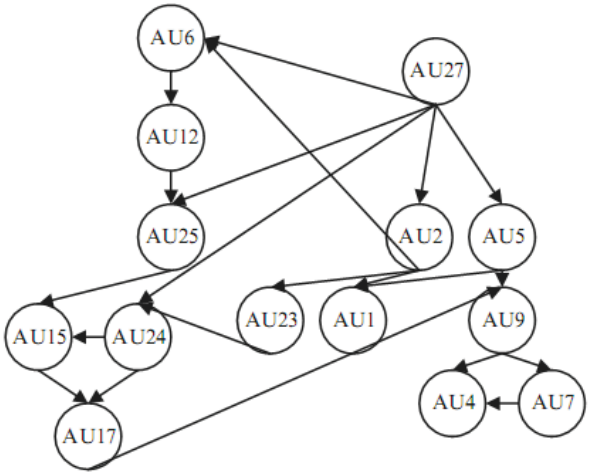


Figura 3.12: Rede Bayesiana treinada com base nas relações encontradas entre AUs.

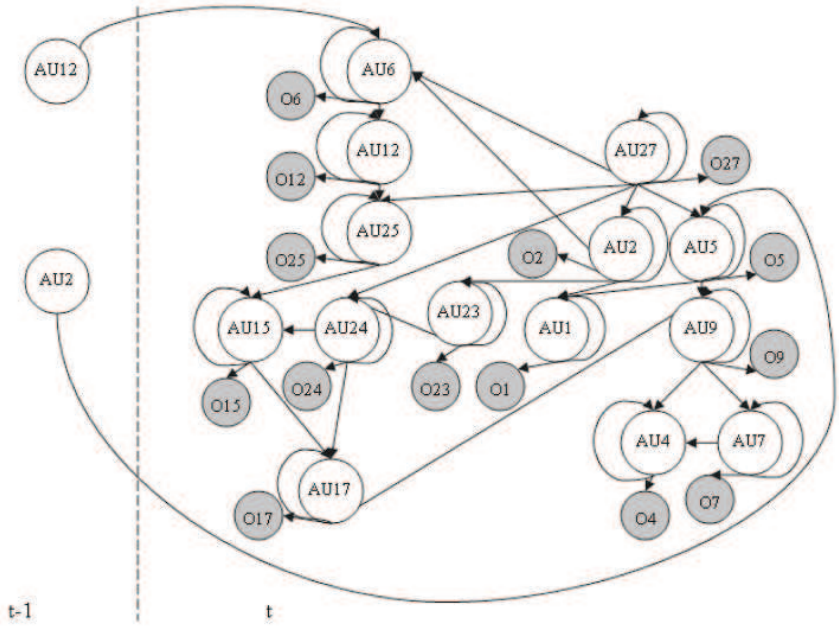


Figura 3.13: Exemplo de relação entre AUs sobre o tempo. Os círculos escuros representam as medidas utilizadas na inferência de relações.

lizando somente Adaboost e 93,33% aplicando RBD. Embora a taxa de reconhecimento não pareça ter obtido grande melhoria com RBD, os resultados mostram que alguns AUs difíceis de serem detectados pelo Adaboost obtiveram melhores detecções pelo RBD, além de menos falsos positivos. Uma possível desvantagem do sistema, é que todo o processo de reconhecimento de AUs ocorre sobre cada frame, o que causa uma sobrecarga de execução em comparação a sistemas que executam algum tipo de rastreamento sobre regiões já detectadas (como detecção da face e dos olhos).

3.9 Reconhecimento de expressões faciais utilizando Raciocínio Baseado em Caso e Lógica Fuzzy

Em (KHANUM et al., 2009) é proposto um sistema para reconhecimento de expressões faciais de emoções básicas (alegria, raiva, surpresa, repulsa, medo e tristeza) através de imagens estáticas. Para atingir este fim, utiliza-se a combinação de Lógica Fuzzy (LF) e Raciocínio Baseado em Casos (RBC) na construção de um sistema.

Após a detecção da face, que é realizada utilizando algoritmos que consideram a cor da pele, dados para o reconhecimento de emoções são extraídos de 8 características faciais (olhos, sobrancelhas, testa, nariz, lábios, dentes, bochechas e queixo). Os dados fornecidos sobre as características faciais tratam-se de pontos (22 pontos) de interesse, que são posicionados sobre regiões específicas e que são obtidos por projeções horizontais sobre divisões da face (Figura 3.14). Porém, para os olhos, lábios e sobrancelhas, métodos de Processamento de Imagens são utilizados para a identificação dos pontos posicionados em posições extremas.

Figura 3.14: Modelo de pontos utilizado. (De Khanum et al. (2009)).

Foram construídas três formas de classificação das emoções: por RBC, classificador Fuzzy e um classificador híbrido baseado em RBC e Fuzzy (Figura 3.15). Com o classificador RBC, os autores atingiram a taxa média de acurácia de 83,5%, com o classificador Fuzzy, 87,7% e com o classificador híbrido, 90,3%. A integração entre Lógica Fuzzy e Raciocínio Baseado em Casos foi realizada combinando as vantagens de cada abordagem. Também o sistema tem a vantagem de apresentar um aprendizado contínuo com o RBC.

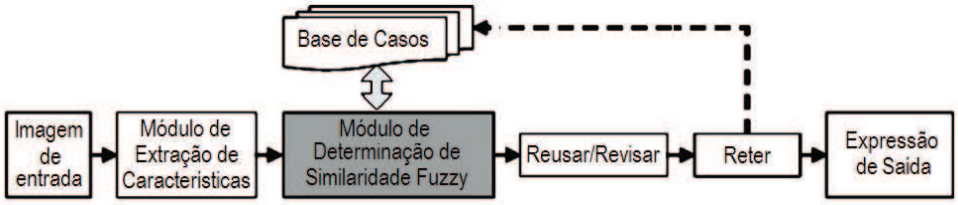


Figura 3.15: Sistema híbrido RBC e Lógica Fuzzy. (De Khanum et al. (2009)).

3.10 Utilizando velocidade e deslocamento no reconhecimento de códigos FACS

É proposto por Brick, Hunter e Cohn (2009) uma nova metodologia para obtenção de códigos FACS com desempenho maior. Além de considerar o deslocamento das características faciais, normalmente utilizado em trabalhos relacionados, são consideradas as medidas de aceleração e de velocidade na execução das ações faciais.

Foram considerados no estudo 16 AUs (1, 2, 4, 5, 6, 7, 9, 12, 15, 17, 20, 23, 24, 25, 26 e 27), que são extraídos de imagens da base facial Cohn-Kanade. Para obtenção das medidas de deslocamentos das características faciais, é utilizado o *Active Appearance Models* (AAMs), que é um modelo matemático que ajusta, como uma espécie de máscara, 68 pontos sobre a face da imagem considerada. O AAM rastreia as movimentações dos pontos considerados relevantes a classificação dos AUs, conforme exibido na Figura 3.16.

Figura 3.16: Rastreamento de pontos com modelo AAM. (De Brick, Hunter e Cohn (2009)).

Para considerar velocidade (mudança na posição de pontos ao longo do tempo) e aceleração (mudança na velocidade ao longo do tempo), é utilizada uma série temporal com os dados de deslocamentos, sobre forma matricial. Então, essa matriz é multiplicada com uma matriz de pesos pré-determinados. O resultado é a estimativa de posicionamento, velocidade e aceleração para uma série temporal considerada.

Durante os experimentos, foram utilizados para classificação SVM e Análise Discriminante Linear (*Linear Discriminant Analysis* - LDA). Foram treinados três casos de classificação: (1) utilizando apenas dados de posicionamento de pontos, (2) utilizando dados de posicionamento e velocidade de deslocamento de pontos e (3) considerando posicionamento, velocidade e aceleração de deslocamento de pontos. Os resultados demostraram que o uso de velocidade e aceleração aumenta o número de verdadeiros positivos e diminui o número de falsos positivos na maioria dos AUs. De forma geral, houve um aumento na média de 4,2% (com SVM) e 4,5% (com LDA) na classificação correta geral, que tem taxas de reconhecimento entre 83,4% e 96,3%.

3.11 Comparativo entre trabalhos

Pode-se ver na Tabela 3.1, um comparativo entre os trabalhos relacionados descritos anteriormente. Entre os itens comparados estão: o método utilizado para obtenção da face; método para extração das características faciais; método para classificação das expressões faciais; seu modo de análise, se é sobre cada imagem/*frame* (estático) ou se considera a sequência de imagens na sua avaliação (dinâmico); as emoções inferidas; número de AUs do sistema FACS extraídos; se houve alguma ação manual sobre o sistema ou se ele é totalmente automático; e seu desempenho.

O método para obtenção de face mais utilizado entre os trabalhos selecionados, foi o de *Haar-like*, seguido pelo método que utiliza a cor da pele para esta detecção. A vantagem do primeiro método é sua rapidez e robustez, porém tem pouca tolerância à rotação e à inclinação, ao contrário do método que utiliza a cor da pele. Já, para a extração das características faciais existe uma variedade de métodos, mas verifica-se que, na maioria dos trabalhos, o objetivo, independente do método, é a localização das extremidades destas características. As expressões faciais são classificadas utilizando, principalmente, RNA e SVM, sendo ambos métodos robustos, embora distintos. A maioria dos trabalhos relacionados utiliza o modo de análise estático, que é o mais encontrado nas referências pesquisadas. Embora a análise dinâmica forneça mais dados para a determinação de expressões faciais, isso não representa necessariamente que essa análise sobreponha a análise estática (PETRIDIS et al., 2009). Metade dos trabalhos focou na extração de emoções, a outra metade na obtenção de códigos do sistema FACS, mas nenhum realizou ambas. Nota-se que somente nos trabalhos mais antigos é declarada a ação manual sobre os sistemas construídos, o que demonstra uma evolução nas técnicas que tornaram mais automáticos os processos de extração de expressões e emoções pela face.

Tabela 3.1: Comparação entre trabalhos relacionados.

Autor	Obtenção da face	Extração das características	Classificação das expressões faciais	Modo de análise	Emoções	AUs	Ação manual	TREC (%)
Kobayashi e Hara (1991)	N/D	extremidades de características faciais (marcação manual)	RNA	estático	6	0	sim	87 - 91
Tian, Kanade e Cohn (2001)	RNA	cor, forma, movimento	RNA	dinâmico	0	16	sim	97
Pantic e Rothkrantz (2004)	cor da pele	*templates*, detecção de bordas e contornos, RNA, regras	regras	estático	0	32	sim	86
Martin et al. (2005)	*Haar-like features*	Projeção integral vertical e horizontal; (extremidades de características faciais)	SVM	dinâmico	6	0	não	N/D
Batista, Gomes e Carvalho (2006)	N/D	1) GF; 2) PCA	1) SVM e k-NN; 2) SVM e RNA	estático	1	0	N/D	1) 81,3; 2) 87,5
Valstar e Pantic (2006)	*Haar-like features*	GF; GentleBoost	SVM	dinâmico	0	15	não	90,2
Sohail e Bhattacharya (2007)	*Haar-like features*	extremidades de características faciais via técnicas combinadas de Processamento de Imagens	SVM	estático	6	0	não	84,8 - 89,4
Tong, Liao e Ji (2008)	RNDA	GF	1) AdaBoost; 2) AdaBoost-RBD	estático	0	14	não	1) 91,2; 2) 93,3
Khanum et al.(2009)	cor da pele	Projeção integral horizontal	1) RBC; 2) Fuzzy; 3) RBC-Fuzzy	estático	6	0	N/D	1) 83,5; 2) 87,7; 3) 90,3
Brick, Hunter e Cohn (2009)	AAM	AAM	1) SVM; 2) LDA	dinâmico	0	16	não	1) 83,4 - 96,3

Não há como afirmar quais trabalhos podem ser considerados os melhores, pois, embora tenham objetivos similares, possuem escopos e metodologias distintas. Dos trabalhos relacionados, a maior taxa de reconhecimento é de 97% na identificação de 16 AUs (TIAN; KANADE; COHN, 2001), porém existe outro trabalho que identifica o dobro de AUs (PANTIC; ROTHKRANTZ, 2004), mas obtém uma taxa de reconhecimento de 86%. Considerando este exemplo, analisar somente o desempenho de detecção não é suficiente para definir o melhor trabalho, pois um abrange uma faixa maior de AUs, porém com resultado inferior em relação ao que identifica menos AUs. Ainda utilizando como exemplo o trabalho que identifica 16 AUs a 97%, existe outro que identifica um número menor de AUs, 14, mas com uma taxa inferior, de 93,3%, porém faz isso automaticamente (TONG; LIAO; JI, 2008). Um outro detalhe interessante é que os trabalhos que identificam emoções aparecem com taxas de sucesso inferior aos que detectam AUs, que pode fornecer indícios de que a tarefa de detecção de emoções é mais difícil do que a detecção de AUs.

Considerando que, pelos números apresentados, identificar AUs é, teoricamente, mais fácil que identificar emoções, e conhecendo os AUs é possível inferir as emoções, a intenção do trabalho proposto é obter os AUs e, por estes códigos, a emoção. Também foi definido que o método para obtenção da face será o de *Haar-like features*, sem ação manual.

4 TRABALHO PROPOSTO

Pretende-se, neste trabalho, construir uma aplicação capaz de detectar automaticamente as seis emoções básicas (alegria, repulsa, tristeza, raiva, surpresa e medo) de uma face humana. Estas emoções são inferidas de códigos FACS a partir da pessoa à frente do computador.

Utiliza-se uma *webcam* para a captura da face, que tem parâmetros dos olhos, sobrancelhas e boca identificados automaticamente. Optou-se por tais características, pois estas são as mais relevantes na classificação de expressões emocionais em uma face. Os dados destas características faciais são submetidos a métodos que, a partir da expressão facial, executam a inferência da manifestação afetiva (emoção) presente.

A Figura 4.1 exibe o fluxo de macro processos realizados pela aplicação proposta: detecção da face, detecção de características faciais, classificação da expressão facial e inferência da emoção. Esse processo se inicia pela obtenção de imagens capturadas por uma *webcam* de usuários à frente de um computador. Estas imagens são submetidas à métodos de Visão Computacional para a localização da face (detecção da face). Sobre a face encontrada, são extraídos dados através de pontos distribuídos sobre as características faciais (detecção de características faciais). Após, o comportamento das coordenadas destes pontos são analisados em busca da presença de ações faciais. Estas ações faciais trazem dados que podem evidenciar a ocorrência de alguma emoção (inferência da emoção).

4.1 Metodologia de trabalho

Para atingir os objetivos definidos para o trabalho, a sua proposta de desenvolvimento foi dividida em seis etapas. Primeiramente, foram realizados estudos bibliográficos. As áreas de pesquisa relevantes para este trabalho abrangem Processamento de Imagens, Visão Computacional, Computação Afetiva, Interação Humano-Computador e Psicologia, sendo as duas primeiras adotadas como foco inicial para realizar FaD e FeD.

Na segunda etapa, estudos bibliográficos concentraram-se em Computação Afetiva e

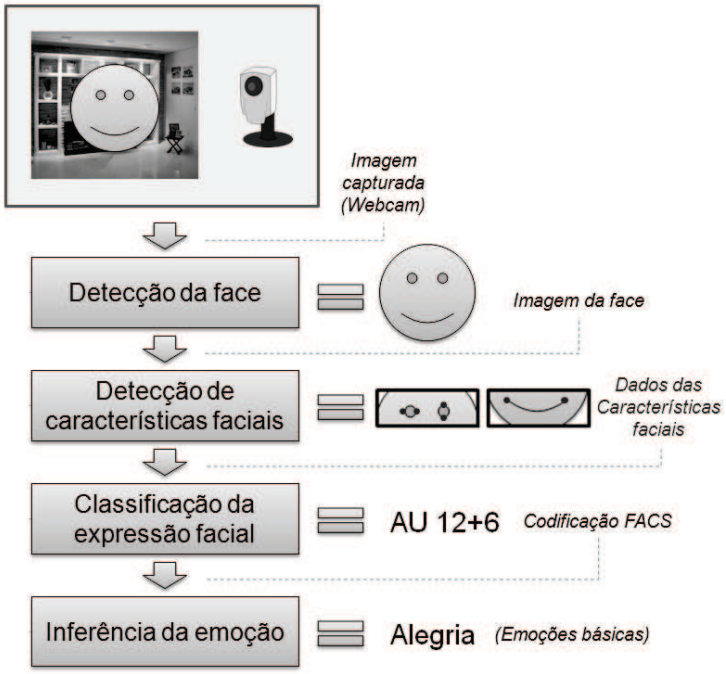


Figura 4.1: Fluxo de processos da aplicação.

Psicologia, resultando na criação da Seção 2.3. Nesta seção foram apresentadas, introdutoriamente, algumas metodologias utilizadas para a classificação de emoções baseada no reconhecimento de expressões faciais. Entre estes estudos, chamam a atenção os que empregam o sistema de codificação FACS na classificação da emoção, devido sua demonstrada capacidade de solução em diversos trabalhos. Por este motivo este sistema será utilizado.

Na etapa seguinte foram estudadas bibliotecas de Visão Computacional para construção da aplicação. Para FaD e FeD, foram consideradas mais adequadas as funções encontradas na biblioteca OpenCV (BRADSKI; KAEHLER, 2008; WILLOW GARAGE, 2010). Na mesma etapa foram decididos os métodos para classificação de expressões faciais e inferência de emoções. Sendo assim, foi adotada rede neural artificial (RNA), fornecida pela biblioteca FANN (NISSEN, 2003), para a classificação das expressões faciais. Para emoções, será utilizada uma árvore de decisão utilizando o software See5 (RULEQUESTRESEARCH, 2001).

A quarta etapa foi reservada para o planejamento e desenvolvimento de uma aplicação experimental que faça automaticamente a inferência da emoção de um usuário à frente de um computador. A aplicação é composta por quatro módulos, correspondentes aos macro processos (Fig. 4.1), executados sequencialmente: (1) detecção da face por *webcam*; (2) detecção de características faciais; (3) classificação da expressão facial; (4) inferência da emoção.

Na quinta etapa houve os testes da aplicação. A base de faces CK+ (LUCEY et al., 2010) foi utilizada como apoio para o treinamento da rede neural classificadora de códigos

FACS, e para os testes da aplicação. Nestes primeiros testes foram levantados pontos que necessitavam de ajustes, que ocorreram na etapa seguinte.

Na sexta e última etapa foram realizados ajustes na implementação, avaliação da aplicação, análise de dados sobre as amostras coletadas em testes e sugestões sobre futuras aplicações para o trabalho desenvolvido. Houve também nesta última etapa, um esforço final para a escrita da dissertação que ocorreu paralelamente em todas etapas, que foi, porém, intensificada nesta última.

4.2 Estrutura da aplicação de inferência de emoções

Conforme descrito anteriormente, a aplicação construída para executar a inferência de emoções é dividida pela seguinte sequência de etapas (módulos): detecção da face, detecção das características faciais, classificação da expressão facial e inferência da emoção. As próximas seções descrevem como ocorre a execução de cada um destes módulos.

4.2.1 Etapa 1: detecção da face

Inicialmente é realizada a busca por uma face (FaD) em uma imagem (Figura 4.2(a)). Isto ocorre sobre cada *frame* até que a face seja detectada (Figura 4.2(b)). A região da face detectada é demarcada e utilizada na etapa seguinte do processo, de detecção das características faciais (Figura 4.2(c)).

A caixa, a seguir, contém o resumo das especificações do primeiro módulo:

```
módulo:   detecção da face.
entrada:  um frame de uma sequência de imagens (webcam, vídeo).
processo: 1) realizar busca por uma face na imagem de entrada (frame);
          2) se houver detecção, seguir ao próximo passo, senão, executar novamente o
             passo anterior desse módulo sobre o frame seguinte;
          3) a região da face detectada é isolada em uma nova imagem.
saída:    imagem contendo a face isolada.
```

4.2.2 Etapa 2: detecção das características faciais

Na segunda etapa, sobre a região da face detectada, cinco subprocessos ocorrem sequencialmente: busca pelo centro dos olhos, correção da inclinação da face, aplicação de modelo

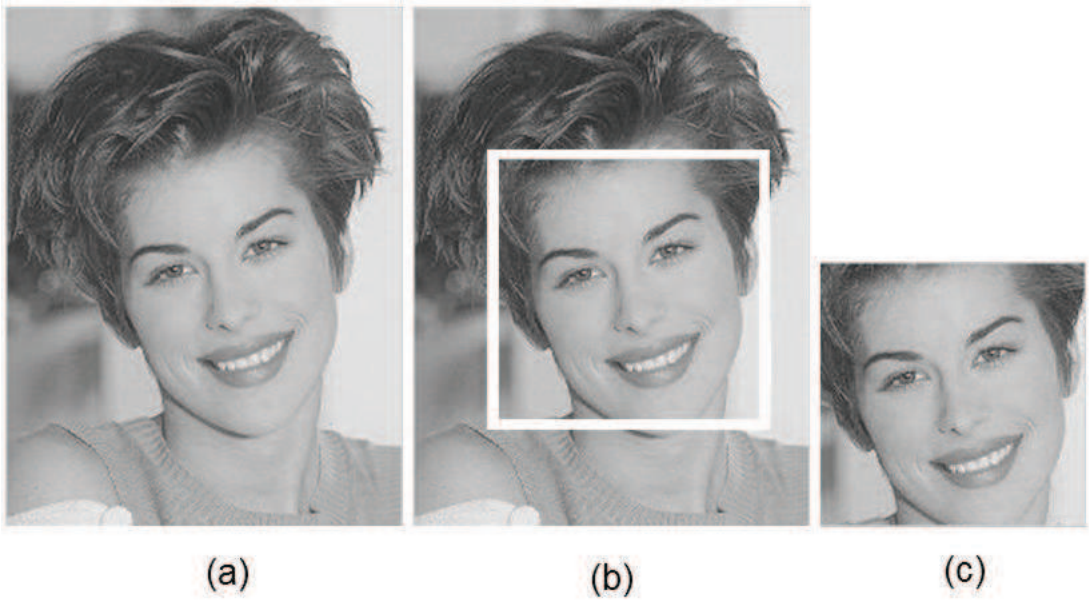


Figura 4.2: Detecção da face. (a) Figura contendo face; (b) figura com a região da face detectada. (c) região da face detectada isolada. (De Rowley, Baluja e Kanade (1997)).

antropométrico, identificação de pontos extremos sobre as características faciais e avaliação desses pontos extremos. Cada um desses subprocessos será visto nas próximas seções.

Busca pelo centro dos olhos

Primeiramente ocorre a divisão da face em regiões de interesse (Figura 4.3(b)). Nessas regiões de interesse são utilizados detectores específicos, para cada um dos olhos, que indicarão as suas coordenadas (Figura 4.3(c)).

Correção da inclinação da face

Os centros dos olhos são utilizados como parâmetro para correção da inclinação da face, que é realizada com a rotação de sua imagem caso ela apresente inclinação relevante. Este processo é ilustrado pela Figura 4.3.

A detecção de olhos é realizada sobre cada imagem submetida a esta etapa e repetida após a correção da inclinação da face. Essa repetição é necessária, pois as coordenadas originais dos centros dos olhos são perdidas após a correção da inclinação da face, mas estes dados são utilizados pelo modelo antropométrico, subprocesso seguinte. Em caso de falha na detecção de um dos olhos, descarta-se o *frame* corrente e reinicia-se o processo aguardando por outra imagem contendo a face isolada.

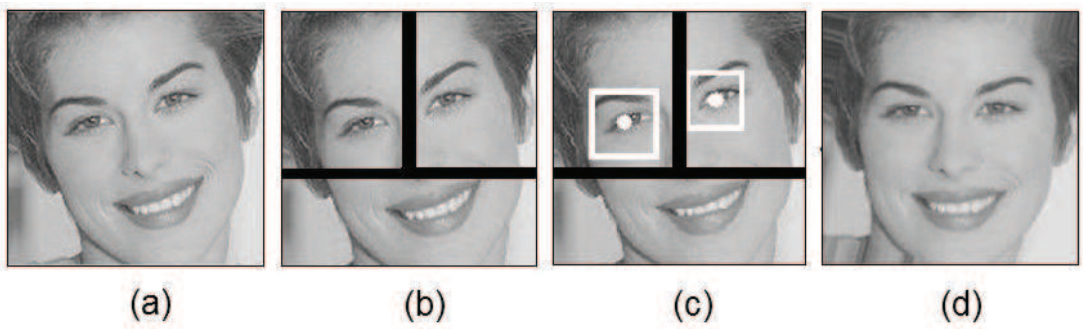


Figura 4.3: Detecção de olhos e correção da inclinação da face. (a) Face detectada - com inclinação; (b) divisão da face em regiões de interesse; (c) olhos detectados dentro das regiões de interesse; (d) face com inclinação corrigida.

Aplicação de modelo antropométrico

Posteriormente, aplica-se um modelo antropométrico que, tendo as coordenadas e a distância entre os olhos, delimita as áreas correspondentes as características faciais olhos, boca e sobrancelhas (Figura 4.4(b)). Com essa delimitação a tarefa de FeD é otimizada, pois somente as regiões que interessam são consideradas.

Identificação de pontos extremos sobre as características faciais

Na FeD, os pontos sobre as extremidades das características faciais são demarcados automaticamente e armazenados em um vetor, chamado **vetor de pontos extremos**. Para olhos e boca, são obtidos primeiramente os limites horizontais (eixo x) e posteriormente os verticais (aproximadamente na metade da distância entre os extremos horizontais). Para as sobrancelhas, são obtidos apenas os pontos horizontais, sendo eles os extremos em sua parte interna, e na externa são os coincidentes sobre o eixo horizontal dos cantos externos dos olhos. A Figura 4.4 exemplifica o processo de obtenção das extremidades das características faciais.

Avaliação dos pontos extremos encontrados

Os pontos extremos, anteriormente identificados, passam por uma validação que tem como objetivo evitar que falhas ocorridas na identificação desses pontos perturbem a execução dos processos seguintes da aplicação. Nesta validação são verificadas as coordenadas dos pontos para que se garanta que os pontos estejam sobre aéreas esperadas. Por exemplo, espera-se que os pontos extremos internos horizontais dos olhos estejam em uma posição (altura) próxima sobre o eixo vertical (y), e abaixo dos pontos das sobrancelhas, considerando uma leitura a partir do canto superior esquerdo da imagem.

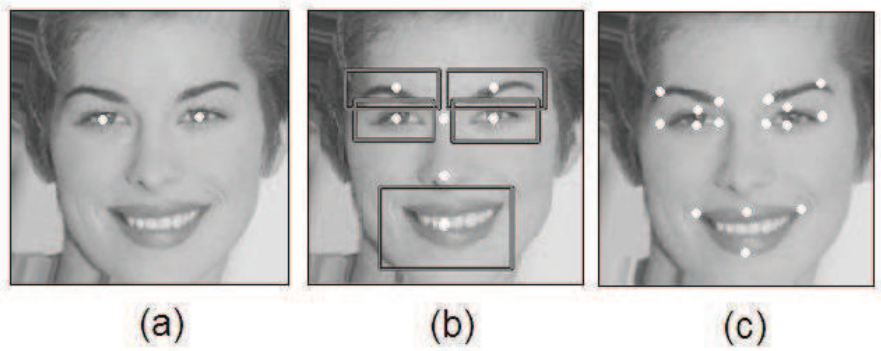


Figura 4.4: Deteção de pontos de interesse sobre características faciais. (a) Imagem com os centros dos olhos detectados; (b) modelo antropométrico aplicado sobre a face; (c) detecção de extremidades nas características faciais de interesse.

Abaixo, especificações resumem o segundo módulo:

```
módulo:   detecção de características faciais.
entrada:  imagem contendo face, gerada pelo módulo anterior.
processo: 1) dividir a imagem de entrada em regiões de interesse e sobre elas fazer
             busca pelo centro dos olhos;
          2) se houver detecção de ambos os olhos, seguir ao próximo passo, senão,
             encerrar a execução deste módulo e executar novamente o módulo anterior;
          3) se há inclinação relevante na face, corrigi-la e retornar a executar o
             passo 1, senão, seguir ao próximo passo;
          4) aplicar o modelo antropométrico sobre a imagem da face;
          5) encontrar os pontos extremos sobre as características faciais;
          6) avaliar posicionamento dos pontos extremos encontrados;
          7) caso haja alguma inconformidade nos pontos extremos, encerrar a execução
             deste módulo e executar novamente o módulo anterior (detecção de faces);
saída:    vetor contendo coordenadas dos pontos extremos das características faciais.
```

4.2.3 Etapa 3: classificação da expressão facial

Se as coordenadas dos pontos das características faciais contidos no vetor de pontos extremos forem corretamente identificadas, executa-se o módulo de classificação de expressões faciais. Inicialmente assume-se o primeiro vetor de pontos extremos como sendo de face neutra e após, obtêm-se os estados das características faciais (boca aberta, olhos fechados etc) pelos dados deste vetor. O resultado será um outro: o **vetor de estados das características faciais** para face neutra. Todos os demais vetores de estados subsequentes são considerados como contendo expressão.

Posteriormente o vetor de pontos extremos é submetido a um algoritmo de rastreamento,

que tenta ajustar estes pontos à imagem do próximo *frame*. Após o rastreamento, um novo vetor de pontos extremos é obtido e seus pontos são avaliados (da mesma forma como ocorreu no módulo anterior) para verificação de anomalias no posicionamento sobre as características faciais. Caso haja alguma anomalia, o processo é interrompido e reiniciado no primeiro módulo. Não havendo inconsistências no posicionamento dos pontos, como ocorreu com a face neutra, a partir do vetor de pontos extremos da face contendo expressão é obtido o vetor de estados das características faciais para face com expressão.

No etapa seguinte é realizada a diferença entre os vetores de estados das características faciais de face neutra e com expressão. Essa diferença resulta no **vetor de características**, que é normalizado e submetido à uma rede neural. Essa rede neural classifica o estado das características faciais de acordo com o sistema FACS, ou seja, sua saída são expressões faciais na forma de um **vetor de AUs**.

A seguinte caixa possui as especificações sobre a terceira etapa:

```
módulo:   classificação da expressão facial.
entrada:  vetor contendo coordenadas dos pontos extremos das características faciais.
processo: 1) caso seja a primeira iteração, considerar o primeiro vetor de pontos
             extremos como sendo de uma face neutra;
          2) caso seja a primeira iteração, obter o vetor de estados da face neutra do
             vetor de pontos extremos da face neutra;
          3) pelo vetor de pontos extremos, rastrear pontos sobre o frame seguinte;
          4) avaliar os pontos rastreados e havendo inconformidades, reiniciar
             execução do primeiro módulo, senão, seguir o próximo passo;
          6) obter vetor de estados de face com expressão;
          7) criar vetor de características da diferença entre vetor de estados da
             face neutra e com expressão;
          8) submeter vetor de características a rede neural.
saída:    vetor de AUs fornecida pela rede neural.
```

4.2.4 Etapa 4: inferência da emoção

A etapa de classificação da expressão facial fornece um vetor de AUs extraídos das expressões faciais obtidas da face capturada. Como visto na seção 2.3.2, pela combinações de AUs é possível determinar a emoção manifestada por uma expressão facial. É desse método que o módulo corrente se utiliza para obter as emoções de uma face.

A inferência de emoções é realizada por uma árvore de decisão, que retorna a emoção correspondente a um conjunto de AUs. Pode-se ver este processo detalhado na caixa abaixo:

```
módulo:   inferência da emoção.
entrada:  vetor de AUs.
processo: 1) submeter à árvore de decisão ao vetor de AUs;
          2) caso exista mais algum frame para ler, executar novamente o módulo de
             classificação da expressão facial, senão, encerrar execução da aplicação.
saída:    emoção.
```

O fluxograma presente na Figura 4.5, mostra todos os módulos que compõem a aplicação a partir da visão dos subprocessos.

4.3 Métodos aplicados

A Figura 4.1 exibe as etapas da aplicação construída para identificar emoções, que na seção anterior foi detalhada passo a passo. Entretanto, e execução dessas etapas depende da aplicação de diversas técnicas e métodos, entre eles, os principais estão ilustrados na Figura 4.6. As próximas seções aprofundam as descrições sobre os métodos empregados em cada uma das quatro etapas da aplicação.

4.3.1 Detecção da face

Para a detecção de face (Figura 4.2), é aplicado um classificador de faces baseado em *Haar-like features* (Método de Viola-Jones (VIOLA; JONES, 2001)). Pela categorização de Yang, Kriegman e Ahuja (2002) este é um método baseado em aparência, pois utiliza um conjunto de imagens como base em sua construção (treino) para obter um padrão que é procurado em sua execução.

Mais detalhes sobre este classificador podem ser encontrados na seção 2.2.3, na subseção Viola-Jones. Este e outros classificadores baseados no mesmo método e que foram empregados neste trabalho são fornecidos pela biblioteca OpenCV.

4.3.2 Detecção de características faciais

Métodos de Processamento de Imagem e de Visão Computacional contidos na biblioteca OpenCV (BRADSKI; KAEHLER, 2008) são utilizados para a detecção de características faciais. Subsequente a obtenção da imagem da face, é realizada sobre ela uma normalização, com o ajuste de inclinação e de escala; segmentação, isolando as regiões de interesse (olhos,

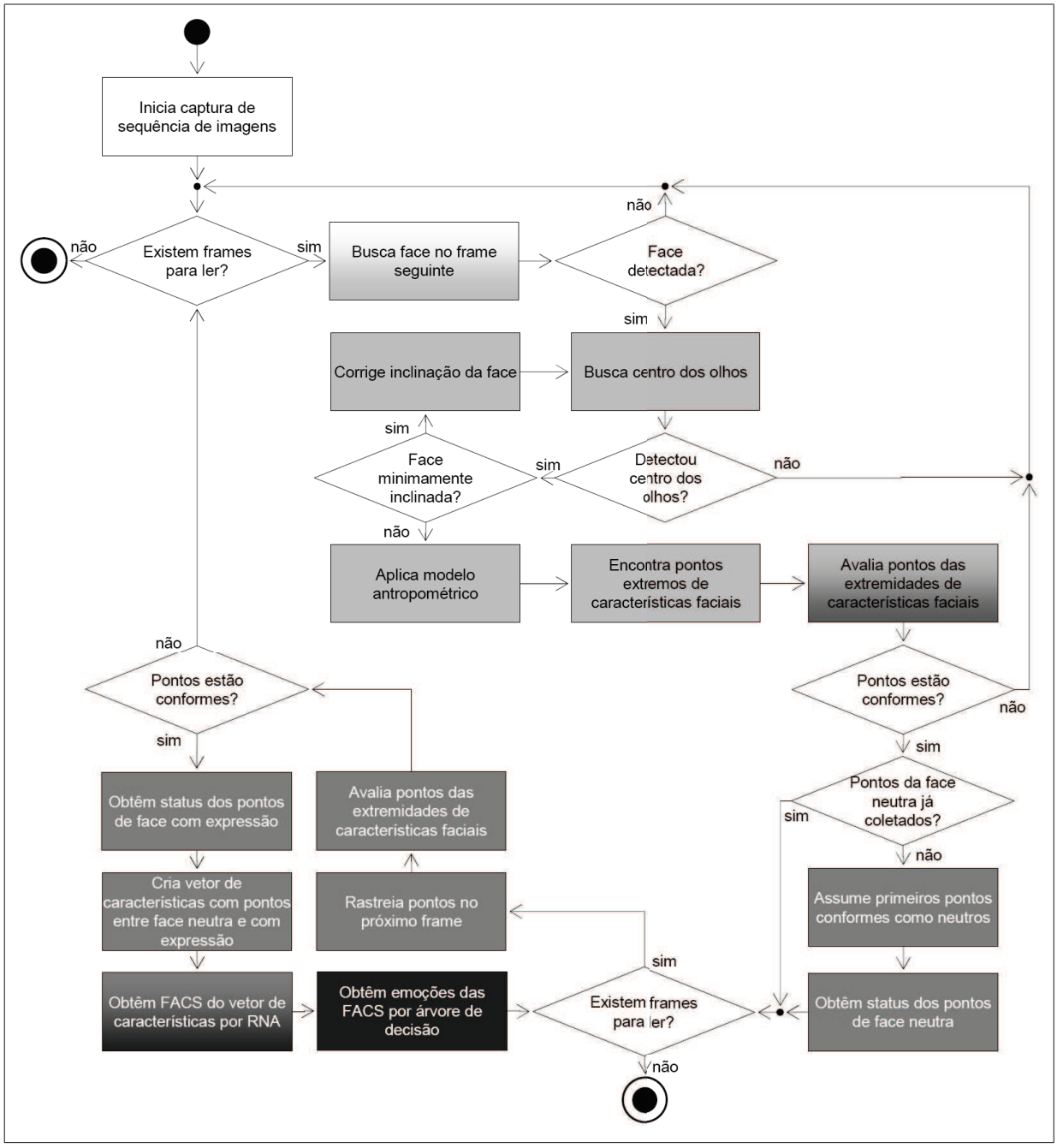


Figura 4.5: Fluxograma da aplicação.

boca, sobrancelhas); e melhorias na qualidade da imagem para favorecer a obtenção de dados das características faciais e seus estados.

O primeiro passo desta etapa é a correção da inclinação da face, e para isto é utilizado o centro dos olhos como referência. Na detecção do centro dos olhos é aplicado como classificador o método de *Haar-like features*, o mesmo da detecção da face. Para simplificar este processo, o classificador é aplicado somente sobre regiões prováveis de ocorrência, que se tratam das partes superiores da face, como mostrado na Figura 4.3(b). Tendo em mãos as

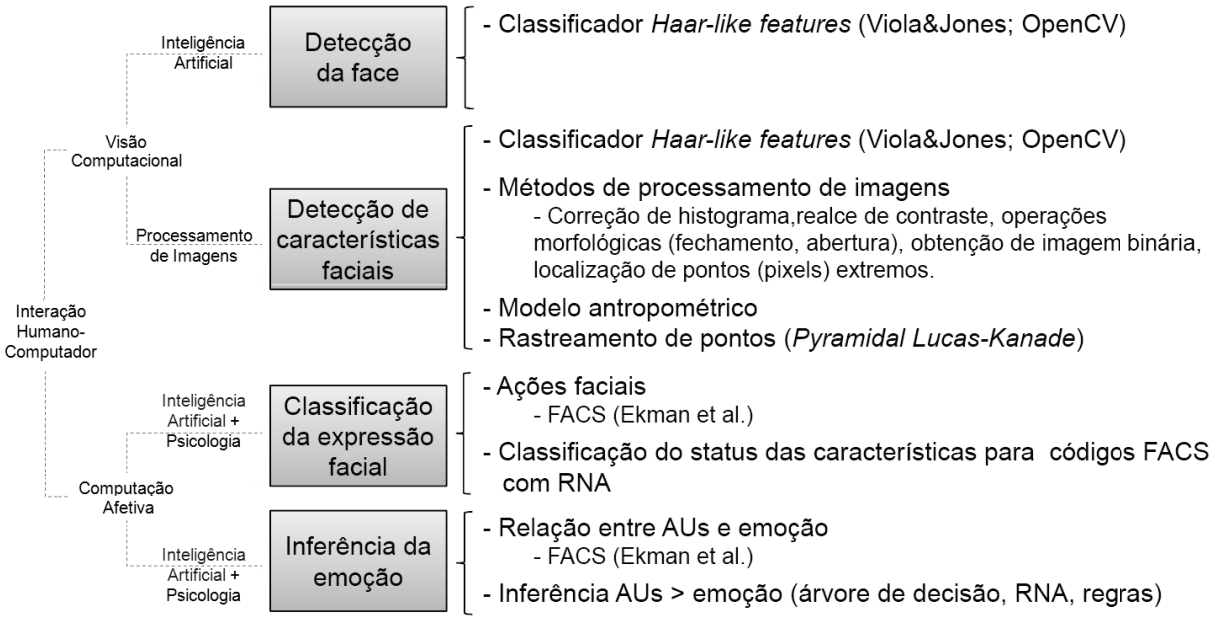


Figura 4.6: Métodos aplicados e suas áreas de pesquisa de origem.

coordenadas dos centros dos olhos obtidas pelo classificador (Figura 4.3(c)), obtém-se o ângulo de inclinação da face pela Equação (4.1) (onde (x_1, y_1) e (x_2, y_2) são coordenadas do centro olho direito e esquerdo respectivamente). Em seguida, aplicando uma transformação afim de rotação de matrizes, executa-se a correção dessa inclinação rotacionando a imagem ao ângulo zero entre os centros dos olhos (Figura 4.3(d)). Neste processo, o centro da imagem é considerado como ponto de origem e o ângulo zero é encontrado quando os dois pontos dos centros dos olhos estão no mesmo no eixo horizontal $(y_1 = y_2)$. Esta correção de inclinação é útil para evitar perturbações que podem ocorrer na identificação dos estados das características faciais, como exemplo, saber se um canto da boca está mais inclinado do que o outro. Ela também é necessária para que o modelo antropométrico seja posicionado sobre as regiões corretas.

$$\theta = \tan^{-1}\left(\frac{|y_1 - y_2|}{x_2 - x_1}\right) \tag{4.1}$$

Após a inclinação corrigida, a imagem da face é redimensionada para o tamanho de 100×100 pixels, que segundo (TIAN; KANADE; COHN, 2005) é uma resolução suficiente para detecção de características faciais. Essa medida é tomada para otimizar o processamento, porém caso a imagem original tenha resolução inferior a essa, o redimensionamento tem pouco efeito. Em seguida, é aplicado o modelo antropométrico de Sohail e Bhattacharya (2006), que partindo dos centros dos olhos (Figura 4.4(a)), permite demarcar as regiões dos olhos, boca e sobrancelhas (Figura 4.4(b)). Sobre a região das características faciais, são empregadas diversas

técnicas de Processamento de Imagem que foram consideradas úteis no auxílio à obtenção de pontos sobre as extremidades das características. Sequencialmente são utilizados os métodos: (*a*) conversão da imagem para tons de cinza, (*b*) correção de histograma, (*c*) realce de contraste, (*d*) filtro bilateral (BRADSKI; KAEHLER, 2008), (*e*) operação morfológica de abertura (apenas sobre os olhos), (*f*) obtenção de imagem binária (pelo método de limiarização adaptativa) e (*g*) eliminação de pequenas ilhas e de (*h*) vales na imagem com o algoritmo de preenchimento *Flood Fill* (BRADSKI; KAEHLER, 2008). Posteriormente é (*i*) demarcada uma área retangular delimitada pelos contornos presentes em uma imagem, em que é realizada a busca pelos (*j*) pontos extremos das características faciais. Uma exceção ocorre com os olhos, que têm os pontos extremos localizados, mas isso não acontece dentro uma região específica. O motivo para isso, é que os olhos nem sempre apresentam, após a aplicação dos métodos de Processamento de Imagens, uma área contínua (principalmente devido a esclera[1]). Dessa forma, uma demarcação poderia eliminar parte dessa característica.

Para as sobrancelhas, são obtidos apenas os pontos horizontais, sendo eles os extremos em sua parte interna (pontos 2 e 3), e na externa (pontos 1 e 4) são os que coincidem, aproximadamente, sobre o eixo horizontal dos cantos externos dos olhos (pontos 5 e 12). Essa medida foi tomada para definir explicitamente um limite externo da sobrancelha, que tem comportamento indefinido (se encontra com os cabelos, é mais curta, se curva em direção aos cantos horizontais externos dos olhos etc). Em relação aos olhos, os pontos extremos verticais (pontos 6, 7 e 10, 11) são posicionados no eixo horizontal, na metade da distância entre os pontos extremos (pontos 5, 8 e 9, 12), mais 1/6 dessa distância para o olho direito e menos 1/6 da respectiva distância para o olho esquerdo. A Tabela 4.1 descreve a posição dos pontos sobre as extremidades das características faciais, que informam o estado dessas características. A Figura 4.7 contém a disposição dos 16 pontos, que depois de posicionados ficarão como mostrado na Figura 4.4(c).

Na Figura 4.8 é possível visualizar a aplicação dos métodos Processamento de Imagens e seus resultados. Esta operação realiza uma melhoria na qualidade das imagens das regiões definidas pelo modelo antropométrico e se estende até a obtenção dos pontos extremos, utilizados para a classificação das expressões faciais.

4.3.3 Classificação da expressão facial

Com as coordenadas de alguns pontos nas extremidades das características faciais, pode-se calcular o deslocamento que ocorre sobre estes pontos em um intervalo de tempo determinado, sendo assim possível encontrar os respectivos AUs definidos em FACS (EKMAN;

[1]Esclera: área do olho de coloração branca adjacente a íris.

Tabela 4.1: Descrição dos pontos extremos de características faciais.

Ponto	Característica	Lado	Posição	Observação
1	sobrancelha	direita	externo	1x = aprox. ponto 5x
2			interno	
3		esquerda	interno	
4			externo	4x = aprox. ponto 12x
5	olho	direita	externo	
6			superior	6x = (dod/2)+(dod/6)
7			inferior	7x = (dod/2)+(dod/6)
8			interno	
9		esquerda	interno	
10			superior	10x = (doe/2)-(doe/6)
11			inferior	11x = (doe/2)-(doe/6)
12			externo	
13	boca	direita	externo	
14		-	superior	
15			inferior	
16		esquerda	externo	

Obs.: doe, dod = distancia entre pontos extremos dos olhos esquerdo e direito

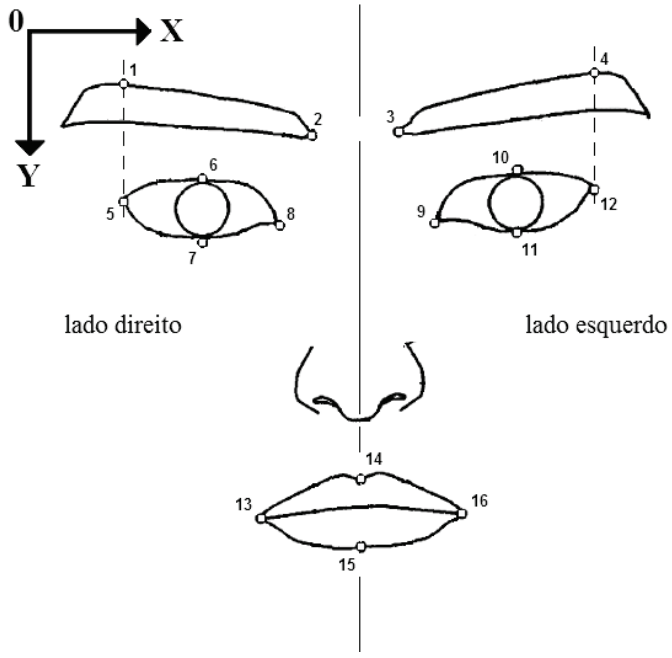


Figura 4.7: Pontos extremos sobre características faciais.

FRIESEN; HAGER, 2002a). Um exemplo que pode ser utilizado é o baixar e o aproximar das sobrancelhas, movimentação que pelo FACS representa o AU 4, que pode ser identificado pelo comportamento de pontos nesta característica. O vetor de pontos extremos, obtidos na FeD, contém essas coordenas necessárias para calcular o estado das características faciais.

Os pontos extremos depois de obtidos passam a ser rastreados, o que reduz o custo computacional da aplicação, que deixa de executar os dois primeiros módulos (FaD e FeD), se

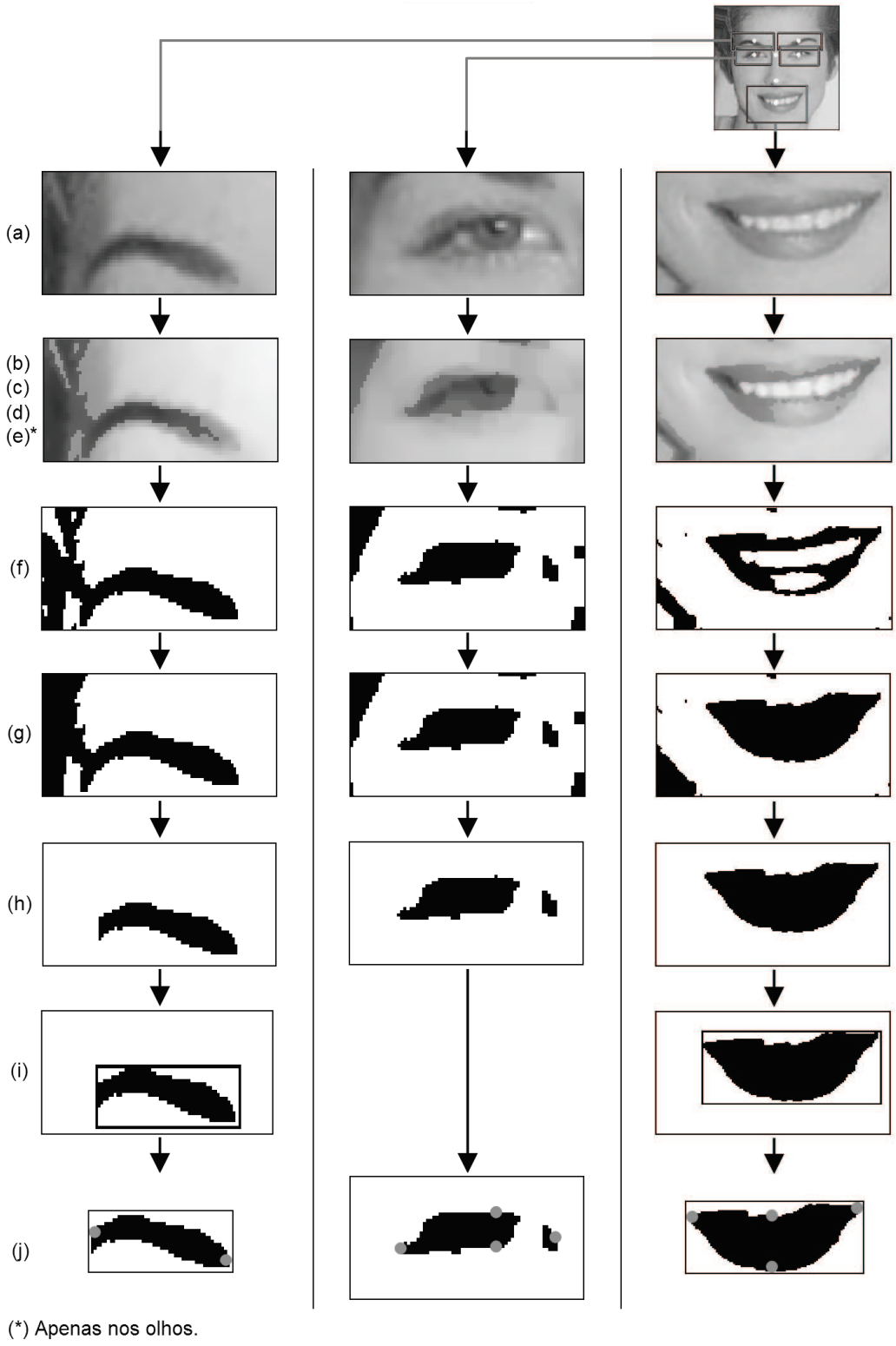


Figura 4.8: Etapas/métodos para obtenção de pontos extremos sobre características faciais. (a) Conversão da imagem para tons de cinza; (b) correção de histograma; (c) realce de contraste; (d) filtro bilateral; (e) operação morfológica de abertura (apenas sobre os olhos); (f) obtenção de imagem binária (limiarização adaptativa); (g) eliminação de pequenas ilhas; (h) eliminação de vales; (i) demarcação de contornos; (j) posicionamento de pontos extremos.

concentrando nos dois seguintes. O rastreamento é realizado utilizando o método *Pyramidal Lucas-Kanade*, implementado na biblioteca OpenCV (WILLOW GARAGE, 2010). Os pontos rastreados são avaliados após cada *frame* e devem obedecer um conjunto de premissas contidas na Tabela 4.2.

Tabela 4.2: Regras de posicionamento dos pontos sobre características faciais.

Regra	Descrição	Exemplo
1	*x* para os pontos do lado direito da face devem ser menores que os *x* para os pontos lado esquerdo da face	(1x,2x,5x,6x,7x,8x,13x)< (3x,4x,9x,10x,11x,12x,16x)
2	*y* dos pontos superiores dos olhos e boca devem ser menores que *y* de seus respectivos pontos inferiores	(6y<7y),(10y<11y),(14y<15y)
3	valores x dos pontos superiores e inferiores dos olhos e boca devem estar entre seus respectivos pontos extremos verticais	(5x<(6x,7x)<8x), (9x<(10x,11x)<12x), (13x<(14x,15x)<16x)
4	*y* dos pontos das sobrancelhas devem ser menores que *y* dos pontos extremos horizontais dos olhos de seu lado	((1y,2y)<(5y,6y,7y,8y)),((3y,4y)< (9y,10y,11y,12y))
5	*y* dos pontos dos olhos devem ser menores que os pontos y da boca	((5y,6y,7y,8y),(9y,10y,11y,12y))< (13y,14y,15y,16y)
6	*x* do ponto externo da sobrancelha direita deve ser menor que o ponto interno da mesma sobrancelha, que deve ser menor que o ponto interno da sobrancelha esquerda, que deve ser menor que o ponto externo da mesma sobrancelha	1x<2x<3x<4x

Obs.: P*x* representa os valores do eixo horizontal e P*y* os valores do eixo vertical para o ponto P

Conforme descrito na Seção 4.2.3, do vetor de pontos extremos, atualizado pelo rastreamento de pontos, se chega ao vetor de estados das características faciais (VECF). A Tabela 4.3 descreve como os elementos do vetor de estados das características faciais são obtidos do vetor de pontos extremos. Considera-se o primeiro vetor resultante sendo de uma face neutra ($VECF_n$) e os subsequentes de face com expressão ou não ($VECF_e$). Da subtração entre os vetores de estados das características faciais de uma face neutra com uma com expressão resulta-se o vetor de características (VC), que tem seus elementos normalizados com base na distância entre os cantos internos dos olhos (DO - pontos 8 e 9), conforme Equação 4.2. A normalização é necessária para corrigir diferenças de escala na imagem, isto é, evitar que sejam comparados dados coletados entre faces com diferentes proximidades da câmera. Já, a opção pelos cantos internos dos olhos foi adotada, pois a distância entre eles não se altera mesmo com a presença de expressões faciais. Após essa primeira normalização, o vetor resultante é novamente normalizado, dessa vez para manter seus valores escalares entre -1 e 1. Neste caso é feita uma busca pelo maior valor absoluto, que tem seu módulo utilizado para normalizar todo o vetor. O vetor de características resultante (VC_N) é então submetido à classificação para obtenção de

AUs presentes na expressão facial.

$$VC = \sum_{i=1}^{17} \left(\frac{VECF_e(i)}{DO_e} - \frac{VECF_n(i)}{DO_n} \right). \tag{4.2}$$

Tabela 4.3: Descrição dos estados das características faciais.

Elemento	Estados da característica facial	Descrição do estados
1	sobrancelha externa direita	movimento vertical do ponto 1
2	sobrancelha externa esquerda	movimento vertical do ponto 4
3	sobrancelha interna direita	movimento vertical do ponto 2
4	sobrancelha interna esquerda	movimento vertical do ponto 3
5	distância entre sobrancelhas	distância euclidiana entre pontos 2 e 3
6	abertura do olho direito	distância euclidiana entre pontos 6 e 7
7	abertura do olho esquerdo	distância euclidiana entre pontos 10 e 11
8	pálpebra superior direita	movimento vertical do ponto 6
9	pálpebra superior esquerda	movimento vertical do ponto 10
10	pálpebra inferior direita	movimento vertical do ponto 7
11	pálpebra inferior esquerda	movimento vertical do ponto 11
12	largura da boca	distância euclidiana entre pontos 13 e 16
13	abertura da boca	distância euclidiana entre pontos 14 e 15
14	lábio superior	movimento vertical do ponto 14
15	lábio inferior	movimento vertical do ponto 15
16	canto direito da boca	movimento vertical do ponto 13
17	canto esquerdo da boca	movimento vertical do ponto 16

Obs.: os movimentos verticais são as diferenças entre os eixos verticais (Px) dos pontos e centro dos olhos.

A classificação dos códigos FACS é realizada por duas redes neurais tipo *multi-layer perceptron feed-forward* com algoritmo de aprendizado *iRPROP-*, fornecidas pela biblioteca FANN (NISSEN, 2003). Após algumas tentativas, a melhor configuração encontrada é similar a utilizada por (TIAN; KANADE; COHN, 2001), que tem duas dessas redes neurais: uma para os AUs superiores (AUs 1, 2, 4, 5, 6, 7 e 9) e outra para os inferiores (AUs 10, 11, 12, 15, 16, 17, 20, 22, 23, 24, 25, 26 e 27). Em sua construção, cada RNA possui uma camada oculta de 28 e 52 neurônios para os AUs superiores (RNA1) e inferiores (RNA2), respectivamente. A configuração da camada oculta foi obtida multiplicando o número de neurônios das saídas de cada rede e avaliando o seu desempenho, sendo quatro vezes o número de saídas a configuração encontrada como mais adequada (RNA1 = 4×7; RNA2 = 4×13), ou seja, que convergia mais próximo ao erro estimado de 1%. Como entrada, as redes recebem os estados das características faciais (Tabela 4.3), formatado no vetor de características normalizado (VC_N), sendo os elementos 1 a 11 para a RNA1, e 12 ao 17 para a RNA2. Os neurônios da camada de saída têm como resultado um valor contínuo no intervalo de 0 a 1, que passam por um limiar que discretiza seu resultado em 0 ou 1. Este limiar está ajustado para 0,5, o que significa que os neurônios que atingem a

partir desse valor são considerados como uma ocorrência positiva (valor igual a 1) do código AU representado por ele. A Figura 4.9 mostra a estrutura das redes neurais construídas, que têm outros detalhes sobre desempenho no Capítulo 5.

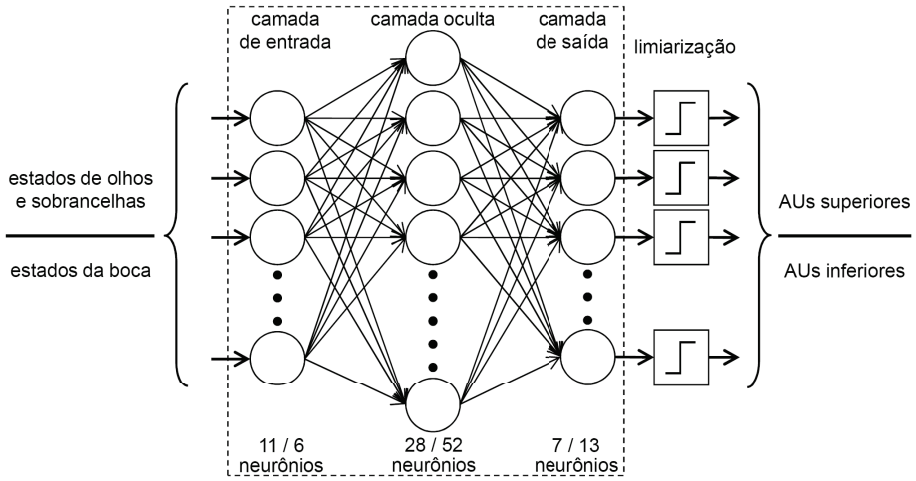


Figura 4.9: Estrutura das redes neurais 1 (para AUs superiores) e 2 (para AUs inferiores).

Em sua aplicação prática, as redes recebem a sua parte do vetor de características (que contém os estados das características faciais) de cada *frame* lido. Da mesma forma, para cada *frame* lido é retornado o vetor de AUs, utilizado posteriormente na inferência da emoção.

4.3.4 Inferência da emoção

Algumas combinações de AUs estão presentes em expressões faciais que ilustram as emoções básicas (EKMAN; FRIESEN; HAGER, 2002b). Conhecidos os AUs contidos em uma expressão facial, é possível obter a emoção por eles representados. Por exemplo, a emoção de medo pode ser caracterizada pela presença dos AUs 1+2+4+5+25, conforme Tabela 2.1. Essa mesma tabela foi a base da construção de uma árvore de decisão encarregada de classificar emoções de AUs, conforme descrito nesta seção.

Todas as possíveis combinações da Tabela 2.1 foram retabuladas considerando os AUs 1, 2, 4, 5, 6, 7, 9, 10, 11, 12, 15, 16, 17, 20, 22, 23, 24, 25, 26 e 27. Nesta nova tabela foram retirados os AUs 54 e 64 e as medidas de intensidade e simetria, por serem dispensáveis (opcionais) na inferência e não detectáveis pela aplicação. A tabela resultante contém 116 possíveis combinações entre as seis emoções básicas (7 para surpresa, 15 para medo, 2 para alegria, 24 para tristeza, 6 para repulsa, e 60 para raiva).

Esta tabela contendo 116 combinações para as seis emoções básicas foi submetida a ferramenta See5 (RULEQUESTRESEARCH, 2001) para a construção da árvore de decisão

(chamada de **em1**) e de um conjunto de regras (chamado de **em2**). See5 é a versão comercial e atualizada do algoritmo de classificação C4.5 (QUINLAN, 1993), que é utilizada para a indução de árvores de decisão ou conjunto de regras com base nas informações contidas em uma base de dados. A Figura 4.10 exibe a árvore de decisão construída pelo See5 sobre a tabela de combinações de AUs. Pode-se ver nesta figura que a existência de determinados AUs (caixas) denuncia uma provável emoção (elipses).

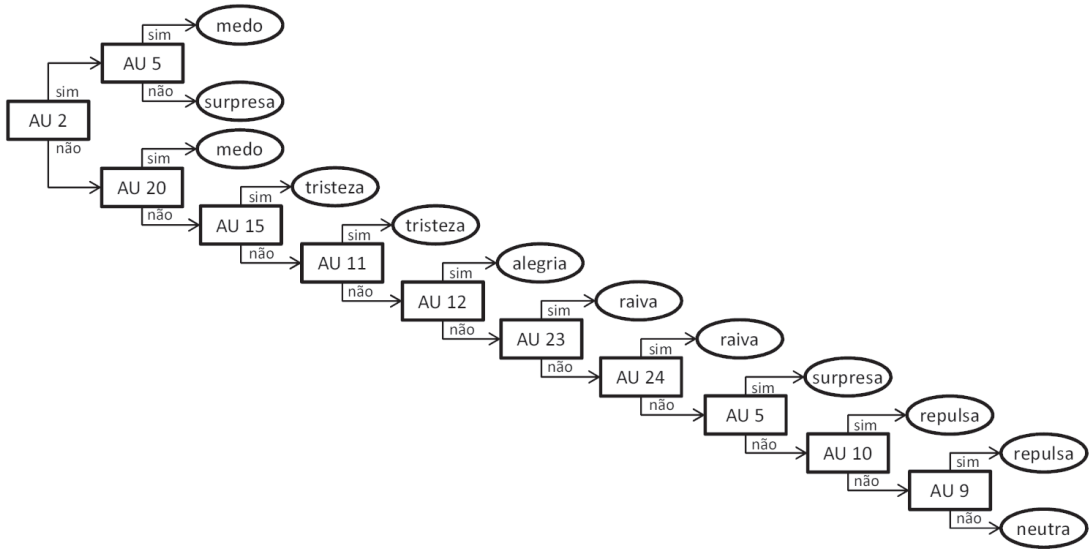


Figura 4.10: Árvore de decisão sobre emoção baseada na presença de AUs

As definições contidas na árvore de decisão (**em1**) (Figura 4.10) e no conjunto de regras (**em2**) foram implementadas dentro do quarto módulo da aplicação, utilizando para isto cadeias de "ifs" para realizar as inferências de AUs para emoções. Estas implementações recebem como entrada 20 AUs considerados na aplicação na forma de um vetor, fornecido pelas redes neurais. O vetor de AUs é constituído por variáveis booleanas em que cada elemento corresponde a um AU específico, conforme pode ser visto na Figura 4.11.

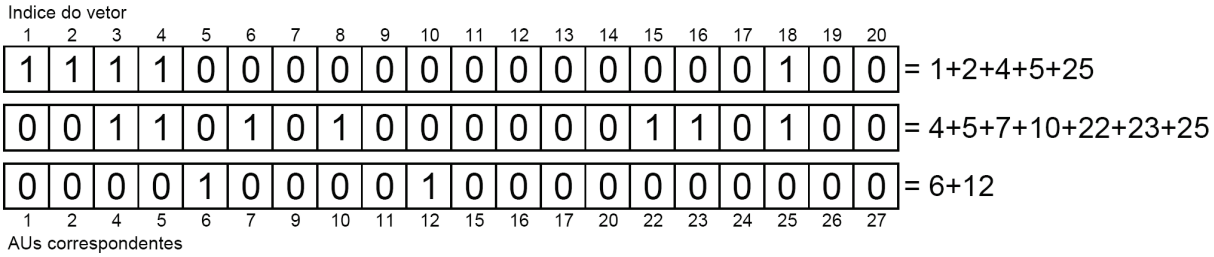


Figura 4.11: Exemplo de vetor de AUs.

Uma terceira forma de inferência de emoções, mais simplificada, foi implementada (renomeada como **em3**). Ela utiliza um conjunto de regras desenvolvido por Lucey et al. (2010),

mostrado na Tabela 4.4. Esse inferidor (**em3**) é utilizando em conjunto com os outros dois (**em1** e **em2**) para obtenção da emoção inferida (**emX**). É realizada uma eleição, sendo eleita como emoção, aquela que ocorrer mais vezes entre os inferidores. Caso não haja concordância entre os três inferidores, é eleita a emoção de **em1**. Havendo condições favoráveis (face e centro de olhos detectados, regras na Tabela 4.2 atendidas), este processo é realizado sobre cada *frame*, em tempo real, durante a execução do sistema.

Tabela 4.4: Regras de emoções de acordo com ocorrência de determinados AUs.

Emoção	Regra de ocorrência
Surpresa	1+2 ou 1+5A/B
Medo	1+2+4 ou 1+2+5E
Alegria	12
Tristeza	1+4+11 ou 1+4+15 ou 6+15
Repulsa	9 ou 10
Raiva	23 ou 24

5 AVALIAÇÃO DO SISTEMA

Para analisar o desempenho alcançado pela aplicação na tarefa de obter os códigos FACS e posteriormente as seis emoções básicas, tornou-se necessário o uso de bases de faces que contivessem esses dados em suas imagens. Nesse sentido, foram selecionadas três bases: JAFFE (*The Japanese Female Facial Expression Database*) (LYONS et al., 1998), CK+ (*The Extended Cohn-Kanade Dataset*) (LUCEY et al., 2010) e MPI-FVD (*Max Planck Institute Face Video Database*) (MPI, 2011).

A base JAFFE contém 213 imagens estáticas, em tons cinza, com resolução de 256×256 pixels, de 10 mulheres japonesas expressando, cada uma, as seis expressões básicas, além da expressão neutra. Já a CK+ contém 593 sequências de imagens estáticas, em tons de cinza e coloridas (poucos exemplos), com resolução de 640×490 ou 640×480 pixels, de 123 pessoas de sexos e etnicidades variadas, que realizam diversas expressões faciais, registradas em forma de códigos FACS e emoções básicas (em alguns casos). A última base utilizada é a MPI-FVD, que apresenta 246 vídeos coloridos, com resolução de 786×576 pixels, de uma pessoa executando códigos FACS em ângulos de captura variados.

Como a aplicação opera sobre imagens em sequência, as imagens estáticas da base CK+ foram transformadas em vídeos. O mesmo não pode ser realizado com a base JAFFE, que não possui imagens da transição entre a face neutra e com expressão. Neste caso, é realizada apenas a comparação entre imagem da face neutra e com expressão, eliminando a etapa de rastreamento de pontos.

Nas próximas seções são encontradas descrições dos experimentos realizados, bem como a análise dos resultados obtidos. Finalizando o capítulo, a Seção 5.2 traz um parecer onde considerações sobre os dados coletados e seus resultados são expostos.

5.1 Experimentos

Desde o início da implementação até as últimas validações, foram utilizados dados das bases de faces consideradas neste trabalho. Sua função foi de auxiliar em ajustes da aplicação, testes de classificadores e ser base para avaliação do desempenho do sistema.

Foram conduzidos alguns experimentos de avaliação realizados dentro de três cenários, que foram construídos para atender a diferentes formas de classificação ou de ambientes, onde foram submetidas as amostras. A Tabela 5.1 contém os dados que identificam cada cenário e seus experimentos, com informações sobre base utilizada, que tiveram como objetivos obter o desempenho na classificação de AUs e na inferência de emoções.

Tabela 5.1: Cenários e experimentos utilizados para avaliação da aplicação.

		Base/Exemplos utilizados			
Cenário	**Experimento**	**CK+**	**MPI-FVD**	**JAFFE**	**Outras**
-	Pré-avaliação	10	1	10	10
-	Teste	468	20	213	-
1	FACS	468	20	-	-
	Emoção	381	-	213	-
2	Emoção - RNA-EMO	381	-	213	-
3	Emoção - webcam	-	-	-	30

Os resultados obtidos pela aplicação nos diversos experimentos foram medidos pelas equações 5.1, quando se utiliza a amostra como referência, e 5.2 quando são tomados os AUs como base (TIAN; KANADE; COHN, 2001).

$$\textit{Taxa de reconhecimento} = \frac{\textit{número total de } \textbf{\textit{exemplos da amostra}} \textit{ reconhecidos corretamente}}{\textit{número total de } \textbf{\textit{exemplos da amostra}}}. \tag{5.1}$$

$$\textit{Taxa de reconhecimento} = \frac{\textit{número total de } \textbf{\textit{AUs}} \textit{ reconhecidos corretamente}}{\textit{número total de } \textbf{\textit{AUs}}}. \tag{5.2}$$

A medida de desempenho de taxa de reconhecimento (TREC), normalmente empregada em trabalhos relacionados, identifica quanto o sistema classificou corretamente os exemplos conforme o esperado, ou seja, seu sucesso[1]. Em se tratando de exemplos da amostra como base, um resultado esperado é uma classificação que retorne a combinação exata de códigos

[1] Sucesso: exemplo é classificado corretamente. Erro: exemplo é classificado erroneamente.

AUs presentes em um exemplo, ou a mesma emoção (conforme Equação 5.1). Utilizando AUs como base, o resultado obtido é a proporção de AUs corretamente classificados em relação ao número de AUs (conforme Equação 5.2). Além da taxa de reconhecimento, será utilizada a taxa de VP e a taxa de VN (KOHAVI; PROVOST, 1998), adotadas por serem consideradas adequadas para expressar o desempenho da aplicação. Mais detalhes sobre os resultados obtidos nos experimentos são descritos nas próximas seções.

5.1.1 Pré-avaliação

O objetivo da pré-avaliação foi realizar, durante a implementação da aplicação, testes sobre o sistema. Para isso, foi utilizada uma amostra com dados das bases de faces consideradas no trabalho. Essa amostra era pequena e composta por pouco mais de 30 imagens, variadas o suficiente para que fossem capazes de fornecer alguns dos desafios contidos na Figura 2.13, como variação de luminosidade (alta ou baixa), baixa qualidade (resolução baixa) e inclinação da face. Estas imagens foram submetidas ao sistema de forma a constatar problemas e realizar ajustes, principalmente em relação a FaD e FeD, mas também no retreinamento das redes neurais utilizadas no trabalho.

Após obtidos valores considerados satisfatórios, durante a implementação e testes iniciais, foi realizado um teste final mais abrangente, utilizando uma amostra de dados maior para que os últimos ajustes fossem feitos. De posse dos resultados desse teste, foi levantada uma série de constatações que levaram a alterações na aplicação, descritas nos próximos parágrafos.

Originalmente, havia a necessidade de utilizar novamente o classificador de detecção de olhos após a inclinação da face ser realizada, para que fosse identificado o novo posicionamento dos olhos. Isso foi dispensado, pois as coordenadas dos olhos na face sem inclinação passaram a ser identificados apenas calculando a rotação específica desses dois pontos ao ângulo correto.

Os dados do vetor de características (o que é enviado às redes neurais) eram normalizados apenas com base na distância entre os cantos internos dos olhos (pontos 8 e 9). Porém, a normalização agora passa a considerar os cantos externos dos olhos (pontos 5 e 12) para reforçar essa distância. Para isso, tomando como base estudos de Farkas e Munro (1987), onde sabe-se que a distância entre os cantos internos dos olhos é aproximadamente a largura de um olho, utiliza-se os pontos externos dos olhos para encontrar de forma mais robusta a distância entre os cantos internos dos olhos. Isso garante uma normalização dos dados mais confiável, pois minimiza efeitos de um ponto mal posicionado sobre o olho.

Foi alterada a normalização dos dados de entrada da RNA. Inicialmente, os valores

eram normalizados pelo módulo do maior valor absoluto do vetor de características, porém foi constatado que o intervalo de -1 a 1, utilizado na normalização, já encontrava-se presente nos dados não normalizados. Nos casos de exceção, em que valores ultrapassarem os limites inferiores ou superiores, estes recebem o valor de limite. Devido a essa alteração, foram treinadas novamente as RNAs com pesos normalizados pela nova definição.

A ordem de avaliação das regras apresentada na Tabela 4.4, utilizada pelo inferidor **em3**, que verificava em sequência as regras para surpresa, medo, alegria, tristeza, repulsa e raiva, foi alterada para a ordem alegria, raiva, repulsa, medo, surpresa e tristeza. Isso foi necessário, pois existem alguns códigos AUs em comum nas emoções de surpresa e medo, e medo e tristeza, que na nova ordenação possibilitam verificação de regras mais complexas antes das mais simples.

5.1.2 Cenário 1

O primeiro cenário foi criado com o intuito de avaliar o desempenho da aplicação construída sobre as bases de faces consideradas, tanto na classificação de códigos FACS, quanto na inferência de emoções . Conforme já exibido na Tabela 5.1, foi executado um experimento para avaliar a classificação de códigos FACS e outro para a inferência da emoção, como descrito a seguir.

Experimento 1: classificação de códigos FACS

Foram consideradas neste experimento, amostras de 468 exemplos sobre a base CK+, e de 20 exemplos sobre a MPI-FVD. Entretanto, devido problemas nas etapas de FaD e FeD (ver seção 5.2), 195 e 3 exemplos das respectivas bases foram descartados das amostras.

Considerando que a amostra válida da base CK+ é composta por 273 exemplos, a taxa de reconhecimento sobre a amostra foi 8,79% (conforme Tabela 5.2). Em outras palavras, em 8,79% da amostra os 20 códigos AUs esperados foram corretamente inferidos, ou seja, foram exatamente os mesmos. Porém, analisando separadamente os AUs inferiores (1, 2, 4, 5, 6, 7 e 9) e superiores (10, 11, 12, 15, 16, 17, 20, 22, 23, 24, 25, 26 e 27), a taxa de reconhecimento sobre a amostra foi de 49,82% e 22,71% respectivamente. Isso demonstra que existe um problema maior na classificação dos AUs inferiores.

Conforme exibido na Tabela 5.3, considerando códigos AUs, a taxa de reconhecimento é de 53,83% na base CK+. Dividindo esse percentual entre AUs superiores e inferiores, as taxas são de 45,43% e 61,61%, respectivamente. Nota-se que, agora, a taxa de reconhecimento para os AUs inferiores obteve um desempenho melhor. Isso se deve ao fato de a representividade

Tabela 5.2: AUs na base CK+: taxas de reconhecimento sobre a amostra.

Base	Escopo	Taxa REC (amostra)	Taxa VP	Taxa VN
MPI-FVD	AUs superiores	58,82	60,00	71,85
	AUs inferiores	11,76	75,00	56,85
	todos AUs	5,88	70,59	62,14
CK+	AUs superiores	49,82	45,43	87,87
	AUs inferiores	22,71	61,61	81,70
	todos AUs	8,79	54,03	83,70

dos AUs inferiores ser maior: são 13 códigos, contra 7.

Tabela 5.3: AUs na base CK+: taxas de reconhecimento sobre total de AUs.

		MPI-FVD				CK+			
	AU	Casos na amostra	Rec. correto	Rec. incorreto	Taxa REC (nro AUs)	Casos na amostra	Rec. correto	Rec. incorreto	Taxa REC (nro AUs)
	0	-	-	-	-	11	4	7	36,36%
superiores	1	1	0	1	0,00%	93	65	28	69,89%
	2	1	0	1	0,00%	69	45	24	65,22%
	4	1	-	-	-	87	24	63	27,59%
	5	1	-	-	-	70	40	30	57,14%
	6	1	1	0	100,00%	62	18	44	29,03%
	7	1	1	0	100,00%	50	9	41	18,00%
	9	1	1	0	100,00%	29	8	21	27,59%
total		7	3	2	**42,86%**	460	209	251	**45,43%**
inferiores	10	1	0	1	0,00%	8	3	5	37,50%
	11	1	1	0	100,00%	8	7	1	87,50%
	12	1	1	0	100,00%	58	35	23	60,34%
	15	1	1	0	100,00%	39	23	16	58,97%
	16	1	1	0	100,00%	10	4	6	40,00%
	17	1	1	0	100,00%	87	63	24	72,41%
	20	1	0	1	0,00%	27	21	6	77,78%
	22	1	1	0	100,00%	2	1	1	50,00%
	23	1	-	-	-	22	10	12	45,45%
	24	1	1	0	100,00%	19	6	13	31,58%
	25	1	0	1	0,00%	155	104	51	67,10%
	26	1	1	0	100,00%	30	12	18	40,00%
	27	1	1	0	100,00%	56	32	24	57,14%
total		13	9	3	**69,23%**	521	321	200	**61,61%**
TOTAL		20	12	5	**60,00%**	992	534	458	**53,83%**

Também foram obtidos resultados de 17 exemplos da base MPI-FVD. A Tabela 5.2 mostra que a taxa de reconhecimento sobre a amostra foi de 5,88%, 58,82% sobre os AUs superiores e 11,76% sobre os AUs inferiores. Em relação a taxa de reconhecimento sobre AUs (Tabela 5.3), o desempenho sobre todos AUs foi de 60,00%, 42,86% sobre os superiores e 69,23% sobre os inferiores. A base MPI-FVD apresentou resultados pouco melhores em relação a CK+, porém a sua amostra é muito pequena. O comportamento entre as bases em relação a AUs superiores e inferiores é similar, demonstrando que de fato há mais problemas na

classificação de AUs inferiores.

Experimento 2: inferência de emoção

No segundo experimento do cenário 1, foram utilizadas amostras de 381 exemplos sobre a base CK+ e de 213 exemplos sobre a base JAFFE. Da mesma forma como ocorreu com o experimento 1 (pelos mesmos motivos), foram descartadas 157 e 56 exemplos das respectivas bases.

A maior taxa de reconhecimento de emoções para a base CK+ foi de 28,57%, com taxa média de reconhecimento por emoção de 37,47% $\pm$ 26,40%. Esse desempenho foi obtido pelo inferidor **em1**, conforme listado pela Tabela 5.4.

Tabela 5.4: Taxa de reconhecimento de emoções na base CK+.

Base	Inferidor de emoção	Taxa REC	Media das taxas de rec. por emoção
JAFFE	em1	24,84	23,88 ± 15,75
	em2	25,50	23,92 ± 15,91
	em3	**27,39**	27,56 ± 18,87
	emX	25,48	25,15 ± 15,74
CK+	em1	**28,57**	37,47 ± 26,40
	em2	24,11	30,79 ± 25,99
	em3	28,57	31,99 ± 32,62
	emX	28,13	36,70 ± 24,60

Pode-se ver na Tabela 5.5 o desempenho do inferidor **em1** sobre cada emoção. É visível, pelos resultados, uma tendência para inferência sobre as emoções de medo e neutra, que foram as únicas a atingirem percentual de reconhecimento dentro do esperado (os sucessos ocorreram em maior número na própria emoção).

Tabela 5.5: Matriz de confusão para o inferidor em1 sobre a base CK+.

		Emoção obtida													
		Raiva		Alegria		Tristeza		Repulsa		Surpresa		Medo		Neutra	
		%	#	%	#	%	#	%	#	%	#	%	#	%	#
Emoção esperada	Raiva	**25,93**	**7**	0,00	0	11,11	3	0,00	0	3,70	1	29,63	8	29,63	8	27
	Alegria	3,64	2	**32,73**	**18**	12,73	7	0,00	0	1,82	1	40,00	22	9,09	5	55
	Tristeza	8,33	2	0,00	0	**25,00**	**6**	0,00	0	8,33	2	25,00	6	33,33	8	24
	Repulsa	16,67	5	0,00	0	10,00	3	**20,00**	**6**	0,00	0	20,00	6	33,33	10	30
	Surpresa	1,69	1	0,00	0	5,08	3	0,00	0	**10,17**	**6**	67,80	40	15,25	9	59
	Medo	0,00	0	0,00	0	5,56	1	0,00	0	5,56	1	**66,67**	**12**	22,22	4	18
	Neutra	0,00	0	0,00	0	9,09	1	0,00	0	0,00	0	9,09	1	**81,82**	**9**	11
			17		18		24		6		11		95		53	

O experimento sobre a amostra da base JAFFE obteve melhor desempenho com inferidor **em3**, que atingiu uma taxa de reconhecimento um pouco menor em relação a base CK+:

27,39%, com taxa média de reconhecimento por emoção de 27,56% ± 18,87% (Tabela 5.4). O desempenho do inferidor **em3** sobre cada emoção pode ser visto na Tabela 5.6. Constata-se que surpresa, tristeza e medo foram inferidas com maior percentual em outras emoções, e que existe uma tendência de inferência nas emoções de raiva ou neutra.

Tabela 5.6: Matriz de confusão para o inferidor em3 sobre a base JAFFE.

		Emoção obtida													
		Raiva		Alegria		Tristeza		Repulsa		Surpresa		Medo		Neutra	
		%	#	%	#	%	#	%	#	%	#	%	#	%	#
Emoção esperada	Raiva	**36,84**	**7**	0,00	0	0,00	0	36,84	7	0,00	0	0,00	0	26,32	5	19
	Alegria	13,04	3	**52,17**	**12**	4,35	1	8,70	2	4,35	1	4,35	1	13,04	3	23
	Tristeza	41,67	10	12,50	3	**12,50**	**3**	12,50	3	12,50	3	0,00	0	8,33	2	24
	Repulsa	21,05	4	5,26	1	15,79	3	**21,05**	**4**	0,00	0	15,79	3	21,05	4	19
	Surpresa	30,43	7	4,35	1	0,00	0	4,35	1	**4,35**	**1**	21,74	5	34,78	8	23
	Medo	20,00	5	4,00	1	0,00	0	12,00	3	12,00	3	**16,00**	**4**	36,00	9	25
	Neutra	33,33	8	12,50	3	0,00	0	4,17	1	0,00	0	0,00	0	**50,00**	**12**	24
			44		21		7		21		8		13		43	

5.1.3 Cenário 2

Considerando que os resultados apresentados no cenário 1 mostraram-se deficientes na inferência de emoção através de códigos FACS, foi criado o cenário 2, para avaliar uma alternativa a atual inferência de emoções. Trata-se de uma rede neural que combina as RNA1 e RNA2 em uma única rede. Ela foi construída utilizando a mesma estratégia de construção e a mesma base de dados utilizada nas RNA1 e RNA2 (JAFFE e CK+), porém, considerando e tendo apenas como saída as emoções básicas. A rede construída possui 17 neurônios de entrada (os dois vetores da RNA1 E RNA2), duas camadas ocultas com 25 neurônios cada e sete saídas (emoções básicas + neutra). Ela foi chamada de **RNA-EMO**.

No cenário 2, foram utilizadas as mesmas amostras selecionadas no cenário 1, exceto a base MPI-FDV, que foi descartada por não apresentar emoções. A taxa de reconhecimento obtida pela RNA-EMO sobre emoções na base CK+ foi de 57,14%, com taxa média de reconhecimento por emoção de 59,61% ± 23,43% (Tabela 5.7). A Tabela 5.8 discrimina o desempenho da RNA-EMO sobre cada emoção, onde constata-se uma tendência geral para classificação de emoção neutra e uma grande confusão na classificação de raiva.

Tabela 5.7: Taxa de reconhecimento de emoções da RNA-EMO sobre a base CK+.

Base	Inferidor de emoção	Taxa REC	Media das taxas de rec. por emoção
JAFFE	RNA-EMO	43,95	44,88 ± 24,29
CK+	RNA-EMO	57,14	59,61 ± 23,43

Tabela 5.8: Matriz de confusão da RNA-EMO sobre a base CK+.

		Emoção obtida														
		Raiva		Alegria		Tristeza		Repulsa		Surpresa		Medo		Neutra		
		%	#	%	#	%	#	%	#	%	#	%	#	%	#	
Emoção esperada	Raiva	**14,81**	**4**	11,11	3	11,11	3	29,63	8	0,00	0	3,70	1	29,63	8	27
	Alegria	1,82	1	**52,73**	**29**	5,45	3	12,73	7	0,00	0	18,18	10	9,09	5	55
	Tristeza	0,00	0	0,00	0	**62,50**	**15**	0,00	0	0,00	0	4,17	1	33,33	8	24
	Repulsa	0,00	0	0,00	0	3,33	1	**86,67**	**26**	3,33	1	0,00	0	6,67	2	30
	Surpresa	6,78	4	0,00	0	22,03	13	0,00	0	**57,63**	**34**	0,00	0	13,56	8	59
	Medo	0,00	0	5,56	1	22,22	4	0,00	0	5,56	1	**61,11**	**11**	5,56	1	18
	Neutra	0,00	0	0,00	0	9,09	1	0,00	0	9,09	1	0,00	0	**81,82**	**9**	11
			9		33		40		41		37		23		41	

A amostra da base JAFFE obteve desempenho inferior em relação a CK+ com a RNA-EMO: sua taxa de reconhecimento foi de 43,95%, com taxa média de reconhecimento por emoção de 44,88% $\pm$ 24,29% (Tabela 5.7). Analisando a Tabela 5.9, nota-se uma tendência na classificação da emoção de raiva, além de uma completa confusão na classificação da emoção de medo.

Tabela 5.9: Matriz de confusão da RNA-EMO sobre a base JAFFE.

		Emoção obtida														
		Raiva		Alegria		Tristeza		Repulsa		Surpresa		Medo		Neutra		
		%	#	%	#	%	#	%	#	%	#	%	#	%	#	
Emoção esperada	Raiva	**73,68**	**14**	0,00	0	21,05	4	5,26	1	0,00	0	0,00	0	0,00	0	19
	Alegria	8,70	2	**65,22**	**15**	4,35	1	0,00	0	8,70	2	8,70	2	4,35	1	23
	Tristeza	41,67	10	4,17	1	**41,67**	**10**	0,00	0	0,00	0	0,00	0	12,50	3	24
	Repulsa	10,53	2	5,26	1	26,32	5	**31,58**	**6**	15,79	3	0,00	0	10,53	2	19
	Surpresa	0,00	0	4,35	1	21,74	5	0,00	0	**47,83**	**11**	0,00	0	26,09	6	23
	Medo	12,00	3	0,00	0	24,00	6	16,00	4	28,00	7	**0,00**	**0**	20,00	5	25
	Neutra	25,00	6	4,17	1	8,33	2	4,17	1	4,17	1	0,00	0	**54,17**	**13**	24
			37		19		33		12		24		2		30	

5.1.4 Cenário 3

Um último cenário foi criado para testar e avaliar a aplicação sobre o ambiente foco de sua construção: aquele em que há captura de imagens por *webcam* de usuários a frente do computador. Para isso foram utilizadas imagens coletadas de uma pessoa a frente do computador, executando expressões faciais emocionais, por diferentes *webcams*, em diferentes configurações.

Foram utilizadas na coleta de dados do Cenário 3, três *webcams*: Logitech QuickCam Pro 5000 (**cam1**), A4 Tech PK-5 (**cam2**) e a câmera embutida do *netbook* Asus EeePC 1000H (**cam3**). Por meio dessas *webcams* foram coletadas cinco amostras, onde em cada uma foi executada a sequência das emoções de medo, surpresa, repulsa, tristeza, alegria e raiva, totalizando 30 exemplos. A **cam1** e a **cam2** capturaram, cada uma, uma sequência com resolução

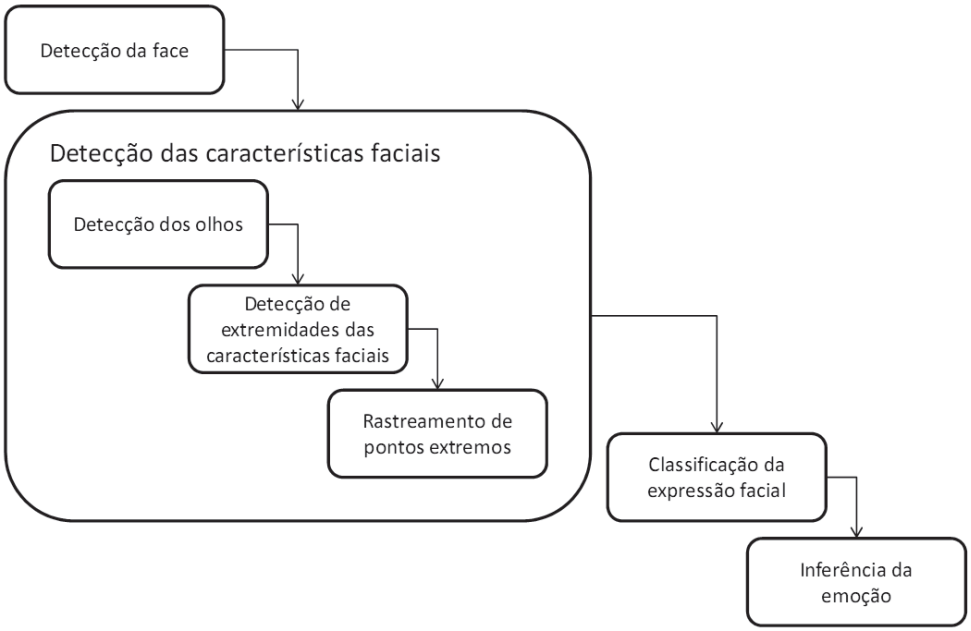
de 320×240 pixels e outra com resolução de 640×480 pixels. A **cam3** capturou apenas uma sequência com resolução de 640×480 pixels.

As amostras do Cenário 3 foram submetidos a inferência da emoção ao **em1**, **em2**, **em3** e RNA-EMO. As taxas de reconhecimento foram, respectivamente, 46,67%, 56,67%, 60,00%, e 63,33%. Em todos os inferidores foi constatado algum grau de tendência na obtenção de emoções, mas a RNA-EMO obteve melhores resultados, justamente por manter um equilíbrio maior na classificação individual das emoções.

5.2 Avaliação dos resultados

Os resultados obtidos pelo trabalho e descritos nas seções anteriores, foram atingidos após a execução de várias etapas da aplicação. Inicialmente, é realizada a detecção da face, em seguida a detecção de características faciais, posteriormente a classificação das características faciais, finalizando com a inferência da emoção. Cada uma dessas etapas é dependente da anterior (Figura 5.1) e pode encontrar obstáculos a ultrapassar, como os problemas de classificação originados por motivos mostrados na Figura 2.13. Isso significa que o desempenho máximo do sistema é relacionado com a combinação dos desempenhos de cada etapa.

Figura 5.1: Dependência entre etapas do sistema.



Considerando essa dependência, é possível ter dois grupos de avaliação: específico, sobre as etapas; geral, do sistema como um todo. A próxima seção avaliará o desempenho específico das etapas de classificação expressão facial e inferência de emoção.

5.2.1 Desempenho dos inferidores e classificadores

Foram construídas duas RNAs para realizar a classificação de expressões faciais sobre códigos FACS, RNA1 e RNA2. A RNA1 obteve taxa de reconhecimento de 27,88% sobre a amostra e 88,61% sobre os AUs. Já a RNA2 obteve taxa de reconhecimento de 40,38% sobre a amostra e 83,65% sobre os AUs, conforme Tabela 5.10.

Tabela 5.10: Desempenho de treinamento das RNAs.

RNA	Taxa REC (amostra)	Taxa REC (nro AUs)	MSE	Amostra de treinamento	Amostra de teste
RNA1 (AUs sup.)	40,38	83,65	0,15	429	104
RNA2 (AUs inf.)	27,88	88,61	0,11	429	104
RNA-EMO	**89,87**	-	0,01	244	79

Comparando os resultados obtidos no Cenário 1, na base CK+, com os obtidos sobre a amostra de teste da RNA1 e RNA2, se torna evidente que o comportamento das redes se reflete nos resultados da aplicação. As taxas de reconhecimento sobre amostra para AUs superiores foram maiores do que para os inferiores; e as taxas de reconhecimento sobre AUs foi maior para os AUs inferiores do que para os superiores, em ambos os casos.

Diferentemente do comportamento similar encontrado nos resultados dos inferidores de emoção e redes neurais classificadores de FACS, em testes e em uso pela aplicação, o desempenho da RNA-EMO sobre a amostra de teste (89,87% de taxa de reconhecimento - Tabela 5.10) é um pouco diferente dos encontrados na aplicação. Os resultados expandidos dos testes da RNA-EMO (Tabela 5.11) apresentam valores equilibrados, entre 80% e 100%, mas na aplicação, essa faixa ficou entre 14,81% e 86,67% (Tabela 5.8). Uma possível explicação a isto é o tamanho da amostra de teste da RNA-EMO, que possui apenas 79 casos.

Tabela 5.11: Matriz de confusão para teste da RNA-EMO sobre a base CK+.

		Emoção obtida													
		Neutra		Surpresa		Medo		Alegria		Tristeza		Repulsa		Raiva	
		%	#	%	#	%	#	%	#	%	#	%	#	%	#
Emoção esperada	Neutra	**100,00**	**2**	0,00	0	0,00	0	0,00	0	0,00	0	0,00	0	0,00	0	2
	Surpresa	0,00	0	**93,75**	**15**	0,00	0	0,00	0	0,00	0	0,00	0	6,25	1	16
	Medo	0,00	0	0,00	0	**90,91**	**10**	9,09	1	0,00	0	0,00	0	0,00	0	11
	Alegria	0,00	0	0,00	0	0,00	0	**100,00**	**15**	0,00	0	0,00	0	0,00	0	15
	Tristeza	8,33	1	8,33	1	0,00	0	0,00	0	**83,33**	**10**	0,00	0	0,00	0	12
	Repulsa	0,00	0	0,00	0	0,00	0	0,00	0	0,00	0	**87,50**	**7**	12,50	1	8
	Raiva	6,67	1	0,00	0	0,00	0	0,00	0	6,67	1	6,67	1	**80,00**	**12**	15
			4		16		10		16		11		8		14	

Ambas redes (RNA1, RNA2 e RNA-EMO) utilizam em sua parametrização as bases de faces JAFFE e CK+. As quantidades utilizadas nos treinamentos e testes, além do MSE

(*Mean Square Error*) obtido no treinamento e as melhores medidas de desempenho podem ser consultadas na Tabela 5.10.

Para inferência de emoções são utilizadas inicialmente a árvore **em1**, o conjunto de regras **em2**, as regras **em3**, e o agregador de inferências **emX**. Embora **em1** e **em2** tenham sido construídos pelo See5, a taxa de reconhecimento do conjunto de regras **em2** obteve um desempenho melhor, de 71,93% (Tabela 5.12). O desempenho de **emX** na avaliação sobre a amostra de teste confirmou o seu comportamento intermediatista, que é de fornecer taxas de reconhecimento melhores que o valor mais baixo, mas piores que o mais alto.

Tabela 5.12: Desempenho dos inferidores de emoção sobre a base CK+.

Inferidor de emoção	Taxa REC (amostra)	Casos de treinamento	Amostra de teste	Erro
em1	69,79	116	374	3,40
em2	**71,93**	116	374	3,40
em3	62,30	-	374	-
emX	69,52	-	374	-

A Tabela 5.13 mostra de forma mais explícita o desempenho do inferidor **em2**, que teve melhores resultados em relação aos demais. Pode-se ver que a taxa de reconhecimento da emoção de surpresa (15,12%) ficou bem abaixo dos demais, sendo bastante confundida com medo (80,23%). Embora a emoção de tristeza (76,19%) tenha apresentado um resultado superior a surpresa, também nota-se uma confusão com raiva (14,29%) e neutra (9,52%). Essas constatações realizadas especificamente sobre os inferidores, explicam o desempenho baixo obtido pela aplicação para algumas emoções.

Tabela 5.13: Matriz de confusão para teste da em2 sobre a base CK+.

		Emoção obtida												
		Raiva		Alegria		Tristeza		Repulsa		Surpresa		Medo		Neutra		
		%	#	%	#	%	#	%	#	%	#	%	#	%	#	
Emoção esperada	Raiva	**97,87**	**46**	0,00	0	0,00	0	0,00	0	0,00	0	0,00	0	2,13	1	47
	Alegria	0,00	0	**95,56**	**86**	0,00	0	0,00	0	0,00	0	2,22	2	2,22	2	90
	Tristeza	14,29	6	0,00	0	**76,19**	**32**	0,00	0	0,00	0	0,00	0	9,52	4	42
	Repulsa	16,92	11	0,00	0	3,08	2	**80,00**	**52**	0,00	0	0,00	0	0,00	0	65
	Surpresa	1,16	1	0,00	0	1,16	1	2,33	2	**15,12**	**13**	80,23	69	0,00	0	86
	Medo	0,00	0	0,00	0	10,34	3	3,45	1	0,00	0	**86,21**	**25**	0,00	0	29
	Neutra	0,00	0	0,00	0	0,00	0	0,00	0	0,00	0	0,00	0	**100,00**	**15**	15
			64		86		38		55		13		96		22	

5.2.2 Desempenho do sistema

Conforme citado anteriormente, existem obstáculos dentro de cada uma das etapas que interferem na utilização da aplicação construída. A Tabela 5.14 exibe o impacto de obstáculos

sobre o sistema, que ocasionou descartes de 41,67% (195 exemplos) dos 468 exemplos da base CK+, para classificação de FACS; 41,21% de 381 exemplos da mesma base para inferência de emoções; 26,29% de descarte sobre a base JAFFE; e 15,00% sobre a base MPI-FVD.

Tabela 5.14: Exemplos considerados e descartados de cada base de faces.

Base	Exemplos	Classificados	Não classificados	Descarte (%)	Objeto de classificação
CK+	468	273	195	41,67	FACS
CK+	381	224	157	41,21	emoções
JAFFE	213	157	56	26,29	emoções
MPI-FVD	20	17	3	15,00	FACS

Pelos levantamentos realizados durante a coleta de dados, os principais problemas ocorrem na detecção de olhos e no posicionamento de pontos sobre extremidades das características faciais. Estes problemas afetam a utilização do sistema, mas para a medida de desempenho estes casos foram descartados, ou seja, entraram na avaliação somente exemplos das bases onde se pôde executar todas as etapas da aplicação. Porém, com menos exemplos, é possível que alguns AUs e emoções tenham sido prejudicados, devido a um possível desequilíbrio nas quantidades de exemplos. Dessa forma, promovendo melhorias (nas etapas de FaD e FeD) que apenas possibilitem o sistema realizar a inferência da emoção, pode-se melhorar os seus resultados de avaliação.

Analisando os resultados obtidos, é possível listar algumas outras medidas que podem ser tomadas para aumentar as taxas de reconhecimento (nas etapas de classificação de expressões e inferência de emoções). Considerando todos AUs, a maior taxa de reconhecimento obtida foi de 8,79%, que é uma medida baixa, embora deva-se considerar que se tratam de 20 códigos neste contexto. Sendo assim, foi analisada, primeiramente, a ocorrência de AUs na amostra, e constatou-se que alguns códigos FACS possuem poucos exemplos na base, como os AUs 10 (15/981), 11 (9/981), 16 (16/981) e 22 (9/981). A partir disso foi realizado um experimento excluindo os AUs com poucos representantes (10, 11, 16 e 22), excluindo o AU 9, que é de difícil detecção direta pela aplicação, terminando por tratar os AUs 25, 26 e 27, devido a sua similaridade (conferir Apêndice A), como sendo apenas o AU 25 (26 e 27 seriam intensidades do 25). A Tabela 5.15 mostra o resultado obtido na base CK+ após estas exclusões, onde pode-se ver que a taxa de reconhecimento sobre a amostra nos AUs superiores passou de 49,82% para 55,68%, nos AUs inferiores de 22,71% para 26,74% e sobre todos AUs de 8,79% para 28,21%.

Outro problema constatado foi a grande ocorrência de falsos positivos (Tabela 5.16). Isso quer dizer que as redes neurais RNA1 e RNA2 estão muito sensíveis em sua classificação, o que dependeria da revisão de seu modelo; ou que os dados submetidos a elas estão sofrendo

Tabela 5.15: AUs reduzidos na base CK+: taxas de reconhecimento sobre a amostra.

Base	Escopo	Taxa REC (amostra)	Taxa VP	Taxa VN
CK+	AUs superiores	55,68	46,64	87,08
	AUs inferiores	26,74	65,93	77,11
	todos AUs	28,21	56,02	81,55

alguma perturbação, que pode ser na normalização, ou mesmo na obtenção dos pontos extremos sobre as características faciais. Entretanto, um motivo que pode explicar parcialmente o grande número de falsos positivos, é a confusão que pode ocorrer entre AUs similares, como AU1 e AU2, AU6 e AU7, e AU25 e AU26 (TIAN; KANADE; COHN, 2001).

Tabela 5.16: Taxas de VP e FP de AUs na base CK+.

AU	1	2	4	5	6	7	9	10	11	12	15	16	17	20	22	23	24	25	26	27	TOTAL
VP	65	45	24	40	18	9	8	3	7	35	23	4	63	21	1	10	6	104	12	32	530
FP	33	24	21	22	28	28	20	33	52	22	66	43	71	57	25	55	47	22	44	17	730

De forma geral, o desempenho do sistema demonstrou certa fragilidade na classificação de códigos FACS. Considerando a base CK+, que é mais completa e possui mais exemplos, a taxa de reconhecimento sobre AUs foi de 53,83%. Esse resultado, consequentemente, afetou o desempenho na inferência de emoção, que obteve como maior taxa de reconhecimento 28,57% com o inferidor **em1**. A rede neural RNA-EMO, que não tem dependência com códigos FACS sobre emoções, foi melhor em relação ao **em1**, com uma taxa de reconhecimento de 57,14%, atingindo 86,67% de reconhecimento na emoção de repulsa e 81,82% na emoção neutra sobre a base CK+. Surpreendentemente o melhor resultado da RNA-EMO ocorreu sobre o Cenário 3, que foi de 63,33% de reconhecimento, porém esta base é pequena e conta com imagens de uma única pessoa.

Embora problemas e dificuldades encontrados pelo sistema na classificação e inferência se reflitam nos resultados já comentados nesta seção, o desempenho dos classificadores é mais animador. Salienta-se o desempenho das RNAs (88,61% (RNA1), 83,65% (RNA2) e 89,87% (RNA-EMO) de taxa de reconhecimento) em seu treinamento, que pode ser considerado satisfatório, pois é próximo ao encontrado na literatura (ver capítulo 3, Trabalhos Relacionados). O mesmo pode ser dito sobre a árvore **em2**, que embora tenha um desempenho inferior (obteve reconhecimento de 71,93%), seu resultado pode ser considerado desejável. Além disso, um outro atenuante que existe é que assim como o sistema confundiu emoções, isso ocorre em outros sistemas de trabalhos relacionados e inclusive com humanos (SCHIANO et al., 2000; SEBE et al., 2005). Entre as causas está a ocorrência de AUs similares em emoções diferentes, como AUs 1 e 2, presentes nas emoções de medo e surpresa.

6 CONSIDERAÇÕES FINAIS

As expressões faciais são formas de comunicação não verbal amplamente utilizadas em nosso cotidiano. Elas externam, consciente ou inconscientemente, nossas manifestações em relação aos estímulos internos e externos a que somos submetidos. Pensando em trazer emoções encontradas nas relações homem-homem para homem-computador, o trabalho se propôs a inferir a emoção expressa na face de um usuário de computador. Mais especificamente, seu objetivo é capturar, por uma *webcam*, imagens do usuário, identificar nestas imagens sua face, e nela localizar boca, olhos e sobrancelhas. Dados dessas características faciais são utilizados para obtenção de expressões faciais, em forma de códigos FACS, que são utilizados para a inferência de emoções básicas contidas na face. Frente a isso, durante o levantamento bibliográfico, foram pesquisados métodos em Processamento de Imagens, Visão Computacional e Computação Afetiva focados no objetivo do trabalho.

Nas pesquisas para construção do referencial teórico e posteriormente nos trabalhos relacionados, constatou-se a existência de várias abordagens a serem adotadas para FaD, FeD, classificação de expressões faciais e inferência de emoções. O método de Viola-Jones (VIOLA; JONES, 2001) mostrou-se satisfatório para FaD e parte do processo de FeD, na classificação de olhos. Diversos métodos de Processamento de Imagens foram necessários para ajustar as imagens das características faciais, possibilitando, assim, a localização de seus pontos extremos e posterior rastreamento com o método *Pyramidal Lucas-Kanade*. Até este ponto, a biblioteca OpenCV (WILLOW GARAGE, 2010) foi de grande valia, fornecendo todos os métodos utilizados na aplicação. Obtidos os dados sobre as características faciais, estes foram submetidos às redes neurais que classificam as expressões faciais em códigos do sistema FACS (EKMAN; FRIESEN; HAGER, 2002a). Os códigos FACS (AUs) obtidos, que podem indicar a existência da manifestação de uma emoção, são utilizados por três inferidores distintos (uma árvore de decisão; um conjunto de regras baseado nessa árvore de decisão; um conjunto de regras construído sobre análise das combinações entre AUs e emoções) que obtêm, com base em um conjunto de AUs uma emoção. Posteriormente, durante as avaliações, foi construída uma rede neural como inferidor de emoções, devido a problemas na obtenção de códigos FACS

que afetavam o desempenho da árvore de decisão. A rede neural foi selecionada como classificador de expressões faciais, devido a seu sucesso em trabalhos anteriores. Optou-se pela árvore de decisão na inferência de emoções, utilizando a implementação da ferramenta See5 (RULEQUESTRESEARCH, 2001), por esta ser de abordagem simbólica e de simples e rápida implementação.

Foram realizadas no Capítulo 5 avaliações sobre o sistema construído, considerando três cenários de teste. Neles foram utilizadas imagens das bases de faces JAFFE (LYONS et al., 1998), CK+ (LUCEY et al., 2010), MPI-FVD (MPI, 2011), além de imagens coletadas por três *webcams* diferentes. A maior taxa de reconhecimento sobre AUs foi de 53,83%, que resultou em 28,57% de reconhecimento de emoções pelo inferidor da árvore de decisão. Já a inferência de emoções pela rede neural obteve como melhor resultado 63,33%. Estes resultados do sistema ficaram abaixo do esperado devido a problemas encontrados, principamente, nas etapas de FeD e classificação de expressões faciais. Porém, o desempenho dos classificadores isoladamente é mais promissor, pois as redes neurais obtiveram entre 83,65% e 89,87% de reconhecimento e a árvore de decisão, 71,93%. Um outro ponto que deve ser considerado é o tamanho das amostras utilizado, que por não ser amplo, permitiu o uso de outra técnica de avaliação que compense esse fato, como validação cruzada. Independente dos resultados, mesmo que o objetivo do trabalho seja inferir emoções básicas, existe uma vantagem da obtenção de códigos FACS, considerando ou não a posterior inferência de emoções. O motivo é que as emoções básicas são mais caricatas e raras no dia-a-dia, diferentemente de expressões mais discretas, como pressionar os lábios quando se está com raiva ou erguer as sobrancelhas para cumprimentar alguém, que são ações comuns e que podem ser registradas nesses códigos (TIAN; KANADE; COHN, 2005). Embora a obtenção de códigos FACS apresente-se mais útil no cotidiano, obtê-los mostrou-se ser uma tarefa não essencial para reconhecimento de emoções, já que melhores resultados foram alcançados neste trabalho desconsideranto essa codificação.

Durante a construção e avaliação deste trabalho, foram identificadas algumas possíveis alternativas e melhorias que poderão ser implementadas futuramente. De forma geral, todas as quatro etapas da aplicação (FaD, FeD, classificação de expressões faciais e inferência de emoções) têm espaço para avanços, mas, devido ao seu desempenho, somente a etapa de FaD é considerada a menos demandante. A etapa de FeD sofre com diversos obstáculos existentes na obtenção de pontos extremos sobre características faciais, como, por exemplo, problemas na intensidade de iluminação, resolução e oclusão. Um outro ponto que exige melhoria em FeD é a detecção de olhos, que pode ser alcançado utilizando projeções integrais (ZHOU; GENG, 2004). Da mesma forma, é necessário melhorar o rastreamento de pontos ou substituir o algoritmo utilizado por um que minimize o problema de abertura (BRADSKI; KAEHLER, 2008),

que ocorre atualmente na aplicação. Outras alternativas são a utilização de um modelo facial, como o AAM, de filtros de Gabor ou Gabor wavelets. A respeito de classificação de expressões faciais e emoções, pode-se avaliar outro tipo de classificador, como o SVM, redes Bayesianas ou modelos ocultos de Markov. Também pode-se considerar o uso de informações sobre cores contidas nas imagens, o que foi ignorado pela aplicação devido as bases de faces existentes apresentarem exemplos apenas em tons de cinza, além de ser mais eficiente e mais simples considerar apenas um espaço de cores, ao invés dos, normalmente, três presentes em imagens coloridas.

Apesar de existirem trabalhos que utilizem o computador para inferir emoções pela face desde a década de 1990, esta tarefa ainda é desafiadora e poucas pesquisas com este foco são encontradas em nosso país. Acredita-se que este já seja um fato que torne este trabalho por si só como uma contribuição científica em nível regional e nacional. Dentre outras contribuições inerentes ao conhecimento de emoções do usuário de computador está a possibilidade de, por exemplo, tentar acalmar uma pessoa que aparenta raiva, tornando o seu ambiente mais agradável e evitando situações conhecidamente irritantes para este usuário. Outras vantagens são a possibilidade de utilizar os dados sobre afetividade como entrada em sistemas tutores inteligentes, além do meio de detecção (*webcam*) ser não intrusivo, pois não necessita de contato físico com o usuário.

REFERÊNCIAS BIBLIOGRÁFICAS

AHLBERG, J. *CANDIDE-3 - an updated parameterized face*. Sweden: Dept. of Electrical Engineering, Linköping University, 2001. Number LiTH-ISY-R-2326.

BATISTA, L.; GOMES, H.; CARVALHO, J. Photogenic facial expression discrimination. In: *International Conference on Computer Vision Theory and Applications*. Setúbal, Portugal: Springer, 2006. p. 166–171.

BOOTH, P. A. *An Introduction To Human-Computer Interaction*. Hove, Reino Unido: Lawrence Erlbaum Associates Ltd, 1995.

BRADSKI, G. R.; KAEHLER, A. *Learning OpenCV, 1st Edition*. [S.l.]: O'Reilly Media, Inc., 2008. ISBN 9780596516130.

BRICK, T.; HUNTER, M.; COHN, J. Get the facs fast: Automated facs face analysis benefits from the addition of velocity. In: *Affective Computing and Intelligent Interaction and Workshops, 2009. ACII 2009. 3rd International Conference on*. [S.l.: s.n.], 2009. p. 1 –7.

CLORE, G.; ORTONY, A. Cognition in emotion: always, sometimes, or never? In: ____. *The Cognitive Neuroscience of Emotion*. Nova Iorque: Oxford University Press, 1999. p. 24–61.

CRAW, I.; TOCK, D.; BENNETT, A. Finding face features. In: *Proceedings of the Second European Conference on Computer Vision*. London, UK: Springer-Verlag, 1992. (ECCV '92), p. 92–96. ISBN 3-540-55426-2.

CROWLEY, J. L.; COUTAZ, J. Vision for man machine interaction. In: *Proceedings of the IFIP TC2/WG2.7 Working Conference on Engineering for Human-Computer Interaction*. London, UK, UK: Chapman & Hall, Ltd., 1995. p. 28–45.

DAI, Y.; NAKANO, Y. Face-texture model based on sgld and its application in face detection in a color scene. *Pattern Recognition*, v. 29, n. 6, p. 1007–1017, 1996.

DARRELL, T. et al. Integrated person tracking using stereo, color, and pattern detection. In: *2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 2001), with CD-ROM, 8-14 December 2001, Kauai, HI, USA*. [S.l.: s.n.], 1998. p. 601–609.

EKMAN, P. Facial expression and emotion. *American Psychologist*, v. 48, n. 4, p. 384–392, 1993.

EKMAN, P. Facial expressions. In: DALGLEISH, T.; POWER, T. (Ed.). *The Handbook of Cognition and Emotion*. Sussex, Reino Unido: John Wiley & Sons Ltd: [s.n.], 1999. p. 301–320.

EKMAN, P.; FRIESEN, W. V.; HAGER, J. C. *Facial Action Coding System*: The manual. Salt Lake City, Estados Unidos: Research Nexus division of Network Information Research Corporation, 2002a.

EKMAN, P.; FRIESEN, W. V.; HAGER, J. C. *Facial Action Coding System*: Investigator's guide. Salt Lake City, Estados Unidos: Research Nexus division of Network Information Research Corporation, 2002b.

FARKAS, L. G.; MUNRO, I. R. *Anthropometric facial proportions in medicine*. Illinois: Charles C.Thomas, 1987. ISBN 0398052611.

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

FREUND, Y.; SCHAPIRE, R. E. A decision-theoretic generalization of on-line learning and an application to boosting. In: *European Conference on Computational Learning Theory*. [S.l.: s.n.], 1995. p. 23–37.

FRIESEN, W. V.; EKMAN, P. *EMFACS-7: Emotional Facial Action Coding System*. 1983. Disponível em: <http://www.face-and-emotion.com/dataface/facs/emfacs.jsp>. Acesso em: 20 fev. 2011.

FRISCHHOLZ, R. W. *The Face Detection Homepage*. 2010. Disponível em: <http://www.facedetection.com/facedetection/techniques.htm>. Acesso em: 20 fev. 2011.

GONZALEZ, R. C.; WOODS, R. E. *Digital Image Processing*. Boston, MA, USA: Addison-Wesley Longman Publishing Co., Inc., 2001. ISBN 0201180758.

HJELMÅS, E.; LERøY, C. B.; JOHANSEN, H. *Detection and Localization of Human Faces in the ICI System*: A first attempt. [S.l.]: Gjøvik College, 1998. Number 6. Disponível em: <http://www.ansatt.hig.no/erikh/papers/hig98_6.pdf>. Acesso em: 20 fev. 2011.

HJELMÅS, E.; LOW, B. K. Face detection: A survey. *Computer Vision and Image Understanding*, v. 83, n. 3, p. 236–274, 2001.

HSU, R.-L.; ABDEL-MOTTALEB, M.; JAIN, A. K. Face detection in color images. *IEEE Trans. Pattern Anal. Mach. Intell.*, v. 24, n. 5, p. 696–706, 2002.

IOANNOU, S. et al. Robust feature detection for facial expression recognition. *European Association for Signal Processing (EURASIP) Journal on Image and Video Processing*, v. 2007, 2007.

JAQUES, P. A.; VICCARI, R. M. Estado da arte em ambientes inteligentes de aprendizagem que consideram a afetividade do aluno. *Revista Informática na Educação: Teoria & Prática*, v. 8, n. 1, p. 15–38, 2005a.

JAQUES, P. A.; VICCARI, R. M. PAT: Um agente pedagógico animado para interagir afetivamente com o aluno. *Revista Novas Tecnologias na Educação (RENOTE)*, v. 3, n. 1, 2005b.

JONES, M.; REHG, J. Statical color model with application to skin detection. In: *IEEE Conference on Computar Vision and Pattern Recognition*. [S.l.: s.n.], 1998. v. 2, p. 1274–1280.

KHANUM, A. et al. Fuzzy case-based reasoning for facial expression recognition. *Fuzzy Sets Syst.*, Elsevier North-Holland, Inc., Amsterdam, Netherlands, v. 160, n. 2, p. 231–250, 2009.

KJELDSEN, R.; KENDER, J. R. Finding skin in color images. In: *Proceedings of the 2nd International Conference on Automatic Face and Gesture Recognition (FG '96)*. Washington, DC, USA: IEEE Computer Society, 1996. (FG '96), p. 312–317.

KOBAYASHI, H.; HARA, F. The recognition of basic facial expressions by neural network. In: *Proceedings of International Joint Conference on Neural Network*. [S.l.: s.n.], 1991. p. 460–466.

KOHAVI, R.; PROVOST, F. Glossary of terms. *Machine Learning*, Kluwer Academic Publishers, Hingham, MA, USA, v. 30, p. 271–274, February 1998. ISSN 0885-6125.

LANITIS, A.; TAYLOR, C. J.; COOTES, T. F. Automatic face identification system using flexible appearance models. *Image Vision Comput.*, v. 13, n. 5, p. 393–401, 1995.

LEUNG, T. K.; BURL, M. C.; PERONA, P. Finding faces in cluttered scenes using labeled random graph matching. In: *International Conference on Computer Vision (ICCV)*. [S.l.: s.n.], 1995. p. 637–644.

LIENHART, R.; MAYDT, J. An extended set of haar-like features for rapid object detection. In: *International Conference on In Image Processing (ICIP)*. [S.l.: s.n.], 2002. p. 900–903.

LOPES, E. C.; FILHO, J. C. B. *Detecção de Faces e Características Faciais*. [S.l.]: Pontifícia Universidade Católica do Rio Grande do Sul - PUCRS, Programa de Pós-Graduação em Ciência da Computação, 2005. Number 45. Disponível em: <www3.pucrs.br/pucrs/files/uni/poa/facin/pos/relatoriostec/tr045.pdf>. Acesso em: 20 fev. 2011.

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

LYONS, M. et al. Coding facial expressions with gabor wavelets. In: *Proceedings of the 3rd. International Conference on Face & Gesture Recognition*. Washington, DC, USA: IEEE Computer Society, 1998. (FG '98), p. 200–. ISBN 0-8186-8344-9. Disponível em: <http://www.kasrl.org/jaffe.html>. Acesso em: 20 fev. 2011.

MA, E. L. H. *Avaliação de Características Haar em Um Modelo de Detecção de Face*. 70 p. Monografia (Graduação) — Departamento de Ciência da Computação, Universidade de Brasília, Brasília, 2007.

MARTIN, O. et al. Multimodal caricatural mirror. In: *ENTERFACE'05 - Proceedings of the first Summer Workshop on Multimodal Interfaces*. Bélgica: [s.n.], 2005. p. 13–20.

MCKENNA, S. J.; GONG, S.; RAJA, Y. Modelling facial colour and identity with gaussian mixtures. *Pattern Recognition*, v. 31, n. 12, p. 1883–1892, 1998.

MENSER, B.; MÜLLER, F. Face detection in color images using principal components analysis. In: *Proc. Seventh International Conference on Image Processing and its Applications IPA'99*. [S.l.: s.n.], 1999. v. 2, p. 620–624.

MIKOLAJCZYK, K.; CHOUDHURY, R.; SCHMID, C. Face detection in a video sequence - a temporal approach. In: *2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 2001), with CD-ROM, 8-14 December 2001, Kauai, HI, USA*. [S.l.]: IEEE Computer Society, 2001. p. 96–101.

M.I.T. Media Labs. *Affective Computing Projects*. 2010. Disponível em: <http://affect.media.mit.edu/projects.php>. Acesso em: 20 fev. 2011.

MPEG Moving Picture Experts Group. *MPEG-4*: Overview v.21, ISO/IEC JTC1/SC29/WG11 N4668. 2002. Disponível em: <http://www.chiariglione.org/mpeg/standards/mpeg-4/mpeg-4.htm>. Acesso em: 20 fev. 2011.

MPI. *Max Planck Institute for Biological Cybernetics*: The Face Video Database of Max Planck Institute for Biological Cybernetics. 2011. Disponível em: <http://vdb.kyb.tuebingen.mpg.de/>. Acesso em: 20 fev. 2011.

NISSEN, S. *Implementation of a Fast Artificial Neural Network Library (fann)*. [S.l.], 2003. Disponível em: <http://fann.sf.net>. Acesso em: 20 fev. 2011.

OLIVEIRA, E. de. *Captura de expressões faciais para identificação de emoções básicas em humanos*. 88 p. Monografia (Graduação) — Universidade do Vale do Rio dos Sinos, São Leopoldo, 2008.

OLIVEIRA, E. de; JAQUES, P. A. Inferindo as emoções do usuário pela face através de um sistema psicológico de codificação facial. In: *Proceedings of the VIII Brazilian Symposium on Human Factors in Computing Systems*. Porto Alegre, Brazil: Sociedade Brasileira de Computação, 2008. (IHC '08), p. 156–165. ISBN 978-85-7669-203-4. Disponível em: <http://portal.acm.org/citation.cfm?id=1497470.1497488>.

ORTONY, A.; CLORE, G. L.; COLLINS, A. *The Cognitive Structure of Emotions*. [S.l.]: Cambridge University Press, 1988. ISBN 0521353645.

OSUNA, E.; FREUND, R.; GIROSI, F. Training support vector machines: an application to face detection. In: *1997 Conference on Computer Vision and Pattern Recognition (CVPR '97), June 17-19, 1997, San Juan, Puerto Rico*. [S.l.]: IEEE Computer Society, 1997. p. 130–136.

PANDZIC, I. S.; FORCHHEIMER, R. (Ed.). *MPEG-4 Facial Animation: The Standard, Implementation and Applications*. New York, NY, USA: John Wiley & Sons, Inc., 2003. ISBN 0470854626.

PANTIC, M.; BARTLETT, M. S. Face recognition. In: ____. Vienna, Austria: I-Tech Education and Publishing, 2007. cap. 20, p. 377–416.

PANTIC, M.; ROTHKRANTZ, L. J. M. Facial action recognition for facial expression analysis from static face images. *IEEE Transactions on Systems, Man, and Cybernetics, Part B*, v. 34, n. 3, p. 1449–1461, 2004.

PETRIDIS, S. et al. Static vs. dynamic modeling of human nonverbal behavior from multiple cues and modalities. In: *ICMI-MLMI '09: Proceedings of the 2009 international conference on Multimodal interfaces*. New York, NY, USA: ACM, 2009. p. 23–30. ISBN 978-1-60558-772-1.

PICARD, R. W. *Affective Computing*. [S.l.]: M.I.T Media Laboratory Perceptual Computing Section Technical Report, 1995. TR 321.

PICARD, R. W. *Affective Computing*. Cambridge, EUA: MIT Press, 1997.

QUINLAN, J. R. *C4.5*: Programs for machine learning. [S.l.]: Morgan Kaufmann, 1993. ISBN 1-55860-238-0.

RAJAGOPALAN, A. N. et al. Finding faces in photographs. In: *Proceedings of the Sixth International Conference on Computer Vision*. Washington, DC, USA: IEEE Computer Society, 1998. (ICCV '98), p. 640–645.

RAOUZAIOU, A. et al. Parameterized facial expression synthesis based on mpeg-4. *EURASIP J. Appl. Signal Process.*, Hindawi Publishing Corp., New York, NY, United States, v. 2002, n. 1, p. 1021–1038, 2002. ISSN 1110-8657.

ROWLEY, H.; BALUJA, S.; KANADE, T. *Rotation Invariant Neural Network-Based Face Detection*. Pittsburgh, PA, December 1997.

ROWLEY, H.; BALUJA, S.; KANADE, T. Rotation invariant neural network-based face detection. In: *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*. [S.l.: s.n.], 1998a.

ROWLEY, H. A.; BALUJA, S.; KANADE, T. Neural network-based face detection. *IEEE Trans. Pattern Anal. Mach. Intell.*, v. 20, n. 1, p. 23–38, 1998b.

RULEQUESTRESEARCH. *Data mining tools See5 and C5.0*. 2001. Disponível em: <http://www.rulequest.com/see5-info.html>. Acesso em: 20 fev. 2011.

SCHAPIRE, R. E. The strength of weak learnability. *Machine Learning*, v. 5, p. 197–227, 1990.

SCHERER, K. Skin colour detection under changing lighting conditions. In: BOROD, J. (Ed.). *The neuropsychology of emotion*. Nova Iorque: Oxford University Press: [s.n.], 2000a. p. 137–162.

SCHERER, K. R. Psychological models of emotion. In: ____. *The neuropsychology of emotion*. J. borod (ed.). Oxford/New York: Oxford University Press, 2000b. p. 137–162.

SCHIANO, D. J. et al. Face to interface: facial affect in (hu)man and machine. In: *Proceedings of the SIGCHI conference on Human factors in computing systems*. New York, NY, USA: ACM, 2000. (CHI '00), p. 193–200. ISBN 1-58113-216-6.

SCHMIDT, K. L.; COHN, J. F. Human facial expressions as adaptations: Evolutionary questions in facial expression. *American Journal of Physical Anthropology (Yearbook of Physical Anthropology)*, v. 44, n. S33, p. 3–24, 2001.

SCHNEIDERMAN, H.; KANADE, T. Probabilistic modeling of local appearance and spatial relationships for object recognition. In: *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. Washington, DC, USA: IEEE Computer Society, 1998. (CVPR '98), p. 45–51. ISBN 0-8186-8497-6.

SEBE, N. et al. Multimodal approaches for emotion recognition: a survey. In: SANTINI, S.; SCHETTINI, R.; GEVERS, T. (Ed.). *SPIE'05: Internet Imaging VI*. San Jose, EUA: [s.n.], 2005. p. 56–67.

SENIOR, A. et al. Face detection in color images. *IEEE Trans. Pattern Anal. Mach. Intell.*, IEEE Computer Society, Washington, DC, USA, v. 24, n. 5, p. 696–706, 2002. ISSN 0162-8828.

SHAPIRO, L. G.; STOCKMAN, G. *Computer Vision*. Upper Saddle River, NJ, USA: Prentice Hall PTR, 2001. ISBN 0130307963.

SOHAIL, A. S. M.; BHATTACHARYA, P. Detection of facial feature points using anthropometric face model. In: *In IEEE/ACM Proceedings of International Conference on Signal-Image Technology & Internet-Based Systems (SITIS 2006)*. Hammamet, Tunisia: [s.n.], 2006.

SOHAIL, A. S. M.; BHATTACHARYA, P. Classifying facial expressions using point-based analytic face model and support vector machines. In: *Proceedings of the IEEE International Conference on Systems, Man and Cybernetics, Montréal, Canada, 7-10 October 2007*. [S.l.: s.n.], 2007. p. 1008–1013.

SOUZA, L. V. de. *Programação genética e combinação de preditores para previsão de séries temporais*. Tese (Doutorado) — Universidade Federal do Paraná, 2006.

STÖRRING, M. *Computer Vision and Human Skin Colour*. Tese (Doutorado) — Faculty of Engineering and Science, Aalborg University, Niels Jernes Vej 14, 9220 Aalborg, Denmark, 2004.

STÖRRING, M.; ANDERSEN, H. J.; GRANUM, E. Skin colour detection under changing lighting conditions. In: ARAUJO, H.; DIAS, J. (Ed.). *7th International Symposium on Intelligent Robotic Systems*. Coimbra, Portugal: [s.n.], 1999. p. 187–195.

SUNG, K. K.; POGGIO, T. Example-based learning for view-based human face detection. *IEEE Trans. Pattern Anal. Mach. Intell.*, v. 20, n. 1, p. 39–51, 1998.

TEKALP, J. O. A. M. Face and 2-d mesh animation in mpeg-4. *Signal Processing: Image Communication*, v. 15, p. 387–421(35), 2000.

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

TIAN, Y.-L.; KANADE, T.; COHN, J. F. Facial expression analysis. In: S.LI, A. (Ed.). *Handbook of Face Recognition*. New York: Springer, 2005.

TOMAZ, F. A. G. *Face detection and recognition*. 2010. Disponível em: <http://w3.ualg.pt/~ftomaz/fr/fr.php>. Acesso em: 20 fev. 2011.

TONG, Y.; LIAO, W.; JI, Q. Affective information processing. In: ____. London: Springer, 2008. cap. 10, p. 159–180.

TURK, M.; PENTLAND, A. Eigenfaces for recognition. *J. Cognitive Neuroscience*, v. 3, n. 1, p. 71–86, 1991.

VALSTAR, M.; PANTIC, M. Fully automatic facial action unit detection and temporal analysis. In: *CVPRW '06: Proceedings of the 2006 Conference on Computer Vision and Pattern Recognition Workshop*. Washington, DC, USA: IEEE Computer Society, 2006. p. 149. ISBN 0-7695-2646-2.

VIOLA, P. A.; JONES, M. J. Rapid object detection using a boosted cascade of simple features. In: *2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 2001), with CD-ROM, 8-14 December 2001, Kauai, HI, USA*. Los Alamitos, CA, USA: IEEE Computer Society, 2001. p. 511–518. ISBN 0-7695-1272-0.

WILLOW GARAGE. *OpenCV*: Open Source Computer Vision Library. 2010. Disponível em: <http://opencv.willowgarage.com>. Acesso em: 20 fev. 2011.

YANG, G.; HUANG, T. S. Human face detection in a complex background. *Pattern Recognition*, v. 27, n. 1, p. 53–63, 1994.

YANG, J.; WAIBEL, A. A real-time face tracker. In: *Workshop Applications of Computer Vision*. [S.l.: s.n.], 1996. p. 142–147.

YANG, M.-H. *Recent Advances in Face Detection*. [S.l.]: IEEE ICPR 2004 Tutorial, Cambridge, United Kingdom, 2004. Disponível em: <http://vision.ai.uiuc.edu/mhyang/papers/icpr04_tutorial.pdf>. Acesso em: 20 fev. 2011.

YANG, M.-H.; AHUJA, N.; KRIEGMAN, D. J. Face detection using mixtures of linear subspaces. In: *4th IEEE International Conference on Automatic Face and Gesture Recognition (FG 2000), 26-30 March 2000, Grenoble, France*. Washington, DC, USA: IEEE Computer Society, 2000. p. 70–76.

YANG, M.-H.; KRIEGMAN, D. J.; AHUJA, N. Detecting Faces in Images: A Survey. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, v. 24, n. 1, p. 34–58, 2002.

YANG, M.-H.; ROTH, D.; AHUJA, N. A snow-based face detector. In: *NIPS*. [S.l.: s.n.], 1999. p. 862–868.

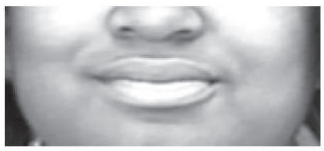
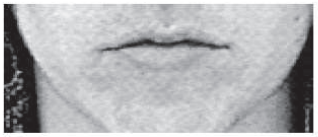
YILMAZ, A.; JAVED, O.; SHAH, M. Object tracking: A survey. *ACM Comput. Surv.*, v. 38, n. 4, 2006.

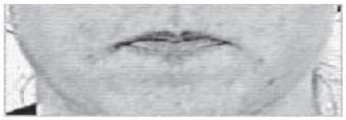
YOW, K. C.; CIPOLLA, R. Feature-based human face detection. *Image Vision Comput.*, v. 15, n. 9, p. 713–735, 1997.

ZHOU, Z.-H.; GENG, X. Projection functions for eye detection. *Pattern Recognition*, v. 37, n. 5, p. 1049–1056, 2004.

APÊNDICE A – TABELA DE AUS

AU	Descrição	Exemplo	Categoria
1	*Inner Brow Raiser*		Relacionado com músculos faciais específicos
2	*Outer Brow Raiser*		
4	*Brow Lowerer*		
5	*Upper Lid Raiser*		
6	*Cheek Raiser*		
7	*Lid Tightener*		
9	*Nose Wrinkler*		
10	*Upper Lip Raiser*		

AU	Descrição	Exemplo	Categoria
11	*Nasolabial Furrow Deepener*		Relacionado com músculos faciais específicos
12	*Lip Corner Puller*		
13	*Cheek Puffer*		
14	*Dimpler Buccinnator*		
15	*Lip Corner Depressor*		
16	*Lower Lip Depressor*		
17	*Chin Raiser*		
18	*Lip Puckerer*		
20	*Lip Stretcher*		
22	*Lip Funneler*		
23	*Lip Tightner*		

AU	Descrição	Exemplo	Categoria
24	*Lip Pressor*		Relacionado com músculos faciais específicos
25	*Lips Part*		
26	*Jaw Drop*		
27	*Mouth Stretch*		
28	*Lip Suck*		
41	*Lid Droop*		
42	*Slit*		
43	*Eyes Closed*		
44	*Squint*		
45	*Blink*		
46	*Wink*		
8	*Lips Toward Each Other*		Ações variadas - Sem base muscular específica
19	*Tongue Out*		
21	*Neck Tightener*		

AU	Descrição	Exemplo	Categoria
29	*Jaw Thrust*		Ações variadas - Sem base muscular específica
30	*Jaw Sideways*		
31	*Jaw Clencher*		
32	*Lip Bite*		
33	*Cheek Blow*		
34	*Cheek Puff*		
35	*Cheek Suck*		
36	*Tongue Bulge*		
37	*Lip Wipe*		
38	*Nostril Dilator*		
39	*Nostril Compressor*		
51	*Head turn left*		Posições de cabeça
52	*Head turn right*		
53	*Head up*		
54	*Head down*		
55	*Head tilt left*		
56	*Head tilt right*		
57	*Head forward*		
58	*Head back*		

AU	Descrição	Exemplo	Categoria
61	*Eyes turn left*		Posições de olhos
62	*Eyes turn right*		
63	*Eyes up*		
64	*Eyes down*		
65	*Walleye*		
66	*Cross-eye*		
68	*Eye Movement*		Movimento de olhos
69	*Eye Movement*		
70	*Brows and forehead not visible*		Visibilidade
71	*Eyes not visible*		
72	*Lower face not visible*		
73	*Entire face not visible*		
74	*Unscorable*		
40	*Sniff*		*Gross behaviors*
50	*Speech*		
80	*Swallow*		
81	*Chewing*		
82	*Shoulder shrug*		
84	*Head shake back and forth*		
85	*Head nod up and down*		
91	*Flash*		
92	*Partial flash*		
83	*Head Movement*		Movimento de cabeça

Baseado em: http://www-2.cs.cmu.edu/afs/cs/project/face/www/facs.htm