

UNIVERSIDADE FEDERAL DO MARANHÃO
CENTRO DE CIÊNCIAS EXATAS E TECNOLOGIA
CURSO DE CIÊNCIA DA COMPUTAÇÃO

Ruy Guilherme Silva Gomes de Oliveira

*Reconhecimento de expressões faciais utilizando árvore de
decisão CART*

São Luís
2015

Ruy Guilherme Silva Gomes de Oliveira

*Reconhecimento de expressões faciais utilizando árvore de
decisão CART*

Monografia apresentada ao Curso de Ciência da
Computação da UFMA, como requisito parcial
para a obtenção do grau de BACHAREL em
Ciência da Computação.

Orientador: Aristófanês Corrêa Silva

Prof. Doutor em Informática

São Luís

2015

Oliveira, Ruy Guilherme Silva Gomes de.

Reconhecimento de expressões faciais utilizando árvore de decisão CART/ Ruy Guilherme Silva Gomes de Oliveira. – São Luís, 2015.

49 f.

Impresso por computador (fotocópia).

Orientador: Aristófanês Corrêa Silva.

Monografia (Graduação) – Universidade Federal do Maranhão, Curso de Ciência da Computação, 2015.

1. Processamento de Imagens - Expressões Faciais. 2. Reconhecimento de padrões.3. Árvore de decisão. I. Título.

CDU 004.383.5

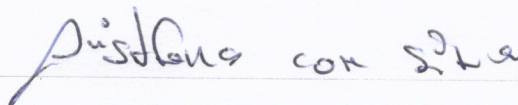
Ruy Guilherme Silva Gomes de Oliveira

*Reconhecimento de expressões faciais utilizando árvore de
decisão CART*

Monografia apresentada ao Curso de Ciência da
Computação da UFMA, como requisito parcial
para a obtenção do grau de BACHAREL em
Ciência da Computação.

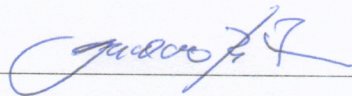
Aprovado em 19 de Janeiro de 2015

BANCA EXAMINADORA



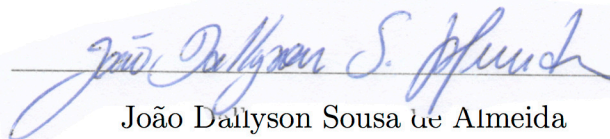
Aristófanés Corrêa Silva

Prof. Doutor em Informática



Geraldo Braz Júnior

Prof. Doutor em Engenharia da Eletricidade



João Dallyson Sousa de Almeida

Prof. Doutor em Engenharia da Eletricidade

À minha família.

Agradecimentos

Ao meus pais, por me guiarem durante toda a minha educação e por suas palavras de sabedoria que me ajudaram a escolher sempre o melhor caminho.

À minha avó por suas orientações e por compor juntamente com meus pais a base da minha formação pessoal e profissional.

À minha namorada Valéria, por me acompanhar durante todos estes anos, sempre me apoiando a alcançar meus objetivos e a superar os meus limites.

Aos meus amigos, pelos momentos de descontração e pelas sugestões fornecidas sobre o trabalho aqui desenvolvido.

Ao meu orientador, Aristófares Corrêa Silva, por sua dedicação e por todos os ensinamentos passados.

E a todos que contribuíram mesmo que indiretamente para o desenvolvimento deste trabalho.

Resumo

O reconhecimento de expressões faciais compõe a base da computação afetiva. Ela possibilita a criação de aplicações tanto no domínio acadêmico como no domínio empresarial. O desafio de desenvolver tais ferramentas, consiste em como realizar o reconhecimento das expressões faciais a partir da análise de imagens digitais. Diversas metodologias de reconhecimento de expressões faciais foram desenvolvidas ao longo dos anos, cada uma com suas restrições e qualidades. Por isso, existe uma demanda contínua por novas metodologias, que sejam capazes de atender aos requisitos de sistemas não abrangidos pelas já desenvolvidas. Neste trabalho propõe-se uma metodologia capaz de reconhecer a expressão neutra e as seis expressões faciais universais, sendo elas: alegria, tristeza, raiva, medo, desgosto e surpresa. A metodologia proposta é composta por série de etapas, que exercem desde a extração das características da imagem até a classificação da expressão facial. A extração das características é realizada utilizando técnicas de processamento de imagens capazes de segmentar e extrair informações das feições do rosto. E a classificação é realizada utilizando-se a árvore de decisão CART, uma técnica de reconhecimento de padrões, que induz a expressão facial contida na imagem a partir da análise das características extraídas. A metodologia resultante alcançou uma acurácia de 54% nos testes realizados, um resultado promissor em se tratando de um problema que envolve sete classes diferentes.

Palavras-chave: Expressões Faciais, Processamento de Imagens, Reconhecimento de Padrões, Árvore de Decisão.

Abstract

Facial expressions recognition is the base of affective computing. It makes possible the development of a variety of applications. Emotionent is an example of these applications, it is a system capable of analyse customers reaction when seeing product prices. The challenge in these tools is recognizing facial expressions on digital images. There is a variety of methodologies developed to recognize facial expressions. Facial expression recognition is performed based on image processing and pattern recognition techniques. On pattern recognition, decision trees proved to be a promising approach on facial expression recognition, being an attractive technique to be applied on the development of a new methodology. Therefore, the main objective of this work is to develop a new facial expression recognition methodology based on decision trees capable of recognizing all six universal facial expressions and the neutral expression. The proposed methodology is composed of a series of stages that performs both facial characteristics extraction and facial expression classification. The resultant methodology reached 54% accuracy on tests performed with an image database. This is a promising result considering that the problem approached has seven classes. But it was verified that the methodology had difficulty to extract mouth characteristis. And based on its performance evaluation, it is possible to improve the results with the inclusion of more face characteristics.

Keywords: Facial Expression, Image Processing, Pattern Recognition, Decision Tree.

“Faça ou não faça. Tentativa não há.”

(Yoda)

Lista de Figuras

2.1	Exemplo de aplicação do <i>Gaussian Smoothing</i> em uma imagem da face . . .	16
2.2	Exemplo de aplicação do <i>Canny Edge Detector</i> para a detectar as arestas da boca	16
2.3	Aplicação do filtro de abertura em uma imagem das arestas da boca previamente segmentados com o <i>Canny Edge Detector</i>	18
2.4	Segmentação do olho por limiarização	19
2.5	(a) Características pseudo-haar de Vértice; (b) Características pseudo-haar de Linha; (c) Características pseudo-haar de quatro retângulos; (INTEL CORPORATION, 2001)	20
2.6	Detecção da face com <i>Haar Cascade</i>	20
2.7	Imagem resultante após a execução do landmark	21
2.8	Exemplo de uma árvore de decisão de reconhecimento de expressões faciais	22
2.9	Conversão de matriz de confusão de n classes para 2 classes	27
3.1	Etapas da Metodologia	28
3.2	Proporções da face humana com base na distância entre os olhos (VUKADINOVIC; PANTIC, 2005)	30
3.3	Exemplo de segmentação dos olhos	31
3.4	Exemplo de segmentação da sobrancelha	32
3.5	Exemplo de segmentação da boca	33
3.6	<i>Facial Characteristic Points</i> (KING; HOU, 1996)	34
3.7	<i>Facial Characteristic Points</i> utilizados na metodologia representados pelos pontos azuis	35
3.8	Cálculo do centroide da sobrancelha	35

3.9	Medidas do rosto, representadas pelas setas azuis, utilizadas para calcular os FAPS da metodologia	36
4.1	<i>Extração de feições</i>	40
4.2	Matriz de Confusão resultante da classificação	41

Sumário

1	Introdução	11
1.1	Objetivos	12
1.1.1	Objetivos Específicos	12
1.2	Trabalhos Relacionados	12
2.1	Técnicas Processamento de Imagens	15
2.1.1	<i>Gaussian Smoothing</i>	15
2.1.2	<i>Canny Edge Detector</i>	16
2.1.3	Operação Morfológica de Abertura	17
2.1.4	Segmentação por Limiarização	18
2.2	<i>Haar Cascade</i>	19
2.3	Flandmark	20
2.4	Classificação de Objetos com Árvores de Decisão	21
2.5	Avaliação de Modelos de Classificação	24
2.5.1	Matriz de Confusão	24
3	Metodologia Proposta	28
3.1	Extração de Feições	28
3.2	Computação dos <i>Facial Characteristic Points</i>	34
3.3	Computação dos <i>Facial Animation Parameters</i>	36

3.4	Classificação das Expressões Faciais	37
4	Resultados e Discussão	39
5	Considerações Finais	43
	Referências	44

Lista de Abreviaturas e Siglas

REF	Reconhecimento de Expressões Faciais
IHC	Interface Humano-Computador
FCP	<i>Facial Characteristic Points</i>
FAP	<i>Facial Action Parameters</i>
DPM	<i>Deformable Part Models</i>
ROI	<i>Region Of Interest</i>

1 Introdução

O Reconhecimento de Expressões Faciais (REF) tem sido uma área de pesquisa bastante atrativa durante as últimas décadas. As expressões faciais têm uma forte relação com o processamento das emoções humanas. Pesquisas recentes apontam que elas carregam uma quantidade considerável de informações relacionadas ao comportamento humano e desempenham um papel ativo na comunicação interpessoal (PERVEEN; GUPTA; VERMA, 2012). Tais aspectos das expressões atraem a atenção dos pesquisadores e até mesmo de grandes empresas, estimulando o desenvolvimento de trabalhos que envolvem o reconhecimento e a análise de expressões faciais.

Hoje, o REF forma a base da computação afetiva e suas aplicações são exploradas por diversas áreas de pesquisa, como a Interface Humano-Computador (IHC) e a Ergonomia. O REF tem diversas aplicações sob o ponto de vista da IHC, como a criação de uma maneira natural de comunicação entre homem e máquina e a extração de dados relevantes de usabilidade de um sistema que podem ser usados para a análise da experiência do usuário.

Até os tempos atuais, diferentes técnicas de REF já foram propostas, cada uma com aspectos diferentes quanto à precisão e à robustez. A precisão é medida pelo índice de acerto das técnicas ao serem aplicadas sobre um mesmo conjunto de pessoas. E a robustez das metodologias está relacionada com a capacidade de fornecer bons resultados quando sujeito a condições como variação de iluminação e diferenças de medidas antropométricas entre etnias.

Sabe-se hoje que determinadas expressões faciais são comuns à todas as culturas, estas são denominadas expressões faciais universais. As expressões faciais universais, que foram encontradas por Paul Ekman Ekman (1989), são alegria, tristeza, raiva, medo, desgosto e surpresa. A fim de tornar os trabalhos que envolvem REF mais abrangentes e menos específicos, as técnicas desenvolvidas na área objetivam reconhecer somente as seis expressões faciais universais.

Assim, a proposta de uma nova metodologia para o reconhecimento de expressões faciais utilizando árvores de decisão mostra-se relevante, pois proporcionará

uma base para o desenvolvimento de soluções que envolvam a análise de expressões faciais. A nova metodologia tem aspectos diferentes quanto ao desempenho, robustez e precisão em comparação às metodologias já propostas, demonstrando que o reconhecimento de expressões faciais é uma área de pesquisa que, mesmo antiga, ainda tem muito a ser explorado.

1.1 Objetivos

O objetivo deste trabalho é desenvolver uma metodologia que aplique técnicas de processamento de imagem e reconhecimento de padrão para o reconhecimento da expressão facial neutra e das seis expressões faciais universais, que são alegria, tristeza, raiva, medo, desgosto e surpresa.

1.1.1 Objetivos Específicos

Os objetivos específicos são:

- Estudar e implementar técnicas para realizar a segmentação de feições e computação de características faciais;
- Estudar e implementar técnicas para realizar o reconhecimento das expressões faciais universais a partir de características faciais utilizando a árvore de decisão CART;
- Avaliar a metodologia de reconhecimento de expressões faciais proposta;

1.2 Trabalhos Relacionados

Desde a década de 90, diversos pesquisadores têm realizado pesquisas em busca de soluções para o reconhecimento das expressões faciais universais. Hiroshi Kobayashi e Fumio Hara Kobayashi e Hara (1992) propuseram uma técnica de reconhecimento de expressões faciais e quantificação da força da expressão facial reconhecida utilizando redes neurais, porém com extração de características faciais manual.

Barlett Donato et al. (1999) explorou em suas pesquisas técnicas que permitissem o reconhecimento de expressões faciais em sequências de imagens. As técnicas

propostas em seus trabalhos incluem análise de fluxo ótico, análise espacial holística, análise local de características, e métodos baseados na saídas de filtros locais.

Tai e Chang Tai e Chung (2007) propuseram uma técnica de reconhecimento automático de expressões também utilizando redes neurais. Neste trabalho, eles implementaram uma técnica chamada canthi-edge para a extração automática dos pontos de características faciais. Eles também optaram por reduzir a quantidade de características faciais utilizadas a fim de diminuir o processamento, mas os resultados obtidos apresentaram uma baixa acurácia.

Juanjuan Juanjuan et al. (2010) criou um algoritmo de reconhecimento de expressões faciais derivado do PCA (*Principal Component Analysis*), denominado *PCA Reconstruction*. A sua proposta se baseava no fato que o PCA, ao extrair características das expressões faciais, trazia junto características referentes à fisionomia da pessoa. Assim, ele propôs uma variação do PCA que se concentra nas regiões dos olhos e da boca, pois têm uma forte relação com às expressões faciais.

Perveen Perveen, Gupta e Verma (2012) propôs uma metodologia de classificação de expressões faciais em imagens estáticas baseada em Gini Index, que é um tipo de árvore de decisão. A metodologia utiliza características da face denominadas *Facial Action Parameters* (FAP) calculadas a partir dos *Facial Characteristic Points* (FCP). Os FCPs nesta metodologia são extraídos da imagem utilizando *template matching*.

Gupta e Perveen Gupta, Verma e Perveen (2012) apresentaram outra metodologia para reconhecer expressões faciais porém utilizando árvores de decisão CART. Ele também utilizou as FAPs para realizar o reconhecimento e *template matching* para extrair as FCPs.

Assim, o resultado destes trabalhos é incentivador pois mesmo com o sucesso de algumas abordagens, todos encontraram resultados com seus pontos positivos e negativos referentes ao desempenho e precisão. Isto abre um grande espaço para novas propostas e experimentos na área.

1.3 Organização do Trabalho

Este capítulo apresentou a introdução e os objetivos deste trabalho e listou os principais trabalhos relacionados.

O Capítulo 2 traz a base teórica sobre as técnicas utilizadas no desenvolvimento da metodologia. Para o entendimento da metodologia proposta, o leitor deve ter um conhecimento básico sobre o Processamento de Imagens. A Seção 2.1 apresenta diversas técnicas de processamento de imagens. A Seção 2.2 apresenta o algoritmo proposto por Viola e Jones (VIOLA; JONES, 2001), que foi utilizado na metodologia para detecção de faces. A biblioteca *flandmark* responsável por detectar algumas características da face foi explicada na Seção 2.3. A classificação com árvores de decisão, que na metodologia é responsável por classificar as expressões faciais, é explicada na Seção 2.4. A Seção 2.5 aborda o método de validação cruzada e as métricas de avaliação que são aplicados na metodologia. A Seção 2.5.1 descreve o funcionamento da matriz de confusão e as métricas que podem ser extraídas a partir da mesma.

O Capítulo 3 apresenta a metodologia desenvolvida a partir da aplicação das técnicas descritas no Capítulo 2. A metodologia é composta por um conjunto de etapas, cada uma descrita em uma das seções deste capítulo.

A Seção 3.1 descreve a etapa na qual é realizada a extração das feições a partir das técnicas de processamento de imagens apresentadas.

A Seção 3.2 mostra os cálculos realizados para encontrar as coordenadas dos FCPs. Na Seção 3.3 são apresentados os FAPs selecionados e as fórmulas utilizadas para calculá-los a partir dos FCPs.

A Seção 3.4 mostra o procedimento realizado para classificar a expressão facial contida na imagem a partir dos valores dos FAPs extraídos da mesma. A classificação é a etapa final da metodologia.

Os resultados da metodologia e a avaliação do seu desempenho são descritos no Capítulo 4. Nele, são apresentadas as métricas extraídas da implementação da metodologia.

No Capítulo 5 são apresentadas as considerações finais acerca da metodologia proposta.

2 Fundamentação Teórica

Neste capítulo são explicados conceitos e técnicas que estão relacionados com a metodologia proposta. O capítulo é dividido nas seções Técnicas de Processamento de Imagens, *Haar Cascade*, Flandmark e Classificação de Objetos com Árvores de Decisão. Esta organização do capítulo visa facilitar a compreensão dos assuntos explorados.

2.1 Técnicas Processamento de Imagens

As técnicas de processamento de imagens consistem na execução de operações matemáticas sobre os dados de uma imagem digital de modo que a imagem resultante seja mais adequada que a imagem original para uma aplicação específica (MENESES; ALMEIDA, 2012).

Existe uma grande quantidade de técnicas disponíveis que possibilitam a criação de soluções para problemas de diversos domínios.

2.1.1 *Gaussian Smoothing*

Gaussian Smoothing ou filtro da Gaussiana, é uma técnica comumente utilizada para diminuição de ruídos em imagens (JAIN; KASTURI; SCHUNCK, 1995). Ele é um filtro de suavização cujo *kernel*¹ é definido a partir da equação da gaussiana, apresentada na Equação 2.1.

$$g(x) = e^{-\frac{x^2}{2\sigma}} \quad (2.1)$$

A gaussiana é aplicada neste trabalho durante o pré-processamento das imagens, para diminuir os ruídos e facilitar a detecção das feições. A Figura 2.1 mostra um exemplo de aplicação da gaussiana da forma como fora realizado neste trabalho.

¹Kernel, também conhecido como matriz de convolução ou máscara, é uma matriz de valores utilizada em diversas técnicas de processamento de imagens.

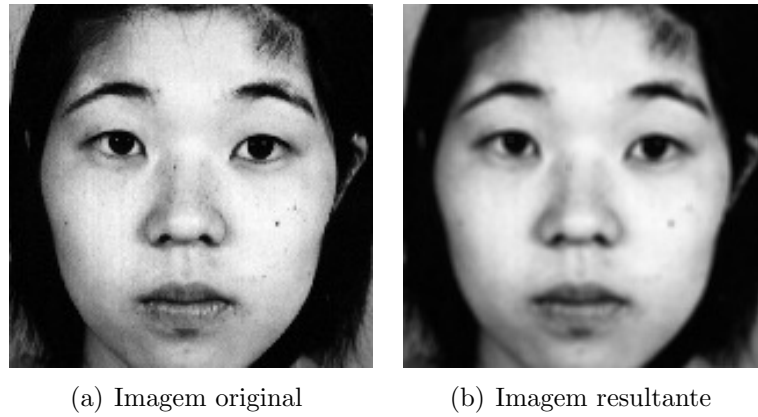


Figura 2.1: Exemplo de aplicação do *Gaussian Smoothing* em uma imagem da face

2.1.2 *Canny Edge Detector*

O *Canny Edge Detector*, também conhecido como filtro de Canny, é um algoritmo que permite a segmentação de arestas em imagens. Este filtro propõe-se a satisfazer três requisitos, baixa taxa de erro, boa localização e resposta mínima. A taxa de erro é referente à frequência com que arestas são detectadas. A localização é referente à distância entre os pixels² da aresta detectados. E a resposta é referente á capacidade de garantir somente uma resposta do detector por aresta (JAIN; KASTURI; SCHUNCK, 1995).

O algoritmo do filtro de Canny opera em quatro etapas: redução de ruídos com filtro da Gaussiana; computação do gradiente de magnitude e de orientação aplicando aproximações de diferenças-finitas às derivadas parciais; supressão de não-máximos do gradiente de magnitude; execução do algoritmo de limiar duplo para detectar e conectar arestas (JAIN; KASTURI; SCHUNCK, 1995).

O filtro de Canny é aplicado neste trabalho na etapa de segmentação das feições para a detecção do contorno da boca. A Figura 2.2 mostra um exemplo desta aplicação.

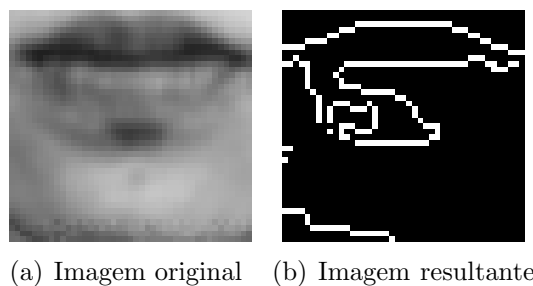


Figura 2.2: Exemplo de aplicação do *Canny Edge Detector* para a detectar as arestas da boca

²Pixel é o menor elemento que compõe uma imagem digital.

2.1.3 Operação Morfológica de Abertura

As operações morfológicas são transformações realizadas em imagens que têm a capacidade de manipular a sua estrutura geométrica. Elas podem ser aplicadas com diversos objetivos distintos como realce, filtragem, segmentação, detecção de bordas, esqueletização, afinamento e outros (FILHO; NETO, 1999).

O elemento responsável pela transformação da imagem é conhecido como elemento estruturante. Em processamento de imagens ele é comumente representado como uma matriz de pixels, e é aplicado sobre a imagem original de acordo com a equação de transformação definida por cada operação.

A dilatação e erosão são operações morfológicas responsáveis, respectivamente por expandir e encolher os elementos de uma imagem. Considerando A a matriz de pixels de uma imagem e B a matriz de pixels do elemento estruturante. A Equação 2.2 apresenta a função de transformação da dilatação e a Equação 2.3 apresenta a função de transformação da erosão.

$$A \oplus B = \{x | [(B_x) \cap A] \subseteq A\} \quad (2.2)$$

$$A \ominus B = \{x | (B_x) \subseteq A\} \quad (2.3)$$

Uma operação morfológica frequentemente utilizada em processamento de imagens é a de abertura. A operação de abertura tem a capacidade de suavizar contornos, quebrar istmos estreitos e eliminar pequenas ilhas e picos em imagens.

A abertura de um conjunto A por um elemento estruturante B denotada por $A \circ B$ definida como:

$$A \circ B = (A \ominus B) \oplus B \quad (2.4)$$

A operação de abertura é aplicada na etapa de segmentação das feições para excluir pequenas falhas no contorno de objetos, incorporando-as. A Figura 2.3 mostra um exemplo da aplicação da operação de abertura em uma imagem da boca, como é utilizado neste trabalho.

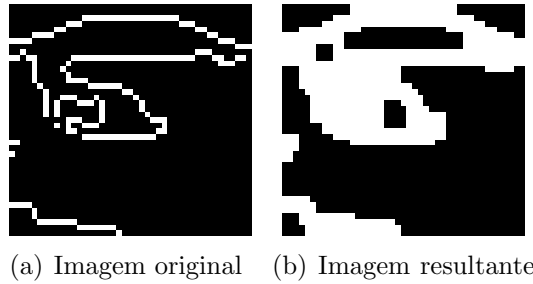


Figura 2.3: Aplicação do filtro de abertura em uma imagem das arestas da boca previamente segmentados com o *Canny Edge Detector*

2.1.4 Segmentação por Limiarização

A segmentação, em processamento de imagens, consiste em dividir uma imagem em diversos conjuntos de pixels que contenham os objetos de interesse nela presentes (FILHO; NETO, 1999).

A limiarização é um método de segmentação baseado em cor comumente utilizado em processamento de imagens. Ela consiste em dividir uma imagem representada em níveis de cinza em duas regiões diferentes, uma com o fundo e outra com os objetos de interesse. Na limiarização, determina-se um valor de nível de cinza, chamado de limiar, e todos os valores de pixels menores ou iguais a esse valor são mapeados na imagem resultante em 0 e os demais são mapeados em um valor diferente de 0, que normalmente é 255 (FILHO; NETO, 1999). A operação de limiarização pode ser representada pela Equação 2.5, na qual x e y são coordenadas dos pixels de uma imagem, $g(x, y)$ é o valor do pixel da imagem resultante e $f(x, y)$ o valor do pixel da imagem de origem.

$$g(x, y) = \begin{cases} 1 & \text{se } f(x, y) > \textit{limiar} \\ 0 & \text{se } f(x, y) < \textit{limiar} \end{cases} \quad (2.5)$$

A segmentação baseada em cor é uma boa alternativa quando se deseja segmentar objetos que variam muito quanto à sua forma geométrica, mas pouco quanto à sua textura. Além disto, a cor é uma característica comum à todas as imagens e é considerada um poderoso descritor das propriedades de um objeto, fatores estes que torna possível a aplicação desta técnica nos mais diversos domínios.

O desafio deste método consiste em encontrar o limiar que melhor segmente o objeto de interesse e que seja adaptável às mais variadas condições de iluminação e para solucionar tais problemas existe uma variedade de técnicas.

Existem técnicas capazes de encontrar o melhor limiar para uma determinada imagem, conhecidas como limiarizações ótimas. Uma técnica de limiarização ótima se baseia nas propriedades estatísticas, como a média dos tons de cinza, da imagem para calcular o seu limiar (FILHO; NETO, 1999).

A limiarização é aplicada na etapa de segmentação das feições para separar as feições. Algumas feições são segmentadas com limiarização simples e outras com a limiarização ótima. Os olhos, por exemplo, são segmentados com limiarização ótima com base na média dos tons de cinza na região da face onde os olhos se encontram, como pode ser visto na Figura 2.4.

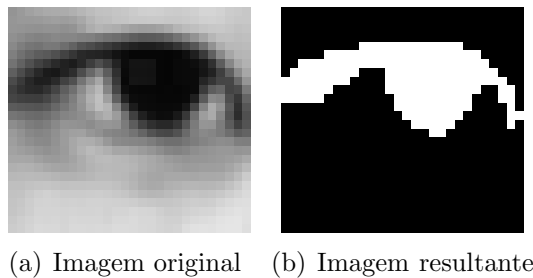


Figura 2.4: Segmentação do olho por limiarização

2.2 *Haar Cascade*

Haar Cascade é um algoritmo de detecção rápida de objetos em imagens digitais idealizado por Viola e Jones (VIOLA; JONES, 2001). A priori o algoritmo foi desenvolvido para a detecção de faces humanas, mas outros pesquisadores provaram que é possível aplicar o *Haar Cascade* para detectar outros tipos de objetos (LIENHART; MAYDT, 2002).

O *Haar Cascade* utiliza o classificador *AdaBoost Cascades* para detectar os objetos. A classificação no *Haar Cascade* é realizada com base nas características pseudo-haar (do inglês *haar-like features*) dos objetos (VIOLA; JONES, 2001). As características pseudo-haar, mostradas na Figura 2.5, são determinadas a partir da diferença de contraste entre grupos retangulares de pixels adjacentes (INTEL CORPORATION, 2001). A variância de contraste entre os grupos de pixels são então utilizadas para determinar áreas claras e escuras. Dois ou três grupos com variância relativa formam uma característica pseudo-haar. Características pseudo-haar podem ser facilmente escaladas possibilitando a detecção de objetos de diversos tamanhos.

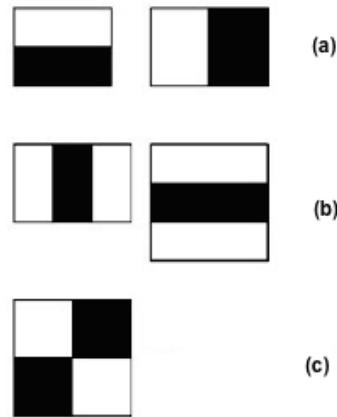


Figura 2.5: (a) Características pseudo-haar de Vértice; (b) Características pseudo-haar de Linha; (c) Características pseudo-haar de quatro retângulos; (INTEL CORPORATION, 2001)

O *Haar Cascade* é utilizado neste trabalho para detecção e segmentação da face nas imagens. A Figura 2.6 mostra um exemplo do resultado da execução do *Haar Cascade* dentro da metodologia proposta.

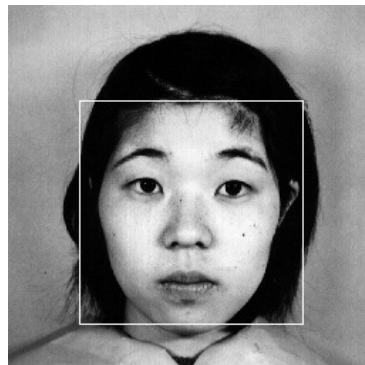


Figura 2.6: Detecção da face com *Haar Cascade*

2.3 Flandmark

Flandmark é uma biblioteca de código aberto escrita em linguagem C que implementa um detector de pontos de referência na face. Ela é capaz de detectar a posição dos cantos da boca, dos cantos dos olhos, da ponta do nariz e do centro da face, mas é preciso fornecer uma imagem da face previamente segmentada para obter uma precisão melhor.

A flandmark usa um classificador de saída estruturada baseado no *Deformable Part Models* (DPM) (DAVID; FELZENSZWALB; HUTTENLOCHER, 2005). No classificador presente na flandmark, a qualidade de uma determinada configuração de

pontos de referência faciais $s = (s_0, \dots, s_{M-1})$ de uma imagem I é medida por uma função de pontuação $f : I \times S \rightarrow \mathbb{R}$. A função de pontuação é definida como a soma do ajuste de aparência e custo de deformação. A função de pontuação pode ser vista na Equação 2.6.

$$f(I, s) = \sum_{i=0}^{M-1} q_i(I, s_i) + \sum_{i=1}^{M-3} g_i(s_0, s_i) + g_5(s_1, s_5) + g_6(s_2, s_6) + g_7(s_0, s_7) \quad (2.6)$$

As funções $q_i(I, s_i)$ e $g_i(s_0, s_i)$, que correspondem respectivamente ao ajuste de aparência e custo de deformação, são parametrizadas. Estes parâmetros são aprendidos pelo algoritmo a partir de exemplos previamente anotados usando o algoritmo de saída estruturada SSVM (Structured Support Vector Machine) (TSOCHANTARIDIS et al., 2005). A maximização de f é realizada com Programação Dinâmica utilizando as restrições de forma dos grafos direcionados acíclicos.

A landmark é usada neste trabalho para a obtenção das Regiões de Interesse (ROI), que são definidas com base nas coordenadas dos pontos de referência faciais e nas medidas antropométricas da face humana. A Figura 2.7 mostra um exemplo do resultado da aplicação do landmark nas imagens utilizadas neste trabalho, no qual os pontos de referência faciais detectados são desenhados na imagem de entrada.

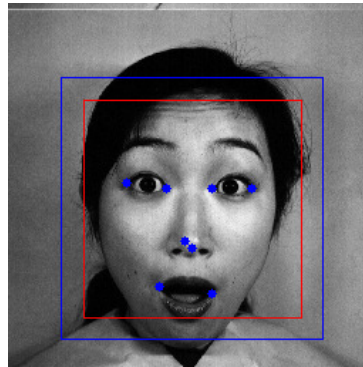


Figura 2.7: Imagem resultante após a execução do landmark

2.4 Classificação de Objetos com Árvores de Decisão

A classificação é uma das estratégias utilizadas para o reconhecimento de objetos (JAIN; KASTURI; SCHUNCK, 1995). Classificar consiste em atribuir uma classe para um elemento de teste através da análise dos valores das suas características (NIXON;

AGUADO, 2008).

A classificação de objetos geralmente envolve um processo de aprendizado no qual são gerados os critérios de decisão do algoritmo de classificação. Existem diferentes algoritmos de classificação e também diferentes maneiras de representar os seus critérios de classificação.

Árvores de decisão são uma forma simples de representação do conhecimento. Elas são amplamente utilizadas em algoritmos de classificação, como um meio eficiente para construir classificadores baseados em um conjunto de valores de atributos (GARCIA; ALVARES, 2000). Para fins de melhor compreensão, um determinado conjunto de valores de atributos será chamado de instância.

As árvores de decisão são árvores binárias³ em que cada nó folha representa uma classe e cada nó não folha representa um atributo utilizado como critério de decisão.

Para prever a classe da instância a partir da árvore de decisão é necessário percorre-la até alcançar um nó folha. A árvore é percorrida a partir do nó raiz, e para cada nó não folha, o valor do seu atributo é comparado com o valor do mesmo atributo da instância. Dependendo do resultado da comparação, percorre-se a árvore pelo caminho da esquerda ou da direita. Este processo se repete até que se alcance o nó folha com o valor da classe resultante da previsão (INTEL CORPORATION, 2001). A Figura 2.8 ilustra um exemplo de árvore de decisão que classifica expressões faciais com base em características como: abertura da boca, altura da sobrancelha e abertura dos olhos.

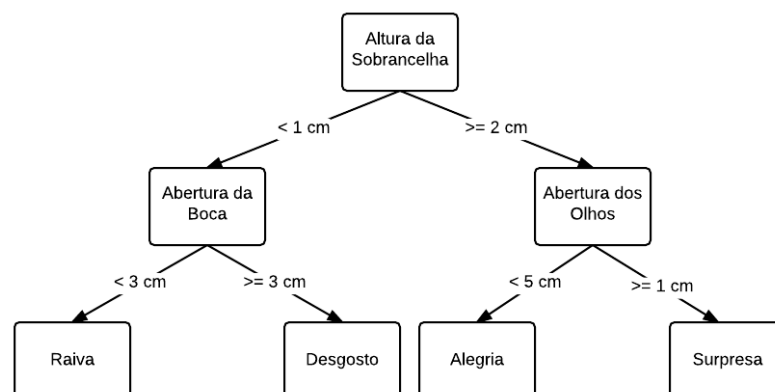


Figura 2.8: Exemplo de uma árvore de decisão de reconhecimento de expressões faciais

Uma árvore de decisão é gerada a partir de um algoritmo que analisa um

³Árvores cujos nós que não são folhas possuem dois nós filhos

conjunto de elementos destinadas para treinamento, previamente classificadas, e cria uma árvore de decisão capaz prever a classe de novos elementos. Existem diversos algoritmos de árvore de decisão, e dentre eles os mais usados são o ID3, o C4.5, o C5.0 e o CART.

O algoritmo ID3 foi proposto por J. Ross Quinlan (MITCHELL, 1997) para a geração de árvores de decisão. Neste algoritmo, constrói-se a árvore de decisão a partir de uma base de treinamento, formada por conjunto de instâncias previamente classificadas. O ID3 separa a base de treinamento em subconjuntos de instâncias da mesma classe. A divisão do subconjunto, operação conhecida como *splitting*, é efetuada a partir de um atributo. O atributo é selecionado a partir de uma propriedade estatística, denominada ganho de informação, que pondera o quanto informativo é um atributo. E então, a base continua a subdividir-se até que todas as instâncias pertençam à uma única classe ou até que o melhor ganho de informação seja menor que zero (ROKACH; MAIMON, 2014).

O C4.5 é o sucessor do ID3 e também foi desenvolvido por Quinlan. Nele, removeu-se a restrição de que os atributos deveriam ser categóricos, ou seja, os atributos deveriam armazenar um valor discreto que particiona os valores contínuos dos atributos em conjuntos discretos de intervalos. Além disso, o C4.5 converte as árvores treinadas em conjuntos de regras. A acurácia de cada regra é avaliada a fim de determinar a ordem na qual elas devem ser aplicadas. A poda da árvore é realizada removendo a pré-condição de uma regra caso a sua acurácia aumente com a remoção. O C5.0 foi a última versão do algoritmo desenvolvido por Quinlan, mas fora publicado sob uma licença proprietária. Ele utiliza menos memória e constrói conjuntos de regras menores que o C4.5 além de ter melhor acurácia.

O CART (*Classification and Regression Trees*) é um algoritmo bastante similar ao C4.5, porém ele difere por suportar variáveis alvo numéricas (regressão) e por não computar conjuntos de regras. Ele constrói árvores binárias buscando a condição de *splitting* que possui o maior ganho de informação para cada nó.

A árvore CART é definida matematicamente da seguinte maneira. Dados vetores de treinamento $x_i \in R^n$, $i = 1, \dots, I$ e um vetor de etiquetas de classe $y \in R^l$, uma árvore de decisão particiona o espaço recursivamente de forma que as amostras com a mesma etiqueta estejam no mesmo grupo. Considerando os dados em um nó m seja representado por Q . Para cada divisão, ou *split*, candidata $\theta = (j, t_m)$ consistindo de um atributo j e um limiar t_m , que particiona os dados nos subconjuntos $Q_{left}(\theta)$ e $Q_{right}(\theta)$.

$$Q_{left}(\theta) = (x, y) | x_j \leq t_m Q_{right}(\theta) = Q / Q_{left}(\theta) \quad (2.7)$$

Neste trabalho, optou-se pela árvore de decisão CART para a realização da classificação das expressões faciais pois este apresenta um bom desempenho se comparado aos outros algoritmos de árvore, além de ser um algoritmo de código aberto.

2.5 Avaliação de Modelos de Classificação

A avaliação de modelos de classificação permite comparar, qualificar e prever o comportamento dos classificadores.

A avaliação da performance de um modelo de classificação não pode ser feita com as mesmas instâncias utilizadas para treinamento, pois um classificador pode facilmente prever corretamente uma instância que foi utilizada no seu treinamento. E isto pode gerar uma visão superestimada do modelo de classificação. Por isso foram desenvolvidos métodos de particionamento de base de dados (instâncias) que utilizam um conjunto de instâncias que não foi utilizada para treinamento para avaliar a performance do classificador (TAN; STEINBACH; KUMAR, 2005).

A validação cruzada é um método de particionamento que divide a base de dados, de forma aleatória, em k conjuntos disjuntos de tamanho igual. Ela busca no processo de particionamento fazer com que cada conjunto possua praticamente a mesma distribuição de classes dentre suas instâncias (ARLOT; CELISSE, 2010).

O classificador é então treinado k vezes, cada vez com um conjunto diferente separado para teste e o resto para treinamento, e os dados de avaliação são armazenados. Por fim, o resultado da avaliação é calculado a partir da média dos k resultados.

2.5.1 Matriz de Confusão

Uma matriz de confusão contém informações sobre previsões corretas e incorretas realizadas por um classificador. E as informações contidas nesta matriz são frequentemente utilizadas para avaliar os classificadores.

Cada entrada $f_{i,j}$ na matriz denota o número de vezes que ao realizar uma

predição uma instância da classe j foi classificada como uma classe i . Considerando-se i a classe correta e j a classe prevista. Se $i \neq j$, então $f_{i,j}$ armazena o número de previsões incorretas da classe i que resultaram na classe j . Se $i = j$, então $f_{i,j}$ armazena o número de previsões corretas da classe i .

Mesmo que a matriz de confusão forneça informações necessárias para determinar o desempenho do modelo de classificação, é possível abstrair estes dados em outras informações relevantes para a avaliação. Tal abstração pode ser realizada utilizando métricas de performance em uma matriz de duas classes, a classe positiva e a classe negativa. Desta duas classes pode-se extrair as métricas acurácia, taxa de erro, especificidade, precisão, sensibilidade e *f-measure*. A definição de cada uma destas métricas é como segue:

- Acurácia (ACU): proporção do total de previsões que estavam corretas, definida pela Equação 2.8;
- Taxa de Erro (TE): proporção do total de previsões que estavam incorretas, definida pela Equação 2.9;
- Precisão (PRE): proporção das previsões que resultaram em positivo que estavam corretas, definida pela Equação 2.10;
- Verdadeiro Negativo (VN): proporção das previsões dos casos negativos que resultaram em negativo, definida pela Equação 2.11;
- Sensibilidade ou *Recall* ou Verdadeiro Positivo (SEN): proporção das previsões corretas da classe positivo, definida pela Equação 2.12;
- Especificidade (ESP): proporção das previsões de casos que resultaram em negativo que estavam corretas, definida pela Equação 2.13;
- *F-measure* (F_β): média ponderada entre a taxa de precisão e de *recall*, definida pela Equação 2.14. Se $\beta = 1$, a precisão e o *recall* possuem o mesmo peso no cálculo do F_β , que agora pode ser chamado de F_1 .

$$ACU = \frac{\text{Número de Previsões Incorretas}}{\text{Número Total de Previsões}} \quad (2.8)$$

$$TE = \frac{\text{Número de Previsões Incorretas}}{\text{Número Total de Previsões}} \quad (2.9)$$

$$PRE = \frac{\text{Número de Previsões Positivas Corretas}}{\text{Número Total de Casos Positivos}} \quad (2.10)$$

$$VN = \frac{\text{Número de Previsões Negativas Corretas}}{\text{Número Total de Casos Negativos}} \quad (2.11)$$

$$SEN = \frac{\text{Número de Previsões Positivas Corretas}}{\text{Número de Previsões Positivas}} \quad (2.12)$$

$$ESP = \frac{\text{Número de Previsões Negativas Corretas}}{\text{Número de Previsões Negativas}} \quad (2.13)$$

$$F_{\beta} = (1 + \beta^2) * \frac{PRE + REC}{(\beta^2 * PRE) + REC} \quad (2.14)$$

A Tabela 2.1 ilustra um exemplo de uma matriz de confusão com duas classes, a classe positivo e a classe negativo, e as equações que definem cada uma das métricas.

Matriz de Confusão		Previsão			
		Positivo	Negativo		
Caso	Positivo	a	b	Precisão	a / a + b
	Negativo	c	d	V.Negativo	c / c + d
		Sensibilidade	Especificidade		
		a / a + c	b / b + d		

Tabela 2.1: Matriz de confusão das classes positivo e negativo

As métricas aplicadas à uma matriz de duas classes pode ser aplicada também à uma matriz de n classes. Mas para isso é necessário realizar uma conversão da matriz de n classes para uma matriz de 2 classes a cada classe analisada. Na conversão, a classe analisada torna-se a classe positivo e a soma das colunas das classes torna-se a classe negativo. Assim, pode-se extrair as métricas da classe individualmente, como é ilustrado na Figura 2.9, que apresenta a conversão de uma matriz de confusão n classes para uma matriz de confusão de 2 classes e o cálculo de suas métricas.

A maioria dos algoritmos de classificação buscam gerar modelos de classificação com alta precisão e baixa taxa de erro.

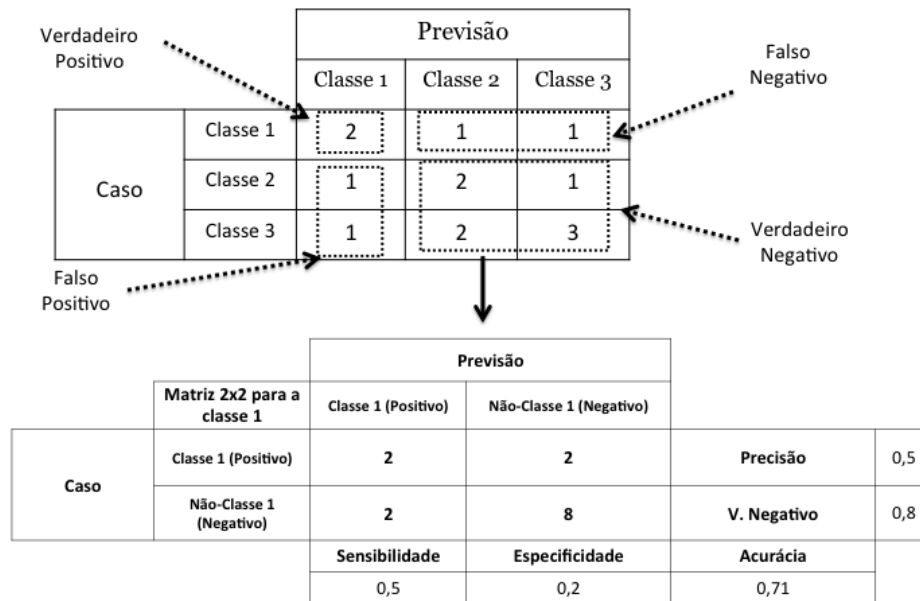


Figura 2.9: Conversão de matriz de confusão de n classes para 2 classes

Neste trabalho, a matriz de confusão é utilizada para a avaliação dos resultados obtidos pela metodologia proposta, pois provê uma visão mais clara do desempenho do modelo de classificação que é gerado na etapa final da metodologia. Ela também possibilita uma posterior comparação do modelo de classificação gerado com outros modelos de classificação.

3 Metodologia Proposta

A metodologia proposta tem como objetivo realizar o reconhecimento das expressões faciais universais a partir de imagens digitais. Ela é composta pela seguinte sequência de etapas: extração de feições, computação dos *Facial Characteristic Points*, computação dos *Facial Action Parameters* e classificação da expressão facial. O esquema básico desta metodologia é apresentado na Figura 3.1. Nas subseções que seguem, cada uma das etapas da metodologia é apresentada de maneira mais aprofundada.

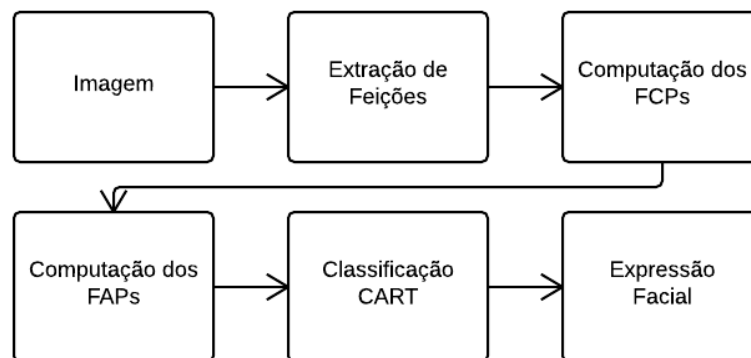


Figura 3.1: Etapas da Metodologia

3.1 Extração de Feições

A etapa de extração de feições é constituída por um *pipeline* de subetapas, cada uma envolvendo a execução de um conjunto de operações específicas. Este *pipeline* é determinado como segue:

- Segmentação da ROI da Face;
- Detecção dos cantos da boca, cantos dos olhos, centro do nariz e centro da face;
- Cálculo das ROIs dos olhos, das sobrancelhas e da boca;
- Segmentação dos olhos;
- Segmentação das sobrancelhas;

- Segmentação do *bounding box*¹ da boca.

Inicialmente, detecta-se a ROI da face. Esta ROI consiste na menor área retangular, ou *bounding box*, da imagem que contém a face. Ela é detectada utilizando o algoritmo detecção de faces *Haar Cascade*, que fornece a origem $O_f(i, j)$, a altura h_f e o comprimento w_f da ROI da face. E a face é por fim segmentada. O objetivo disto é facilitar a busca das feições, pois sabe-se que estas estão localizadas na área delimitada pela face. A segmentação é realizada a partir da imagem de origem utilizando as medidas da ROI.

Na subetapa seguinte, a *flandmark* é aplicada na imagem segmentada da face para detectar os cantos direito e esquerdo de cada um dos olhos, cantos direito e esquerdo da boca, origem do nariz e centro da face. A execução da *flandmark* retorna as coordenadas de cada um destes pontos de interesse.

Continuando a sequência, as ROIs dos olhos, das sobrancelhas e da boca são calculadas a partir das informações previamente extraídas. A ROI dos olhos é a primeira dentre estas ROIs a ser calculada, pois a medida da distância entre os olhos é utilizada para determinar as demais ROIs. Assim como a ROI da face, ela é determinada a partir do comprimento, da altura e da origem dos olhos. O comprimento w_e é determinado através da distância no eixo x entre o canto esquerdo P_{ee} e o canto direito P_{ed} do olho, como na Equação 3.3. A altura h_e é definida a partir da Equação 3.2. Segundo testes realizados, o melhor valor de α encontrado foi $\frac{1}{6}$. O cálculo da origem da cada ROI de cada um dos olhos leva em consideração a origem do nariz P_{ny} , o canto esquerdo do olho P_{ee} e a altura do olho h_e , como mostrado na Equação 3.3.

$$w_e = P_{eex} - P_{edx} \quad (3.1)$$

$$h_e = \alpha * h_f \quad (3.2)$$

$$O_e(i, j) = (P_{eex}, P_{eey}) \quad (3.3)$$

Segundo Vukanidovic (VUKADINOVIC; PANTIC, 2005), existe uma relação de proporção na face entre a localização das feições e a distância entre os olhos. A Figura 3.2 ilustra esta relação de proporção. E por isso, a distância entre os olhos ED é utilizada como unidade de medida para calcular as demais ROIs. A distância entre os olhos é

¹*bounding box*, em duas dimensões, é menor retângulo capaz de conter um conjunto de elementos.

calculada pela distância no eixo x entre o canto mais próximo do nariz do olho esquerdo e do olho direito.

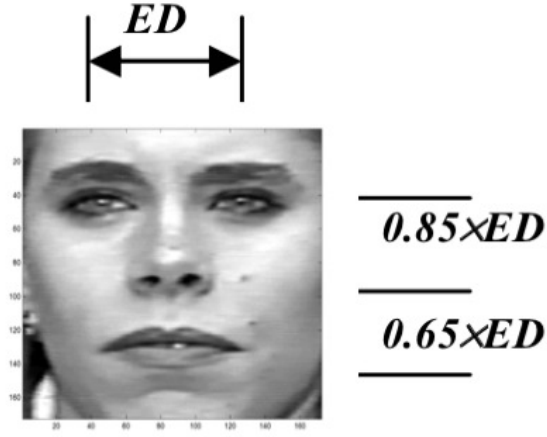


Figura 3.2: Proporções da face humana com base na distância entre os olhos (VUKADINOVIC; PANTIC, 2005)

A ROI da boca é calculada com base na distância entre os olhos e nas medidas da ROI da face. O comprimento da boca w_b é definido pela distância no eixo x entre os cantos da boca, como mostrado na Equação 3.4. A altura h_b é definida pela Equação 3.5, que leva em consideração a distância entre os olhos ED . Definiu-se o β igual à $1/6$, pois este foi o melhor valor encontrado nos testes realizados sobre a base JAFFE. A origem da boca $O_b(i, j)$ é encontrada a partir da Equação 3.6, que leva em consideração a origem do nariz P_n e o canto esquerdo da boca P_{be} .

$$w_b = P_{be_x} - P_{bd_x} \quad (3.4)$$

$$h_b = 0,65 * ED \quad (3.5)$$

$$O_b(i, j) = (P_{be_x}, P_{ny}) \quad (3.6)$$

A ROI das sobrancelhas é formada por uma fatia da sobrancelha, que fica localizada logo acima dos olhos. O comprimento e a altura da ROI da sobrancelha é igual à dos olhos. E a origem desta ROI é definida pela Equação 3.7, considerando P_{eex} e P_{eey} as coordenadas x e y do canto esquerdo do olho, respectivamente.

$$O_s(i, j) = (P_{eex}, P_{eey} - h_e) \quad (3.7)$$

Os olhos são segmentados a partir da análise da sua textura. Como as ROIs dos olhos contém pixels relacionados à pele, aos olhos e à sobrancelha, optou-se por utilizar o método utilizado para realizar a segmentação (MOREIRA; BRAUN; MUSSE, 2010). Segundo (KOURKOUTIS; PANOULAS; HADJILEONTIADIS, 2007), a pele possui altos níveis de intensidade de cor vermelha em imagens RGB² e por isto, esta característica é explorada na imagem para realizar a remoção.

Assim, a remoção é realizada da seguinte maneira. Ressalta-se intensidade do vermelho em cada pixel, zerando o valor dos canais azul e verde de cada pixel e atribuindo ao valor do canal vermelho o complemento do valor que havia nele, i.e., subtrai-se 255 da intensidade do vermelho. Em seguida, aplica-se o operador exponencial para aumentar o contraste da imagem resultante. E por fim, a imagem é limiarizada com base no valor médio de intensidade do canal vermelho dos pixels contidos na ROI. A Figura 3.3 apresenta um exemplo da execução desta sequência de passos.

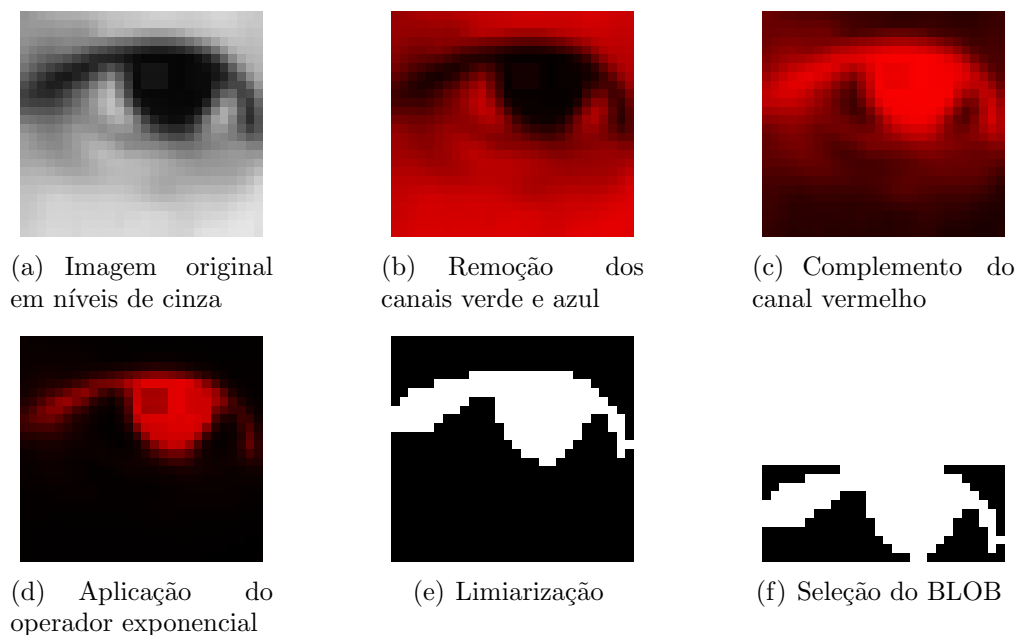


Figura 3.3: Exemplo de segmentação dos olhos

Uma vez a pele removida, é necessário selecionar as regiões da ROI que possuem somente pixels relacionados aos olhos, pois é possível que a imagem também contenha pixels relacionados às sobrancelhas. Para isto, divide-se a imagem resultante em BLOBS³. Segundo o método, a BLOB que contém os olhos é a que está mais próxima do centro do ROI, e por isto ela é selecionada dentre as outras e em seguida segmentada

²RGB é um padrão de cores de imagens digitais em que todas as cores são definidas por combinações dos tons vermelho, verde e azul

³BLOBS, ou *Binary Liked Objects*, são grupos de pixels da mesma cor conectados em sequência

da imagem.

As sobrancelhas são segmentadas da mesma maneira que os olhos, pois as ROIs das sobrancelhas contém pixels relacionados à pele, às sobrancelha e aos olhos. A Figura 3.4 contém um exemplo de segmentação da sobrancelha, com os resultados de cada um dos passos executados. Somente a parte da sobrancelha que fica acima dos olhos é segmentada, pois é nesta parte que estão as informações necessárias para a metodologia. Assim como na ROI dos olhos, ao remover a pele, é possível que restem pixels relacionados aos olhos na imagem. O critério de decisão utilizado para discernir do BLOB da sobrancelha é a proximidade ao centro da ROI. Após selecionado o BLOB correto, a sobrancelha é segmentada da imagem.

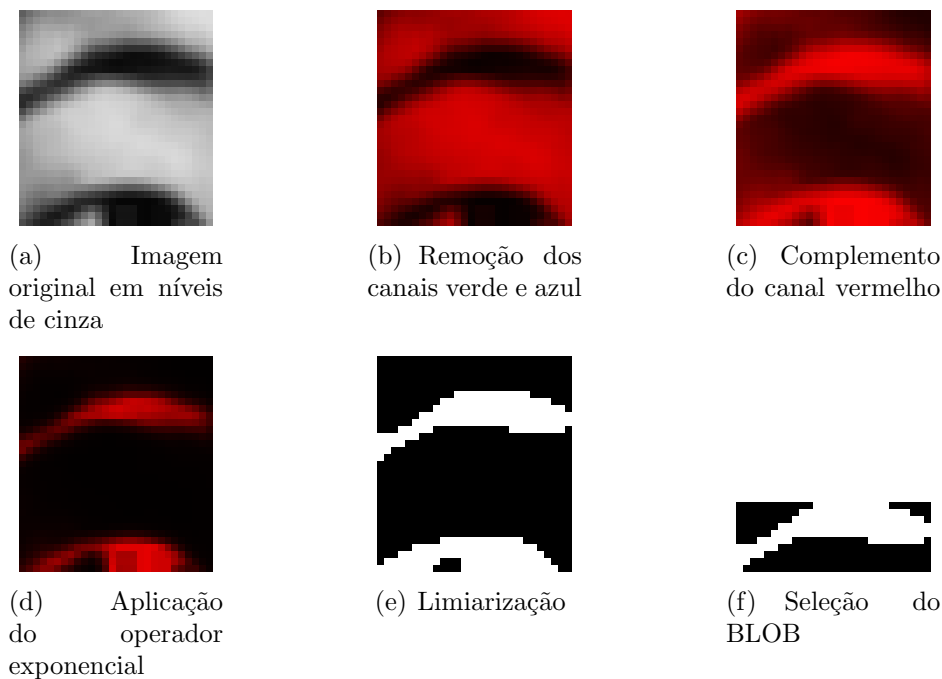


Figura 3.4: Exemplo de segmentação da sobrancelha

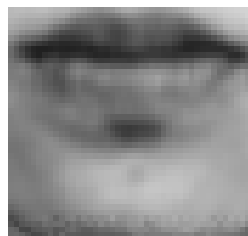
Sabe-se que a boca está localizada na região delimitada pelos lábios. Por isto, as características de textura dos lábios são utilizados para a segmentação do *bounding box* da boca.

Esta segmentação é realizada seguindo esta operações:

- Conversão em níveis de cinza;
- Limiarização;
- *Canny Edge Detector*;

- Operação de Abertura;
- Seleção do BLOB.

A imagem é convertida para níveis de cinza a fim de facilitar a aplicação dos filtros. Em seguida, ela é limiarizada, utilizando-se como limiar a intensidade do nível de cinza igual a 100. O *Canny Edge Detector* é aplicado para segmentar o contorno dos lábios. Para tal, o limiar mínimo e máximo do Canny utilizado é 50 e 100, respectivamente. Vale ressaltar que o valor dos limiares da limiarização e do *Canny Edge Detector* foram definidos a partir de testes realizados sobre a base de imagens. A operação de abertura com elemento estruturante de cruz com tamanho 4 é aplicada para excluir pequenos detalhes e linhas muito finas que poderiam atrapalhar na detecção dos lábios. A Figura 3.5 mostra um exemplo de segmentação da boca utilizando a metodologia proposta. A partir da imagem resultante da segmentação, geram-se BLOBs contendo os conjuntos de linhas detectados. E seleciona-se o BLOB que possui o centroide mais alto, i.e, com maior valor de y . O *bounding box* capaz de conter o BLOB é calculado. Por fim, o *bounding box* da boca é segmentado.



(a) Imagem original em nível de cinza



(b) Limiarização



(c) *Canny Edge Detector*



(d) Operação de Abertura



(e) Seleção do BLOB

Figura 3.5: Exemplo de segmentação da boca

3.2 Computação dos *Facial Characteristic Points*

Segundo King (KING; HOU, 1996), existem 30 pontos característicos da face humana, do inglês *Facial Characteristic Points* (FCP's), que estão relacionados à expressão da emoção humana. Eles podem ser vistos na Figura 3.6.

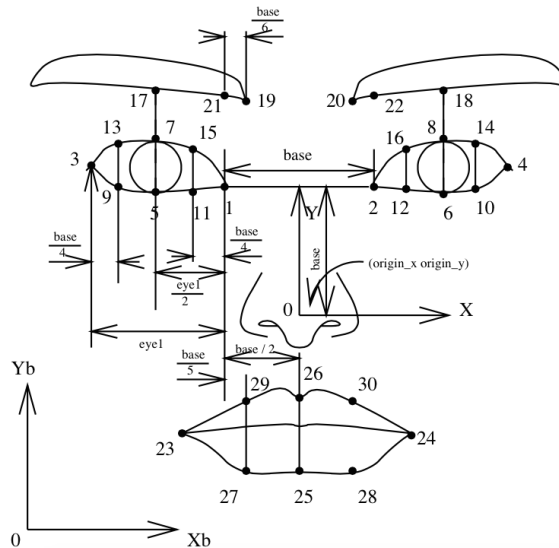


Figura 3.6: *Facial Characteristic Points* (KING; HOU, 1996)

A metodologia proposta neste trabalho utiliza 12 FCPs e dois pontos na face correspondentes aos centroides das sobrancelhas. Os FCPs são representados a partir da nomenclatura $FCP + numeroFCP$, que leva em consideração a numeração utilizada na Figura 3.6. Os FCPs selecionados são FCP1, FCP2, FCP3, FCP4, FCP5, FCP6, FCP7, FCP8, FCP23, FCP24, FCP25, FCP26 e FCP28, que podem ser vistos em destaque na Figura 3.7. Os centroides da sobrancelha esquerda e direita são representados respectivamente por $CEN1$ e $CEN2$, e são mostrados na Figura 3.8.

Os 12 FCPs e dos centroides das sobrancelhas são encontrados a partir das informações das feições obtidas nas etapas anteriores. Sendo que o valor de cada um destes pontos corresponde à sua coordenada (x, y) na imagem. A metodologia utiliza os centroides das sobrancelhas no lugar do FCP19 e do FCP20. O objetivo desta escolha é gerar características menos específicas à sobrancelha de cada indivíduo.

O resultado retornado pela execução do landmark contém as coordenadas dos FCPs: FCP1, FCP2, FCP3, FCP4, FCP23 e FCP24, correspondentes aos cantos dos olhos, e dos FCPs: FCP23 e FCP24, correspondentes aos cantos da boca. O restante dos FCPs são calculados a partir das informações contidas nos BLOBs dos olhos, da

sobrancelha e da *bounding box* da boca. Os limites no eixo x e y, o perímetro, o conjunto de pontos que o compõe e o centroide são algumas das informações disponibilizadas pelos BLOBs.

As coordenadas dos centroides das sobrancelhas são as mesmas dos centroides dos BLOBs das sobrancelhas. FCP5, FCP6, FCP7 e FCP8 são os limites verticais dos olhos, e são obtidos partir dos limites, em x e em y, do BLOB dos olhos. E da mesma maneira obtêm-se o FCP25 e o FCP26, correspondentes aos limites verticais da boca. Na próxima seção será explicado como os valores dos FCPs são utilizados para calcular as características da face que serão utilizadas para a classificação.

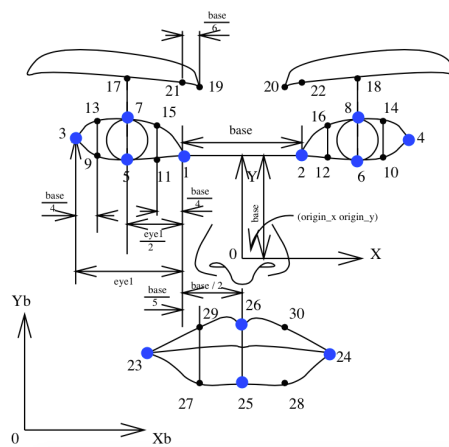


Figura 3.7: *Facial Characteristic Points* utilizados na metodologia representados pelos pontos azuis

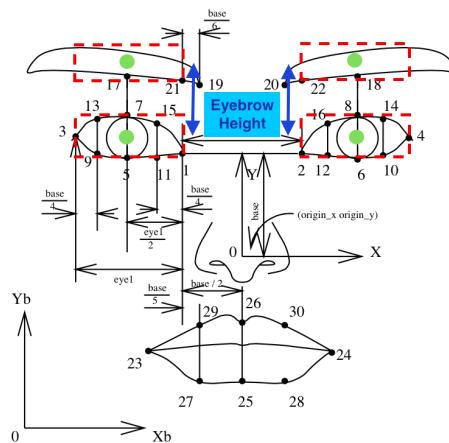


Figura 3.8: Cálculo do centroide da sobrancelha

3.3 Computação dos *Facial Animation Parameters*

Os *Facial Animation Parameters* (FAP) são uma lista de atributos definidos no padrão ISO MPEG-4 da Adobe, que servem para a reprodução de emoções, expressões e pronúncias em vídeos (PANDZIC; FORCHHEIMER, 2002). Nesta metodologia optou-se por calcular o valor dos FAPs utilizados por Perveen (PERVEEN; GUPTA; VERMA, 2012). Segundo ele, de todos os FAPs encontrados no rosto humano, somente cinco são efetivamente relevantes para o reconhecimento de expressões faciais, são eles a abertura dos olhos AO , o comprimento dos olhos CO , a abertura da boca AB , o comprimento da boca CB e altura das sobrancelhas AS . Sendo que estas medidas correspondem à média entre os valores dos FAPs do lado esquerdo e direito do rosto. A Figura 3.9 mostra como os FAPs são calculados a partir das medidas do rosto, sendo que o valor referente ao lado esquerdo e direito é diferenciado pelo caractere e e d , respectivamente, ao fim do nome do FAP.

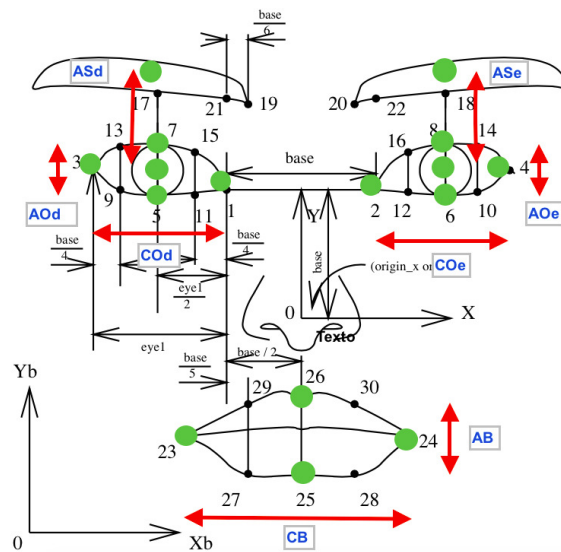


Figura 3.9: Medidas do rosto, representadas pelas setas azuis, utilizadas para calcular os FAPS da metodologia

Os valores dos FAPs são calculados a partir dos valores dos FCPs adquiridos na etapa anterior, segundo as Equações 3.8, 3.9, 3.10, 3.11 e 3.12. Nestas equações, a nomenclatura $FCP(numero)_x$ e $FCP(numero)_y$ é empregada para representar, respectivamente, as coordenadas x e y de cada um dos 30 FCPs. Os elementos $cenEsq_y$ e $cenDir_y$ correspondem às coordenadas y do centroide da sobrancelha esquerda e direita, respectivamente. A altura da face h_f e o comprimento da face c_f são utilizados para normalizar os valores dos FAPs. As medidas referentes ao comprimento são normalizadas

com o c_f e as medidas referentes à altura são normalizadas com h_f .

$$AO = \frac{((FCP7_y - FCP5_y) + (FCP8_y - FCP6_y))}{2h_f} \quad (3.8)$$

$$CO = \frac{((FCP1_x - FCP3_x) + (FCP4_x - FCP2_x))}{2c_f} \quad (3.9)$$

$$AB = \frac{(FCP26_y - FCP25_y)}{h_f} \quad (3.10)$$

$$CB = \frac{(FCP24_y - FCP23_y)}{c_f} \quad (3.11)$$

$$AS = \frac{((cenEsq_y - (FCP7_y - FCP5_y)) + (cenDir_y - (FCP8_y - FCP6_y)))}{2h_f} \quad (3.12)$$

3.4 Classificação das Expressões Faciais

A classificação é a etapa da metodologia responsável por definir a que classe de expressões faciais pertence a expressão contida na imagem. Esta classificação é realizada a partir da análise dos valores dos FAPs adquiridos nas etapas anteriores.

A árvore de decisão CART foi o classificador escolhido para ser utilizado nesta metodologia. Árvores de decisão são ferramentas poderosas e populares de classificação (XIAO; MA; KHORASANI, 2006), além de serem fáceis de compreender e interpretar. Elas requerem pouca preparação dos dados, por exemplo, não requerem normalização. E possuem a capacidade de manusear bases de dados categóricas e numéricas, e de abordar problemas com múltiplas saídas. As árvores utilizam o modelo de classificação denominado caixa branca, que permite compreender claramente como se dá o processo de classificação (PEDREGOSA et al., 2011). Assim, é possível visualizar os critérios utilizados e o caminho percorrido até a resposta final do classificador. Isso ocorre por que elas são uma forma de representar regras (GUPTA; VERMA; PERVEEN, 2012), e regras são facilmente compreendidas por humanos. Essa característica das árvores as difere de outros classificadores, como as redes neurais, que utilizam um modelo de caixa preta.

A tarefa do classificador é atribuir etiquetas de classes para os indivíduos de um conjunto. Mas para isso inicialmente é necessário definir todas as classes existentes no problema em questão. Em seguida, é preciso treinar o classificador para que seja capaz de identificar as características que definem cada uma das classes definidas.

Nesta metodologia são definidas sete classes, as seis expressões faciais universais e uma expressão neutra. A expressão neutra representa a inexistência de expressão facial no rosto de uma pessoa. Em seguida o classificador é treinado com uma base de características. As características contém os valores dos FAPs extraídos de uma base de imagens de expressões faciais previamente classificadas. Com o modelo de classificação treinado, o classificador é utilizado para prever a que classe de expressões faciais pertence a expressão presente em uma imagem.

4 Resultados e Discussão

Neste capítulo serão apresentados os resultados e a avaliação de desempenho da metodologia proposta para o reconhecimento de expressões faciais em imagens digitais.

A base de imagens JAFFE (*Japanese Female Facial Expression*) (LYONS et al., 1998) foi selecionada para ser utilizada na avaliação da metodologia. A base contém 217 imagens de rostos de japonesas em escala de cinza previamente classificadas em 7 expressões faciais, que são as seis expressões faciais universais e a expressão neutra. Esta base é comumente utilizada em trabalhos relacionados ao reconhecimento de expressões faciais (PERVEEN; GUPTA; VERMA, 2012). A base JAFFE contém o rosto de 11 japonesas diferentes. Isto permite verificar a capacidade da metodologia de reconhecer uma mesma expressão facial realizada por diferentes indivíduos.

Nos testes, cada uma das etapas fora executada separadamente, mas respeitando a sequência da metodologia. Inicialmente realizou-se a extração das feições de todas as imagens da base. A execução dos procedimentos desta etapa gerou resultados promissores, 82 % das imagens foram bem segmentadas, sendo que todos os erros obtidos estavam relacionados à segmentação da boca. A Figura 4.1 apresenta um caso de extração de feições bem sucedido e um caso mal sucedido.

Os FCPs e os FAPs foram computados somente sobre as extrações bem sucedidas, pois busca-se avaliar o desempenho de cada etapa individualmente e o uso de casos mal sucedidos no treinamento poderia causar prejuízo à etapa de classificação. Uma base de características foi criada com os valores resultantes da computação dos FAPs, que é utilizada na etapa de classificação.

O método da validação cruzada é aplicado na realização da avaliação do desempenho da etapa de classificação. Ele foi aplicado utilizando 10 *folds*, ou seja, $k = 10$, pelos seguintes motivos. Um valor de k maior que 10 resultaria em conjuntos de teste pequenos, o que poderia gerar uma avaliação tendenciosa. E um valor de k menor que 10 resultaria em conjuntos de treinamento pequenos, o que poderia prejudicar o processo de aprendizado do algoritmo de classificação.

Uma árvore de decisão é gerada a cada iteração da validação cruzada. Cada

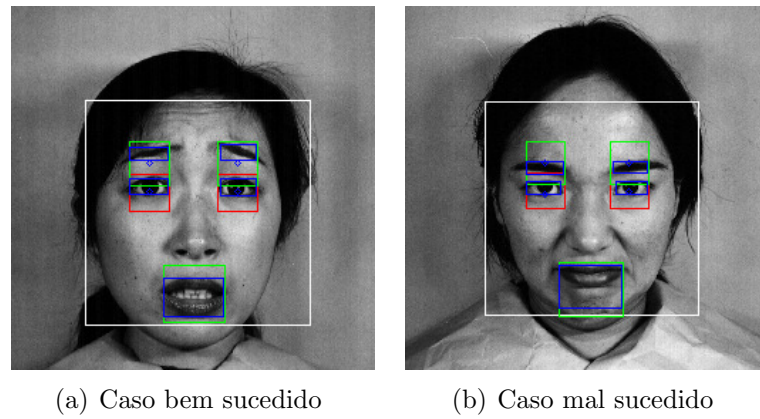


Figura 4.1: *Extração de feições*

nó não-folha destas árvores possui o atributo utilizado como critério de decisão, podendo ser a abertura dos olhos, a altura da sobrancelha, a largura da boca ou a abertura da boca. E estes nós também armazenam o índice GINI e número de indivíduos da base de treinamento que o alcançaram no processo de formação da árvore. Estas árvores permitem visualizar todo o processo de tomada de decisão realizado durante a classificação de um elemento.

Os nós folha da árvore, que são os nós que não possuem filhos, contém o resultado da classificação. Estes nós possuem um vetor de inteiros, no qual cada índice do vetor corresponde a uma classe, e o valor armazenado em cada posição corresponde ao número de ocorrências da classe. Os índices do vetor de valores variam de 0 a 6, correspondendo respectivamente às classes Alegre, Triste, Surpresa, Raiva, Desgosto, Medo e Neutra. A árvore de decisão gerada permitiu identificar que a classificação de nenhuma das classes depende somente de um atributo, mas sim de conjunto de atributos, ou seja, todos os atributos são utilizados no processo de classificação.

A matriz de confusão apresentada na Figura 4.2 corresponde à matriz acumulada durante todas as iterações da classificação. Esta matriz mostra a quantidade total de previsões corretas e incorretas realizadas pelo classificador nas execuções da validação cruzada. As linhas da matriz correspondem à classe correta do elemento classificado, as colunas correspondem à classe prevista pelo classificador e o valor das células corresponde ao número de ocorrências de cada combinação entre classe correta e classe prevista. O número de ocorrências de cada célula da matriz é representado com intensidades de cores. A barra ao lado da matriz mostra a correspondência entre os valores e as cores utilizadas na matriz. As métricas de avaliação calculadas a partir da matriz foram: a precisão, o *recall*, a sensibilidade, a acurácia e o *F-measure*. Elas fornecem

informações quantitativas acerca do desempenho do classificador. Os valores das métricas, que foram extraídas durante a classificação, estão elencados na Tabela 4.1. O trabalho mostrou-se promissor com resultado de 55% de acerto.

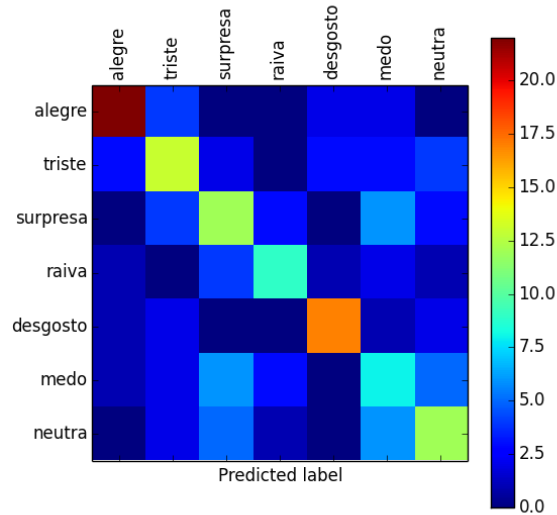


Figura 4.2: Matriz de Confusão resultante da classificação

A precisão e *recall* médio entre as classes foi de aproximadamente 55%. Este valor de precisão indica que as classes obtiveram uma quantidade parecida de classificações corretas. E o valor de *recall* indica que para cada classe obteve-se uma quantidade equivalente de verdadeiros positivos e falsos positivos.

Analisando o desempenho de cada classe individualmente, é possível ver que a classe Surpresa apresentou a maior taxa de precisão com 83% e de *recall* também com 83%. A precisão desta classe leva a concluir que a maioria das classificações corretas foram realizadas sobre indivíduos com a expressão Surpresa. E o *recall* encontrado indica que dentre todas as classes, esta foi a que menos indivíduos de outras classes foram classificados erroneamente como pertencente à classe Surpresa. Já a classe Triste obteve os menores valores de *recall* e precisão, 33% e 36%, respectivamente. Já as demais classes apresentaram resultados próximos à média. O valor do *F-measure* não diferiu muito dos valores das métricas de de precisão e *recall*, indicando um certo equilíbrio do classificador.

As métricas avaliadas indicam que o classificador demonstrou um desempenho relevante, e cada uma das expressões, com exceção das expressões Surpresa e Triste, contribuiu com a acurácia geral do classificador de forma equivalente. E o bom desempenho da classificação da expressão Surpresa foi balanceado com o desempenho

abaixo da média da expressão de Triste.

expressão	precisão	recall	f-measure	ocorrências
alegre	0.60	0.58	0.59	26
triste	0.33	0.36	0.35	25
surpresa	0.83	0.83	0.83	23
raiva	0.45	0.50	0.47	18
desgosto	0.44	0.43	0.44	28
medo	0.50	0.46	0.48	28
neutra	0.73	0.73	0.73	30
média	0.55	0.55	0.55	29

Tabela 4.1: Métricas de avaliação extraídas da classificação

5 Considerações Finais

Este trabalho apresentou a proposta de uma metodologia para o reconhecimento das seis expressões faciais universais em imagens digitais baseada no uso da árvore de decisão CART.

A etapa de extração das feições apresentou bons resultados, necessitando somente de melhorias na extração da boca. As técnicas utilizadas para a segmentação da boca são suscetíveis à iluminação e o que causou a maioria das falhas na sua segmentação.

A etapa de classificação com árvore de decisão CART obteve uma boa taxa de acurácia, levando em consideração que o problema abordado trata-se de uma classificação com múltiplas classes. Também foi possível verificar que a expressão Surpresa é mais facilmente reconhecida a partir das características abordadas na metodologia, enquanto a expressão Triste apresentou uma certa dificuldade para ser reconhecida.

Não foi possível realizar a comparação com os resultados de outras metodologias por pelo menos um dos seguintes motivos: expressões faciais diferentes; base de imagens para teste diferente ou incompleta; método de classificação omitido.

Por fim, a metodologia proposta obteve um resultado promissor na tarefa de reconhecimento das expressões faciais universais e fornece amparo tecnológico para o desenvolvimento de tecnologias futuras. Em trabalhos futuros, propõe-se a inclusão de novas características da face relacionadas às expressões faciais, a utilização de outros algoritmos de classificação para comparação de desempenho e a elaboração um novo método para a extração das feições, que seja adaptável às variações de iluminação, atentando-se principalmente á segmentação da boca. Outra proposta interessante consiste em incluir o reconhecimento de micro-expressões, que podem ser utilizadas para a detecção de mentiras (EKMAN, 2009).

No geral, a metodologia proposta mostrou-se promissora para o desenvolvimento de sistemas que tomem por base o reconhecimento de expressões faciais universais.

Referências

- ARLOT, S.; CELISSE, A. A survey of cross-validation procedures for model selection. *Statist. Surv.*, The American Statistical Association, the Bernoulli Society, the Institute of Mathematical Statistics, and the Statistical Society of Canada, v. 4, p. 40–79, 2010. Disponível em: <http://dx.doi.org/10.1214/09-SS054>.
- DAVID, C.; FELZENSZWALB, P.; HUTTENLOCHER, D. Spatial priors for part-based recognition using statistical models. *Computer Vision and Pattern Recognition(CVPR 2005)*, 2005.
- DONATO, G. et al. Classifying facial actions. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, IEEE, v. 21, n. 10, p. 974–989, 1999.
- EKMAN, P. The argument and evidence about universals in facial expressions of emotion. In: *Handbook of social psychophysiology*. [S.l.]: University of California, 1989. cap. 6, p. 143–164.
- EKMAN, P. Lie catching and microexpressions. *The philosophy of deception*, p. 118–133, 2009.
- FILHO, O. M.; NETO, H. V. *Morfologia Matemática*. [S.l.]: Brasport, 1999. 129–143 p.
- GARCIA, S. C.; ALVARES, L. O. Árvores de decisão – algoritmos id3 e c4.5. *Cadernos de Informática - I Workshop Interno sobre Descoberta de Conhecimento em Bases de Dados*, p. 52–55, 2000.
- GUPTA, S.; VERMA, K.; PERVEEN, N. Facial expression recognition system using facial characteristic points and id3. *International Journal of Computer & Communication Technology (IJCCT) ISSN (ONLINE): 2231 - 0371*, 2012.
- INTEL CORPORATION. *Open Computer Vision Library Reference Manual*. USA, 2001.
- JAIN, R.; KASTURI, R.; SCHUNCK, B. G. Edge detection. In: *Machine Vision*. [S.l.]: McGraw-Hill, 1995. cap. 5, p. 168–173.
- JUANJUAN, C. et al. Facial expression recognition based on pca reconstruction. In: IEEE. *Computer Science and Education (ICCSE), 2010 5th International Conference on*. [S.l.], 2010. p. 195–198.
- KING, I.; HOU, H. Radial basis network for facial expression synthesis. In: CITESEER. *Proceedings of the International Conference on Neural Information Processing*. [S.l.], 1996. p. 1127–1130.
- KOBAYASHI, H.; HARA, F. Recognition of six basic facial expression and their strength by neural network. In: IEEE. *Robot and Human Communication, 1992. Proceedings., IEEE International Workshop on*. [S.l.], 1992. p. 381–386.

- KOURKOUTIS, L. G.; PANOULAS, K. I.; HADJILEONTIADIS, L. J. Automated iris and gaze detection using chrominance: Application to human-computer interaction using a low resolution webcam. In: IEEE. *Tools with Artificial Intelligence, 2007. ICTAI 2007. 19th IEEE International Conference on*. [S.l.], 2007. v. 1, p. 536–539.
- LIENHART, R.; MAYDT, J. An extended set of haar-like features for rapid object detection. *IEEE ICIP 2002*, v. 1, p. 900–903, September 2002.
- LYONS, M. J. et al. *The Japanese female facial expression (JAFFE) database*. 1998.
- MENESES, P. R.; ALMEIDA, T. de. Distorções e correlações dos dados da imagem. In: *Introdução ao Processamento de Imagens de Sensoriamento Remoto*. [S.l.]: Instituto Geociências - Universidade de Brasília, 2012. cap. 6, p. 82–83.
- MITCHELL, T. [S.l.]: McGraw-Hill, 1997. 52–81 p.
- MOREIRA, J. L.; BRAUN, A.; MUSSE, S. R. Eyes and eyebrows detection for performance driven animation. In: IEEE. *Graphics, Patterns and Images (SIBGRAPI), 2010 23rd SIBGRAPI Conference on*. [S.l.], 2010. p. 17–24.
- NIXON, M. S.; AGUADO, A. S. [S.l.]: ELSEVIER, 2008. 329–347 p.
- PANDZIC, I. S.; FORCHHEIMER, R. Mpeg-4 facial animation. *The standard, implementation and applications*. Chichester, England: John Wiley&Sons, Wiley Online Library, 2002.
- PEDREGOSA, F. et al. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, v. 12, p. 2825–2830, 2011.
- PERVEEN, N.; GUPTA, S.; VERMA, K. Facial expression recognition using facial characteristic points and gini index. *Students Conference on Engineering and Systems (SCES 2012)*, March 2012.
- ROKACH, L.; MAIMON, O. [S.l.]: World Scientific Publishing, 2014. 165–192 p.
- TAI, S.; CHUNG, K. Automatic facial expression recognition system using neural networks. In: IEEE. *TENCON 2007-2007 IEEE Region 10 Conference*. [S.l.], 2007. p. 1–4.
- TAN, P.-N.; STEINBACH, M.; KUMAR, V. [S.l.]: Addison-Wesley, 2005. 186–193 p.
- TSOCHANTARIDIS, I. et al. Spatial priors for part-based recognition using statistical models. *Journal of Machine Learning Research* 6, 2005.
- VIOLA, P.; JONES, M. Rapid object detection using boosted cascade of simple features. *IEEE Conference on Computer Vision and Pattern Recognition*, 2001.
- VUKADINOVIC, D.; PANTIC, M. Fully automatic facial feature point detection using gabor feature based boosted classifiers. In: IEEE. *Systems, Man and Cybernetics, 2005 IEEE International Conference on*. [S.l.], 2005. v. 2, p. 1692–1698.
- XIAO, Y.; MA, L.; KHORASANI, K. A new facial expression recognition technique using 2-d dct and neural networks based decision tree. In: IEEE. *Neural Networks, 2006. IJCNN'06. International Joint Conference on*. [S.l.], 2006. p. 2421–2428.