MARCELO JÚNIOR COSSETIN

Reconhecimento De Expressões Faciais Utilizando Redução De Dimensionalidade Para Estratégia De Classificação Um-Contra-Um

Dissertação apresentada ao Programa de Pós-Graduação em Informática da Pontifícia Universidade Católica do Paraná como requisito parcial para obtenção do título de Mestre em Informática.

CURITIBA

2015

MARCELO JÚNIOR COSSETIN

Reconhecimento De Expressões Faciais Utilizando Redução De Dimensionalidade Para Estratégia De Classificação Um-Contra-Um

Dissertação apresentada ao Programa de Pós-Graduação em Informática Aplicada da Pontifícia Universidade Católica do Paraná como requisito parcial para obtenção do título de Mestre em Informática.

Área de Concentração: Ciência da Computação

Orientador: Prof. Dr. Júlio Cesar Nievola
Co-orientador: Prof. Dr. Alessandro Lameiras Koerich

CURITIBA
2015

Dados da Catalogação na Publicação
Pontifícia Universidade Católica do Paraná
Sistema Integrado de Bibliotecas – SIBI/PUCPR
Biblioteca Central

C836r 2015

Cossetin, Marcelo Júnior
Reconhecimento de expressões faciais utilizando redução de dimensionalidade para estratégia de classificação um-contra-um / Marcelo Júnior Cossetin ; orientador: Júlio César Nievola ; co-orientador: Alessandro Lameiras Koerich. – 2015.
xvii, 120 f. : il. ; 30 cm

Dissertação (mestrado) – Pontifícia Universidade Católica do Paraná, Curitiba, 2015
Bibliografia: f. [113]-120

1. Visão por computador. 2. Processamento de imagens. 3. Expressão facial. I. Nievola, Júlio César. II. Koerich, Alessandro Lameiras. III. Pontifícia Universidade Católica do Paraná. Programa de Pós-Graduação em Informática Aplicada. IV. Título.

CDD 21. ed. – 006.37

Pontifícia Universidade Católica do Paraná
Escola Politécnica
Programa de Pós-Graduação em Informática

PUCPR
GRUPO MARISTA

ATA DE DEFESA DE DISSERTAÇÃO DE MESTRADO PROGRAMA DE PÓS-GRADUAÇÃO EM INFORMÁTICA

DEFESA DE DISSERTAÇÃO DE MESTRADO Nº 04/2015

Aos 28 dias do mês de Agosto de 2015 realizou-se a sessão pública de Defesa da Dissertação **" Reconhecimento de Expressões Faciais utilizando Redução de Dimensionalidade para Estratégia de Classificação um-contra-um"** apresentado pelo aluno **Marcelo Júnior Cossetin,** como requisito parcial para a obtenção do título de Mestre em Informática, perante uma Banca Examinadora composta pelos seguintes membros:

Prof. Dr. Júlio César Nievola
PUCPR (Orientador) Julio Cesar Nievola (assinatura) APROV (Aprov/Reprov)

Prof. Dr. Emerson Cabrera Paraiso
PUCPR (assinatura) APROV (Aprov/Reprov)

Prof. Dr. Alessandro Lameiras Koerich
Universidade Du Quebec (assinatura) APROV (Aprov/Reprov)

Prof. Dr. Luiz Eduardo Soares de Oliveira
UFPR (assinatura) APROV (Aprov/Reprov)

Conforme as normas regimentais do PPGIa e da PUCPR, o trabalho apresentado foi considerado APROVADO (*aprovado/reprovado*), segundo avaliação da maioria dos membros desta Banca Examinadora. Este resultado está condicionado ao cumprimento integral das solicitações da Banca Examinadora registradas no Livro de Defesas do programa.

Prof.ª Dr.ª Andreia Malucelli.
Coordenadora do Programa de Pós-Graduação em Informática.

Programa de Pós-Graduação em Informática PPGIa PUCPR

Agradecimentos

Em primeiro lugar agradeço a Deus por ter me dado condições de lutar e alcançar os objetivos pretendidos. Agradeço o orientador Júlio César Nievola e o co-orientador Alessandro Lameiras Koerich pela paciência, estímulo e confiança indispensáveis para o desenvolvimento deste trabalho. Quero agradecer a empresa Hi Technologies na qual trabalhei desde o início desta jornada, pela compreensão e apoio, em especial os diretores Carlos Eduardo Chaves, Marcus Mazega Figueredo e Sérgio Renato Rogal. Também agradeço a Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), que me concedeu uma bolsa de estudos. Agradeço os meus amigos Leonardo Alves Ferreira, Luiz Giovanini e Marcos Hara que me apoiaram durante o desenvolvimento deste trabalho. Por fim quero agradecer as pessoas mais importantes da minha vida, meus pais Janir e Idene e meu irmão Eduardo, pela motivação e confiança passados a mim durante esta pesquisa

Dedico este trabalho aos meus pais Janir
e Idene e ao meu irmão Eduardo.

Sumário

Lista de Figuras

Lista de Tabelas

Lista de Abreviaturas

AAM	*Appearance Active Model*
CFS	*Correlation-based Features Selection*
CK	*Cohn-Kanade Database*
DCT	*Discrete Cosine Transform*
ESOM	*Efficient Second-order Matching*
FEED	*Facial Expression and Emotion Database*
HOG	*Histogram Oriented Gradient*
IG	*Information Gain*
IHM	Interação Humano-Computador
JAFFE	*Japanese Female Facial Expression*
kNN	*k Nearest Neighbor*
KPCA	*Kernel Principal Analysis Component*
KW	*Kruskal Wallis*
LBP	*Local Binary Pattern*
LTP	*Local Ternary Pattern*
MBP	*Medium Binary Pattern*
MLP	*Multilayer Perceptron*
MTP	*Medium Ternary Pattern*
NN	*Neural Network*
PCA	*Principal Component Analysis*
REF	Reconhecimento de expressões faciais
SMV	*Support Machine Vector*
TFEID	*Taiwanese Facial Expression Image Database*
WFT	*Wavelet Fusion Technique*
WLD	*Weber Local Descriptor*

Resumo

O reconhecimento de expressões faciais é um problema desafiador no campo de visão computacional e nas últimas décadas várias alternativas foram propostas. Recentemente foi verificado que extratores de características baseados em geometria são poucos robustos às variações que podem ocorrer no processo de aquisição da imagem, como iluminação, posição e ambiente, dificultando a localização precisa dos pontos fiduciais, com isso os métodos baseados em textura estão ganhando força. Um dos problemas inerentes aos métodos baseados em textura é a geração de vetores de característica com alta dimensionalidade e este trabalho propõe uma abordagem que ameniza esse problema. Para identificar as expressões faciais de alegria, nojo, medo, neutro, surpresa, raiva e tristeza, foram implementadas as etapas de detecção facial, extração de características baseada em textura, redução de dimensionalidade e classificação. Na detecção facial a face do indivíduo é extraída da imagem original eliminando objetos de fundo que possam prejudicar na aprendizagem do modelo. Para a extração de características toda face é dividida zonas ou sub-regiões, sendo que em cada zona é aplicado um algoritmo como WLD ou LBP para obtenção das características, ao final os vetores de atributos extraídos são concatenados para representar a imagem. A redução de dimensionalidade proposta seleciona atributos utilizando exemplos compostos por duas expressões faciais, desta forma se faz necessário aplicar a seleção de atributos para todos os pares possíveis de expressões faciais, considerando as 7 expressões faciais, existem 21 combinações. Cada vetor de atributos gerado é aplicado em um classificador binário especializado nas duas respectivas expressões do vetor de entrada. O resultado da classificação é atribuído à expressão facial com mais votos no conjunto de classificadores binários especializados. Os resultados obtidos foram comparados com a tradicional seleção de atributos que faz uso de todas as expressões e com todo o conjunto de atributos. Testes estatísticos ($p < 0.05$) demonstraram que as taxas de reconhecimento obtidas pelo método proposto competem com as melhores da literatura com a vantagem de possuir dimensionalidade reduzida e não necessitar de pontos fiduciais anotados sobre as imagens de face. Para os conjuntos JAFFE, CK e TFEID foi obtido 99,05%, 98,07% e 99,63% de taxa de acerto respectivamente. Apesar da proposta utilizar 21 classificadores, a seleção de características em pares é capaz de fornecer um número menor atributos do que as demais abordagens.

Palavras-Chave: expressões faciais, redução de dimensionalidade, classificação um-contra-um.

Abstract

In the last few decades, several alternatives have been proposed in the literature regarding the problem of facial expression recognition, which belongs to the field of Computer Vision. It was recently pointed that geometry-based features are sensitive to some variations that may occur during the image acquisition process, such as lighting, position and environment, so the precise location of facial points becomes more difficult. In this context, texture-based methods are highlighting. An important issue related to texture-based methods is the generation of high-dimensional feature vectors, and this work proposes an approach to reduce their dimensionalities without decrease the expression recognition performance. In order to identify the facial expressions of happy, disgust, fear, neutral, surprise, anger and sadness, the stages of facial detection, based-texture feature extraction, dimensionality reduction and classification were implemented. In face detection stage, the face of each individual was cropped from the original image, which removes background objects that might compromises the learning classifiers and also provides alignment base between different faces. For the feature extraction stage, the whole face was first divided into zones or sub-regions, and then an algorithm as Weber Local Descriptor (WLD) or Local Binary Patterns (LBP) was applied in each area to obtain the features. Finally, all feature vectors from different sub-regions were concatenated to represent the image. The dimensionality reduction approach proposed in this work select features using examples composed by two facial expressions. Thus, it was necessary to employ the features selection for all possible pairs of facial expression; considering the seven expressions aforementioned, 21 combinations were possible. In this context, each feature subset was used in a binary classifier which was specialized in the two facial expression described by the input subset. The classification result was assigned to the most voted facial expression by the specialized classifiers. Next, the results were compared with the traditional features selection method that uses all facial expressions as well as all features set. The features selection approach proposed in this work allows to select a smaller subset of attributes than the traditional approach and, in addition, to increase the recognition performance of the seven facial expressions considered in the experiments. Statistical tests ($p < 0.05$) showed that recognition rates achieved by the proposed method competes with the best results of literature with the advantage of low computational cost and does not require recorded fiducial points on the face images. For JAFFE, CK and TFEID datasets, 99.05%, 98.07%, and 99.63% accuracies were achieved with the presented approach, respectively. Although 21 classifiers were considered in this work, the proposed method provide a lower number of features than the traditional approach.

Keywords: facial expressions, dimensionality reduction , classification one-against-one

Capítulo 1

Introdução

Com os avanços na Visão Computacional nos últimos anos, a análise de expressões faciais humanas ganhou atenção. O reconhecimento de expressões faciais é hoje um campo de pesquisa ativo por mais de duas décadas [1].

Na conversa entre humanos, as expressões faciais formam um canal de comunicação trazendo informações importantes sobre o estado mental, emocional e físico dos indivíduos envolvidos. As expressões faciais mais simples de serem distinguidas referem-se aos sentimentos de alegria e raiva. Em uma visão mais detalhada, pode ser verificado que um comentário ou uma opinião a respeito de algo produz pequenas expressões sobre o que se está tentando transmitir. Com base nisso uma pesquisa recente mostrou que as expressões faciais podem ser utilizadas para detecção de mentiras [2].

As expressões faciais também fornecem informações relevantes sobre o comportamento de uma pessoa, sejam seus sentimentos, processos cognitivos ou comunicação social. O reconhecimento automático de expressão facial tem gerado interesses em processamento de sinais, visão computacional, reconhecimento de padrões e Interação Homem-Computador (IHC). Uma das aplicações mais importantes de reconhecimento de expressão facial é fazer a IHC mais semelhante à humana, pois computadores com a capacidade de reconhecer expressões faciais poderiam detectar e rastrear estados afetivos de um usuário e iniciar comunicações com base nessas informações, ao invés de simplesmente responder aos comandos do usuário [3].

O reconhecimento de expressões faciais consiste em identificar a emoção de um indivíduo a partir da imagem de sua face. Conforme apresentado em Zhang et al. [4], um sistema automático precisa resolver os seguintes problemas:

1. Detecção e localização da face em uma cena;
2. Extração de características da face;
3. Redução de dimensionalidade;
4. Classificação da expressão.

De maneira geral, a detecção facial é realizada com o objetivo de eliminar os objetos de fundo e processar somente a região relevante para o reconhecimento da expressão. Na extração de características são obtidas as informações para representar as expressões faciais e existem dois possíveis caminhos para isso, um deles é utilizando pontos fiduciais, que são pontos mapeados nos olhos, boca e sobrancelhas, os quais permitem identificar o formato geométrico dos elementos faciais para cada tipo de expressão. O segundo caminho é a utilização de características baseadas em textura, em que informações como rugas e vincos da pele são obtidos para representar uma expressão facial. A etapa de redução de dimensionalidade não é utilizada em todos os trabalhos, mas tem como objetivo eliminar características redundantes e que não ajudam na distinção entre as classes. Tradicionalmente para redução de dimensionalidade é procurado por um subconjunto que consiga maior discriminação entre todas as expressões faciais, neste trabalho é utilizado uma abordagem diferente, sendo baseada em pares de expressões. A redução de dimensionalidade em pares procura por um subconjunto que consiga maior distinção entre duas expressões faciais, sendo necessário encontrar um subconjunto de atributos para cada possível par. Na etapa de classificação as características ou os atributos selecionados são aplicados em algoritmos de aprendizagem de máquina para fazer predições e classificar as expressões faciais.

1.1. Motivação

Existem várias razões para que o reconhecimento de expressões seja amplamente estudado. Um trabalho muito citado na literatura, conduzido por Mehrabian [5], aponta que a comunicação humana é 7% verbal, 38% vocal e 55% expressão facial.

Segundo Rosário [6], o ser humano tem uma grande facilidade em reconhecer e distinguir expressões faciais. Muitas destas expressões têm características que as tornam

universalmente compreensíveis entre pessoas de diferentes proveniências e culturas. A expressão facial é, assim, um dos métodos mais poderosos e eficientes para partilha de emoções e intenções entre as pessoas.

Ainda, com a popularização e a intensificação da internet em quase todas as atividades cotidianas, a falta de recursos para demonstrar sentimentos deixa muito a desejar. *Emoticons* e outras formas de representações gráficas foram desenvolvidas para tentar suprir essa necessidade, no entanto, o processo é muito artificial, uma vez que o usuário deve procurar o ícone em uma lista de figuras ou utilizar atalhos. O uso de avatares que expressem emoções sem a intervenção do usuário seria muito importante para este tipo de interação.

Na medicina, as expressões faciais sempre foram objetos para estudos de doenças e síndromes. Em 1999, Hamann e Adolphs [7] investigaram a influência do reconhecimento facial em pessoas com lesão na amígdala cerebelosa. Em 2003, Sprengelmeyer et al. [8] conduziram um estudo sobre a dificuldade de pessoas com Síndrome de Parkinson em reconhecer certas emoções. Em 2005, Critchley et al. [9] apontaram algumas reações que ocorrem no corpo em função de repostas para algumas expressões faciais. Em 2008, Corcoran et al. [10] identificaram que pessoas que sofrem de Transtorno Compulsivo-Obsessivo podem desenvolver incapacidade de reconhecer alguns sentimentos. Outros vários estudos com expressões foram desenvolvidos, o que demonstra ser um assunto muito relevante para a sociedade. Ainda, as expressões faciais poderiam ser estudadas para mencionar a Escala Visual Analógica (EVA) utilizada pelos fisioterapeutas para quantificar a dor de um paciente [11].

Áreas de pesquisa como Interação Humano-Computador e a Computação Afetiva tem explorado as expressões faciais como um recurso para melhorar a comunicação entre homem e máquina, além de compreender o estado emocional do usuário, as expressões faciais contêm informações relevantes sobre comportamento humano e desempenham um papel crucial na comunicação interpessoal [12].

Os trabalhos da literatura têm apresentado diferentes propostas para reconhecer expressões faciais, normalmente os trabalhos focam nas etapas de pré-processamento da face [13], extração de características[2][14][15][16] ou classificação [17][2], sendo a redução de dimensionalidade menos explorada. Entre as propostas sugeridas está o uso de classificadores ou etapas especializadas em expressões faciais [18][2], ou seja, cada expressão facial é trabalhada individualmente para que ao final do processo o resultado obtido com cada tipo expressão facial seja combinado e forneça a saída do sistema. Diante deste cenário, o presente

trabalho propões a utilização de redução de características utilizando a abordagem Um-Contra-Um, em que são selecionados os atributos mais discriminantes para cada par de classe ou expressões faciais.

1.2. Desafios

Um dos desafios na literatura é a determinação de pontos fiduciais (Figura 1.1), que são conjuntos de pontos faciais utilizados para representação da face, geralmente localizados nos cantos externos dos olhos, sobrancelhas e boca [19]. O Face++ [20] é uma solução que consegue obter tais pontos com alta precisão e em diferentes situações, no entanto por se tratar de um produto comercial o algoritmo utilizado pela técnica não está acessível. Em muitos trabalhos esses pontos são determinados manualmente [4][17], enquanto outros tentam resolver esse problema de forma automática [21]. Determinar esses pontos é uma tarefa muito difícil devido à dinâmica da face de cada indivíduo, e também a precisão é uma peça essencial para o sucesso do reconhecimento [22]. Para superar este desafio recentemente a literatura tem focado em extratores de características baseados em textura, não necessitando conhecer a posição de elementos faciais [15].

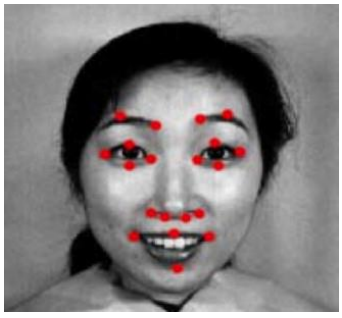


Figura 1.1: Pontos fiduciais

Trabalhar com resolução muito baixa pode ser um problema, pois a medida que a resolução diminui, a expressão vai se degradando. Existem técnicas que não conseguem operar de forma adequada quando aplicadas em imagens pequenas, com isso alguns autores evitam esse tipo de situação [23], enquanto outros procuram por soluções [24].

Em Huang et al. [25], uma das dificuldades no estudo de reconhecimento de expressões faciais é a alta dimensão dos dados gerados pelas técnicas de extração de características, sendo necessário utilizar seleção de atributos para reduzir a dimensionalidade, o que torna os métodos mais complexos. Um exemplo de alta dimensão pode ser observado no trabalho de Bashar et

al. [15], em que foram utilizados 21504 atributos. Vetores extremamente grandes aumentam o custo computacional e também levam ao problema chamado A Maldição da Dimensionalidade [25], em que quanto mais atributos são utilizados, maior é o número de instâncias necessárias para representação do modelo de aprendizagem. A consequência disso é que classificadores com muitas entradas apresentam um desempenho ruim, ou seja, como a dimensão do espaço de entrada é alta, o classificador usa quase todos os seus recursos para representar partes irrelevantes do espaço.

1.3. Objetivos

O objetivo principal deste trabalho é desenvolver um método para redução de dimensionalidade utilizando pares de emoções e classificação Um-Contra-Um no contexto do reconhecimento de expressões faciais, e os objetivos específicos são constituídos por:

- Avaliar o desempenho no reconhecimento de expressões faciais utilizando seleção de atributos em pares, seleção de atributos com todas expressões faciais e sem seleção de atributos;

- Avaliar a dimensão dos subconjuntos selecionados pela redução de atributos em pares e redução de atributos com todas as classes.

- Comparar o desempenho da classificação utilizando estratégia Um-Contra-Um com outros trabalhos da literatura;

- Verificar se o custo computacional produzido pela redução de dimensionalidade e classificação Um-Contra-Um é superior que outras propostas da literatura;

1.4. Hipóteses

As hipóteses deste trabalho são:

i. *"O reconhecimento de expressões faciais através da classificação com estratégia Um-Contra-Um consegue obter desempenho superior em relação aos trabalhos da literatura";*

ii. "*A redução de atributos e classificação Um-Contra-Um possui custo computacional maior do que os trabalhos da literatura";*

iii. *Na classificação com estratégia Um-Contra-Um, a seleção de atributos em pares de expressões faciais consegue selecionar atributos mais discriminantes levando a uma maior taxa de acerto do que a abordagem considerando todas as expressões*".

1.5. Escopo

Está no escopo deste projeto propor um método automático para o reconhecimento de expressões faciais que utiliza a seleção de atributos em pares, ao invés da seleção tradicional amplamente difundida na área. A validação é executada com os conjuntos *Cohn-Kanade Database* (CK) [26], *Japanese Female Facial Expression* (JAFFE) [27] e *Taiwanese Facial Expression Image Database* (TFEID) [28]. O procedimento será composto pela detecção da face, extração de características da imagem, redução de dimensionalidade e classificação da expressão.

Um estudo conduzido por Ekman e Friesen [29] muito citado na literatura e utilizado como base para o desenvolvimento deste trabalho afirma que as expressões faciais pertencem a seis grupos básicos: alegria, tristeza, nojo, medo, raiva e surpresa. No presente estudo, além das seis expressões básicas mencionadas, será considerada também a expressão neutro.

Os resultados obtidos com as classificações serão comparados com outras abordagens já desenvolvidas, indicando a contribuição deste trabalho para a comunidade científica.

1.6. Contribuições

A análise automática de expressões faciais pode levar a interação entre humano e máquina à uma nova modalidade tornando a comunicação mais natural e mais eficiente, pois desta forma seria possível conhecer o estado emocional de uma pessoa. Também um sistema amplamente acessível com capacidade de identificar expressões faciais pode vir a ser utilizado como uma ferramenta para a pesquisa em ciência comportamental e medicina.

Este trabalho traz como contribuição científica uma proposta para selecionar subconjuntos de atributos para o reconhecimento de expressões faciais com base na redução de características em pares de emoções. O método é diferente da abordagem tradicional utilizada na literatura e permitiu obter taxas de acerto equivalentes à menores custos computacionais. Esse fato é muito importante para a literatura, pois recentemente os extratores de características baseados em textura têm sido a atenção dos trabalhos por serem mais robustos a iluminação e exigir um menor custo computacional do que os métodos baseados em geometria, no entanto produzem vetores de características de alta dimensão. Um método eficiente para redução de

atributos pode auxiliar que trabalhos futuros consigam elevar a taxa de acerto e melhorar a eficiência computacional.

1.7. Organização do Trabalho

Neste capítulo foram apresentados a motivação para este trabalho, os desafios encontrados na literatura, os objetivos gerais e específicos, o escopo, as hipóteses de teste e as contribuições deste estudo. As descrições das técnicas utilizadas pelo método proposto são abordadas no Capítulo 2. No Capítulo 3 são apresentados recentes trabalhos desenvolvidos para reconhecer expressões faciais, são detalhas a abordagem implementada por cada um, assim como os resultados obtidos. O Capítulo 4 detalha o método proposto para reconhecer expressões a partir de imagens. Os procedimentos seguidos para realização dos experimentos são abordados no Capítulo 5 e os resultados alcançados são apresentados e discutidos no Capítulo 6. Por fim, no Capítulo 7 é feito o fechamento do trabalho comentando os resultados mais importantes.

Capítulo 2

Fundamentação Teórica

Este capítulo aborda as principais técnicas utilizadas nas propostas para reconhecer expressões faciais do capítulo seguinte. As técnicas consistem em algoritmos para localizar a face em uma imagem (Seção 2.1), extratores de características para obter informações da face (Seção 2.2), seletores de atributos para redução de dimensionalidade (Seção 2.3) e classificadores (Seção 2.4) para determinar o tipo da expressão [16]. As etapas são ilustradas na Figura 2.1 e descritas a seguir.

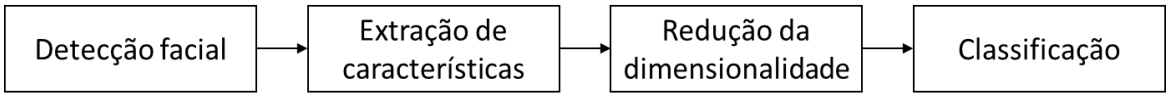


Figura 2.1: Etapas envolvidas para o REF

2.1. Detecção de Face

Normalmente os trabalhos de reconhecimento de expressões faciais necessitam localizar a face dentro de uma imagem de entrada. Este processo melhora a extração de características, uma vez que toda a informação necessária provém da face do sujeito. A seguir é detalhado a técnica que será utilizada na proposta deste trabalho.

2.1.1. Classificador em Cascata baseado em Características Haar

O Classificador em Cascada baseado em Características Haar foi proposto por Paul Viola e Michael Jones [37], e tem sido largamente utilizado para reconhecimento de objetos. Basicamente o método consiste na extração de atributos com o uso de características Haar e imagem integral, e AdaBoost's em cascata para classificação.

(a) Imagem de entrada

(b) Característica Haar para determinar a região dos olhos.

Figura 2.2: Extração de características Haar

As características Haar são compostas por retângulos utilizados para detectar regiões com diferentes intensidades [63], como bordas (Figura 2.3), linhas (Figura 2.4) e centros (Figura 2.5). Por exemplo, no método de Viola-Jones o retângulo da Figura 2.2 é utilizado para encontrar a região dos olhos e a região das bochechas. Os retângulos rotacionados não estão na proposta inicial de Viola e Jones [37], sendo adicionados posteriormente em Lienhart e Maydt [64].

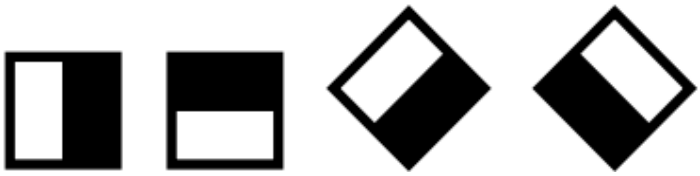


Figura 2.3: Características de borda

Figura 2.4: Características para detecção de linhas

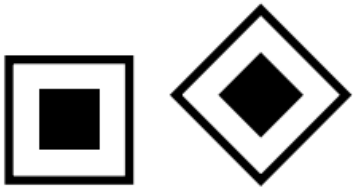


Figura 2.5: Características de centro

As características Haar podem ser localizadas em qualquer área de uma imagem através de uma janela de varredura e são determinadas pela diferença entre o somatório dos pixels da região de preto e o somatório dos pixels da região branca [63]. Então para determinar rapidamente os somatórios da janela de varredura, Viola e Jones [37] utilizaram uma representação intermediária da imagem chamada imagem integral. Considerando que a origem da imagem $(0,0)$ está localizada no canto superior esquerdo, a imagem integral ii fornece para cada posição (x, y) o somatório das intensidades dos pixels acima de y e a esquerda de x . Considerando a imagem de entrada I, a imagem integral $ii(x, y)$ é gerada pela Equação

(2.1).

$$ii(x, y) = ii(x, y-1) + ii(x-1, y) + I(x, y) \tag{2.1}$$

Usando imagem integral a soma das intensidades dos pixels de qualquer retângulo pode ser calculada com apenas quadro referências, ou seja, o somatório dos valores dos pixels em uma área (Figura 2.6) definida por a, b, c e d, pode ser feito rapidamente através da Equação (2.2).

$$\sum_{x_a, y_b} i(x_a, y_b) = ii(c) + ii(a) - ii(b) - ii(c) \tag{2.2}$$

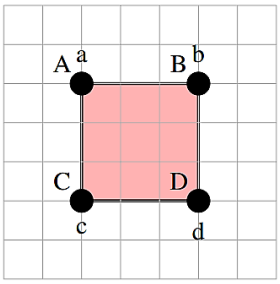


Figura 2.6: Somatório em uma imagem integral

Como as características Haar podem ser extraídas de diferentes formas e tamanhos, a dimensão do vetor de atributos é alta. Em uma imagem de 24×24 pixel são geradas aproximadamente 180,000 características. Viola e Jones [37] utilizaram uma variante do AdaBoost com classificadores fracos para selecionar características únicas que melhor separam os exemplos positivos (com face) e negativos (sem face). Para cada característica, o classificador fraco determina o limiar óptimo, de modo que o menor número de exemplos sejam

classificados incorretamente. Um classificador fraco $h_j(x)$ consiste de uma característica $f_j(x)$, um limiar θ_j e uma paridade p_j indicando a direção da inequalidade (Equação (2.3)), e x é uma região da imagem de entrada com 24×24 pixels. Caso os exemplos positivos sejam calculados abaixo do limiar, é atribuída à polaridade o valor 1. Caso contrário, é atribuída à polaridade o valor -1.

$$h_j(x) = \begin{cases} 1, & se\ p_j f_j(x) < p_j \theta_j \\ 0, & caso\ contrário \end{cases} \tag{2.3}$$

No Quadro 2.1 é apresentado o treinamento do AdaBoost. As imagens com face são rotuladas como positivas, caso contrário como negativas. As características selecionadas nas primeiras interações apresentam taxas de erro entre 0,1 e 0,3. Enquanto que as interações seguintes a tarefa de classificação torna-se mais difícil, em que as taxas de erro ficam entre 0,4 e 0,5.

ALGORITMO *construçãoAdaBoost*(x[0..n-1], y[0..n-1], T)

// Entrada: exemplos de imagens $(x_1, y_1), \dots, (x_n, y_n)$ *em que* $y_i = \{0,1\}$ *para exemplos negativos e exemplos positivos respectivamente.* T *é o número de classificadores fracos*

// Saída: classificador forte $h(x)$

// Inicializar os pesos $w_{1,i}$ *para* $y_i = \{0,1\}$ *respectivamente, em que* m *e* l *são o número de exemplos negativos e positivos respectivamente.*

$$w_{1,i} \leftarrow \frac{1}{2m}, \frac{1}{2l}$$

while t < T **do**

// Normalizar os pesos,

$$w_{t,i} \leftarrow \frac{w_{w,i}}{\sum_{j=1}^{n} w_{w,j}}$$

// Para cada característica j*, treinar um classificador* h_j *restrito ao uso de uma única característica. O erro é avaliado de acordo com* w_t

$$\epsilon_j \leftarrow \sum_i w_i |h_j(x_i) - y_i|$$

// Seleciona o classificador h_t *com menor erro* ϵ_t

$$\underset{h}{arg\ min} \quad \epsilon$$

// Atualiza os pesos. $e_i \leftarrow 0$ *se o exemplo é classificado corretamente e* $e_i \leftarrow 1$ *caso contrários, e* $\beta_t \leftarrow \frac{\epsilon_t}{1-\epsilon_t}$

$$w_{t+1,i} \leftarrow w_{t,i} \beta_t^{1-e_i}$$

end while

// determina o classificador forte final, em que $\alpha_t = log \frac{1}{\beta_t}$.

return $h(x) \leftarrow f(x) \leftarrow \begin{cases} 1, & \sum_{t=1}^{T} \alpha_t h_t(x) \geq \frac{1}{2} \sum_{t=1}^{T} \alpha_t \\ 0, & caso\ contrário \end{cases}$

Quadro 2.1: Pseudocódigo para construção do AdaBoost. A cada iteração uma característica é selecionada.

A partir do AdaBoost proposto é gerado uma cascata de classificadores para realizar a detecção da face. Como ilustrado na Figura 2.7, a cascata de classificadores é composto de vários estágios, cada um contendo um AdaBoost. Um resultado positivo no estágio $AdaBoost_1$ leva ao segundo estágio $AdaBoost_2$. Um resultado positivo no estágio $AdaBoost_2$ desencadeia a avaliação da imagem no $AdaBoost_3$, e assim por diante. Um resultado negativo em qualquer ponto leva à rejeição imediata da região x.

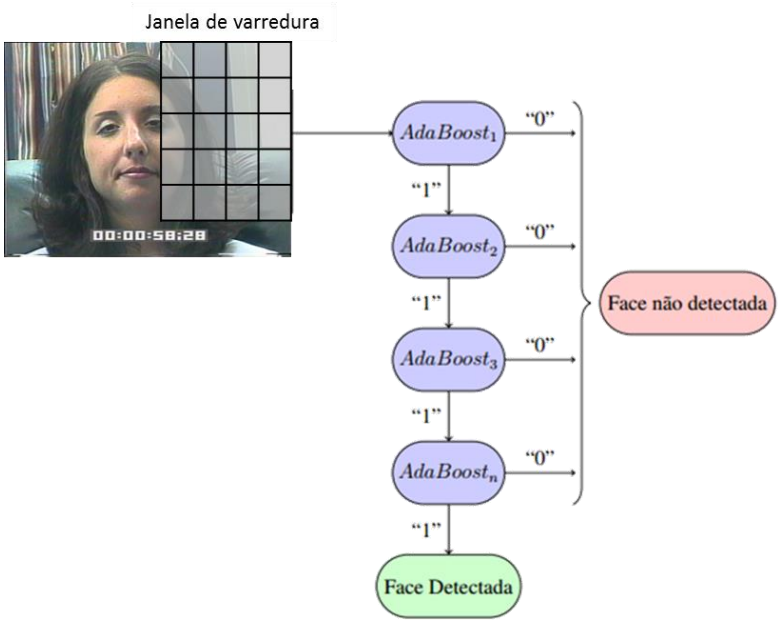


Figura 2.7: Classificador em cascata para detecção da face

Os estágios da cascata são construídos de tal forma que a etapa seguinte seja mais específica e complexa que a anterior. Isso permite que o algoritmo descarte rapidamente regiões que não possuam as características de interesse, como objetos no fundo da imagem. Somente faces e elementos semelhantes são processados pelos estágios seguintes que são mais complexos e necessitam mais tempo de processamento. Desta forma, para que seja considerada que há uma face na região de varredura x, todas os estágios da cascata devem produzir resultados positivos.

Em cada estágio de classificação são adicionadas características até que a taxa de detecção e taxa de falsos negativos sejam satisfeitas por uma condição pré-estabelecida. As taxas são determinadas testando o detector sobre um conjunto de validação. Na proposta de

Viola e Jones [37] foram utilizados 68 estágios em cascata com aproximadamente 6000 características.

Como não é conhecido o tamanho da face em uma imagem, um dos parâmetros a serem definidos é tamanho mínimo da janela de varredura, assim como seu fator de escala para ser redimensionado até o tamanho máximo preestabelecido.

2.2. Extração de Características

Os algoritmos de extração de características utilizados permitem obter informação de textura da face. Essas informações são utilizadas por um algoritmo de aprendizagem para determinar a emoção da face. Conforme apresentado na seção 3.2, devido ao desempenho e custo computacional, foram utilizadas as técnicas LBP e WLD para extração de características, que são descritos a seguir.

2.2.1. Local Binary Pattern

Um dos extratores utilizados para obter informação das faces é o LBP. Este é um algoritmo rápido e eficiente, tendo sido aplicado em sistemas de tempo real [14]. O LBP é um extrator de textura local e foi introduzido por Ojala et al. [65] como um método invariante de escala de níveis de cinza para análise de textura. Esta técnica foi mais tarde aplicada no reconhecimento de expressões faciais [17][14][45] e tem alcançando até 99% de acerto. O $LBP_{P,R}$ consiste em selecionar os P vizinhos ao redor de cada pixel em um raio de operação R e gerar um código binário de 2^P bits, através da diferença de intensidade dos pixels vizinhos com o respectivo centro. O código binário para uma dada imagem i é obtido com o uso da Equação (2.4).

$$LBP_{P,R}(x_c, y_c) = \sum_{p=1}^{P} s(i_c - i_p),\ s(x) = \begin{cases} 1 & x > 0 \\ 0 & caso\ contrário \end{cases} \tag{2.4}$$

Na Equação (2.4), i_c denota o nível de cinza do elemento central (x_c, y_c) e i_p corresponde ao nível de cinza dos pixels vizinhos. A Figura 2.8 ilustra um exemplo de geração de código binário que produz o valor 00111110.

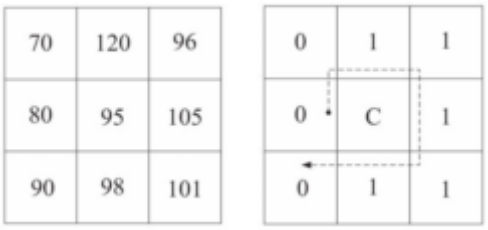


Figura 2.8: Exemplo de geração de código binário [15]

Depois de aplicado o método em todos os pixels da imagem, um histograma de 2^P *bins* é construído a partir dos códigos gerados. Este histograma contém informações sobre a distribuição dos micro padrões locais, tal como bordas, pontos e superfícies (Figura 2.9) ao longo de toda a imagem, de modo que possa ser utilizado estatisticamente para descrever a imagem [45].

Figura 2.9: Exemplos de arestas detectadas com o LBP [14]

De acordo com Shan et al. [45] um caminho interessante para obter melhores resultados é dividir a face em pequenas zonas para então extrair os histogramas. Em seguida as características de cada zona são concatenadas para formar um único vetor

No LBP as características são representadas em um histograma, o que pode resultar na perda de informação espacial. Utilizando LBP Uniforme com 8 vizinhos, existem 59 *bins* para representar padrões, à medida que mais padrões são obtidos, o histograma passa a degradar as informações já obtidas para comportar os novos padrões. Para reter a informação espacial é necessário dividir a face em vários pequenos blocos, em seguida, o histograma LBP de cada pequena região pode ser encontrado, e, finalmente, todas estas características são concatenados para formar o descritor global da face [66] (Figura 2.10). Uma imagem dividida em 7×6 zonas, produzirá um vetor de características de $7 \times 6 \times 256 = 10752$ dimensões, considerando que $P = 8$.

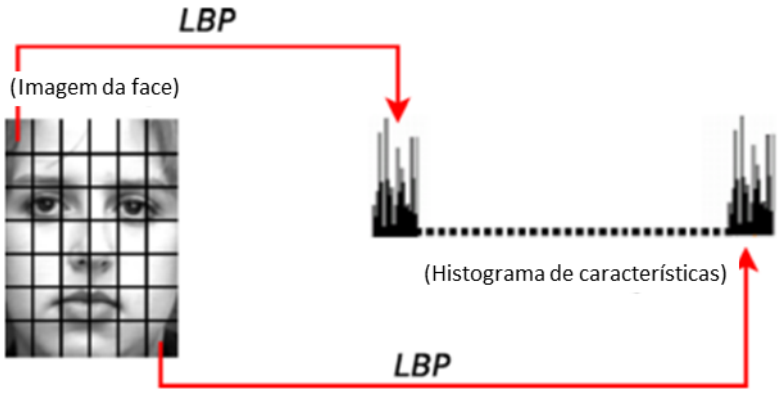


Figura 2.10: Características extraídas com LBP em uma face zoneada [45]

O operador LBP produz 2^P valores de saída correspondentes aos diferentes padrões binários que podem ser formado pelos *P* pixel no grupo da vizinhança. No entanto, certos *bins* do histograma contém mais informações do que outros, portanto, é possível utilizar apenas um subconjunto dos 2^P *bins*. Esses padrões fundamentais são conhecidos como padrões uniformes. Um LBP é chamado uniforme se ele contém no máximo duas transições de '0' para '1' ou vice-versa. Por exemplo, 00000000 (0 transições), 001110000 (2 transições) e 11100001 (2 transições) são padrões uniformes. Um padrão uniforme representa quase 90% de todos os padrões com $LBP_{(8,1)}$ e cerca de 70% para $LBP_{(16,2)}$ [54]. Adicionando todos os padrões que tem mais do que duas transições (não uniformes) em um único *bin* produz um operador LBP chamado $LBP_{P,R}^{u2}$ com menos de 2^P *bins* [67][68]. Por exemplo, o número de *bins* para uma vizinhança de 8 pixels é de 256 para o LBP padrão e 59 para $LBP_{8,1}^{u2}$. Pela questão da alta dimensionalidade este trabalho utiliza do $LBP_{P,R}^{u2}$.

2.2.2. Weber Local Descriptor

Este extrator de características foi proposto por Chen et al. [56] e tem sido aplicado recentemente com sucesso no reconhecimento de expressões faciais [53][16][2]. O método consiste em descrever as características de textura de uma imagem a partir de duas componentes: *diferencial de excitação* e *orientação* (Figura 2.11). Primeiramente, o WLD calcula micro padrões através da *diferencial de excitação*, e constrói estatísticas sobre esses padrões junto com a *orientação* do gradiente.

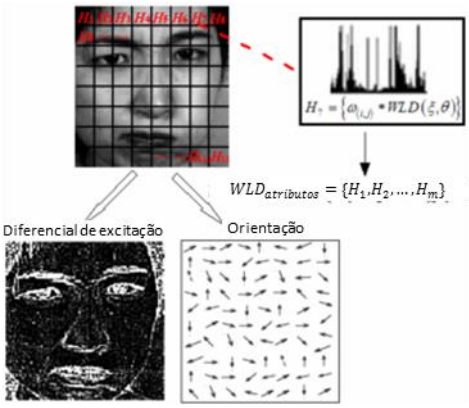


Figura 2.11: Extração e representação de características geradas pelo WLD [16]

Assim como o LBP este método utiliza os P vizinhos em um raio R do pixel central para extrair o *diferencial de excitação* da expressão facial da imagem. O *diferencial de excitação* $\xi(x_c)$ de cada pixel x_c é dada pela razão de V_1 e V_2, em que V_1 e V_2 são saídas dos filtros f_1 e f_2 respectivamente como é ilustrado na Figura 2.12.

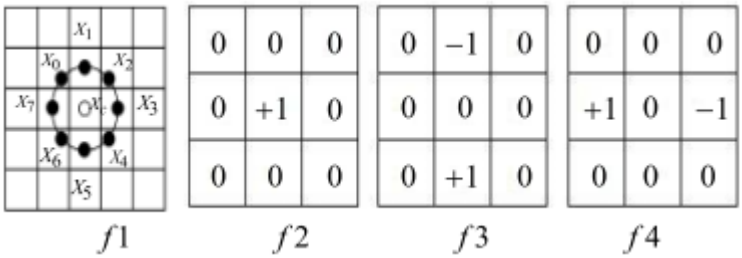


Figura 2.12: Janela de filtros do WLD [53]

O valor de V_1 é calculado pela soma das diferenças entre o pixel x_c e seus vizinhos x_i $(i = 0,1, \ldots, p-1)$, e V_2 é o valor de x_c. É possível calcular V_1 e V_2 através da Equação (2.5).

$$\begin{cases} V_1 = \sum_{i=0}^{p-1} (x_i - x_c) \\ V_2 = x_c \end{cases} \tag{2.5}$$

Sendo p o número de pixels num raio R do centro X_c. O *diferencial de excitação* é mapeada para $[-\frac{\pi}{2}, \frac{\pi}{2}]$ com a Equação (2.6).

$$\xi(x_c) = \arctan(G_1) = \arctan(\frac{V_1}{V_2}) \tag{2.6}$$

A *orientação* é a segunda componente utilizada pelo método e representa a orientação do gradiente. O cálculo é apresentado na Equação (2.7).

$$\begin{cases} V_3 = x_5 - x_1 \\ V_4 = x_7 - x_3 \\ \theta(x_c) = \arctan(\frac{V_3}{V_4}) \end{cases} \tag{2.7}$$

Os valores de V_3 e V_4 são resultados dos filtros f_3 e f_4 da Figura 2.12. θ é limitado em $\left[-\frac{\pi}{2}, \frac{\pi}{2}\right]$. Conforme os valores de V_3 e V_4, θ pode ser restringido para $\theta' \epsilon\ [0,2\pi]$ pela Equação (2.8).

$$\theta' = \begin{cases} \theta & V_3 < 0, V_4 < 0 \\ \theta + \pi & V_3 >, V_4 > 0 \\ \theta + \pi & V_3 < 0, V_4 > 0 \\ \theta + 2\pi & V_3 > 0, V_4 < 0 \end{cases} \tag{2.8}$$

θ' é ainda linearmente quantificado em *T orientações* pela Equação (2.9). Nos trabalhos [53][16] foram utilizados $T = 8$, assim a orientação é $\Phi_t = \frac{t\pi}{4}, (t = 0,1, \dots, 7)$, ou seja, as *orientações* são escaladas no intervalo $[\Phi_t - \frac{\pi}{8}, \Phi_t + \frac{\pi}{8}]$.

$$\Phi_t = f_q(\theta') = \frac{2t}{T}\pi, \qquad t = mod\left(\left[\frac{\theta'}{\frac{2\pi}{T}} + \frac{1}{2}\right], T\right) \tag{2.9}$$

Após calcular o *diferencial de excitação* $\xi(x_c)$ e a *orientação* ϕ_t de cada pixel, é gerado o histograma $WLD(\xi_j, \phi_t), (j = 0,1, \dots, N-1, t = 0,1, \dots, T-1)$, em que N é a dimensionalidade da imagem e T é o número das orientações dominantes. Este histograma possui $w = T \times C$ dimensões, em que C é o número de células em cada orientação, ou seja, cada coluna corresponde a uma orientação dominante ϕ_t e cada linha corresponde a um

histograma de *diferencial de excitação* com C posições (*bins*). Assim, a intensidade de cada célula corresponde às frequências de um certo intervalo de *diferencial de excitação* para uma *orientação* dominante (Figura 2.13). Ao final o histograma é convertido em um vetor de tamanho w.

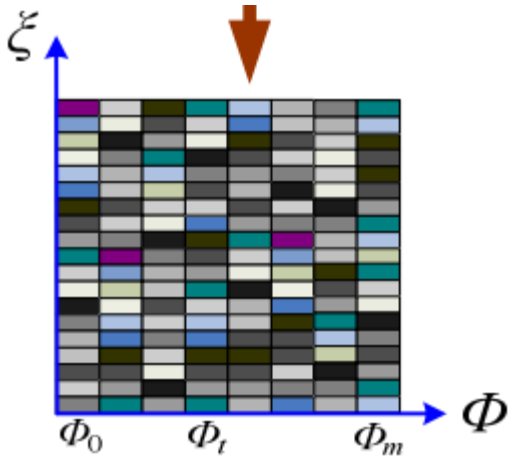


Figura 2.13: Histograma gerado por WLD [56]

A partir do histograma $WLD(\xi_j, \phi_t)$ pode ser gerado um histograma H com características mais discriminantes. Entretanto, assim como em Liu et al. [53], este trabalho não prosseguiu com a abordagem, pois seria introduzido o parâmetro M e encontrar o valor óptimo consumiria mais tempo nos experimentos, além do tempo de cálculo do algoritmo ser maior. No método utilizado por Liu et al. [53] foi alcançado uma taxa de acerto de 96% sobre o conjunto de dados JAFFE.

Para obter um descritor mais discriminante, o histograma $WLD(\xi_j, \phi_t)$ pode ser codificado em um histograma unidimensional H. Dado um histograma $WLD(\xi_j, \phi_t)$ de uma imagem, como mostrado na Figura 2.14, cada coluna é projetada para formar um vetor de uma dimensão $H(t)$, $(t = 0,1, \dots, T-1)$. Assim, os *diferenciais de excitação* ξ_j são reagrupadas em T sub-histogramas $H(t)$ e cada sub-histograma $H(t)$ corresponde a uma *orientação dominante* θ_t. Posteriormente cada $H(t)$ é dividido em M segmentos. Todos os segmentos de $H_{m,t}$ formam uma matriz, em que cada coluna corresponde a uma *orientação dominante* e cada linha corresponde a um *diferencial de excitação*. Então a matriz formada pelos segmentos de $H_{m,t}$ são reorganizados como um histograma unidimensional, em que cada linha é concatenada

produzindo um histograma H_m. Concatenando os resultados dos M sub-histogramas é obtido o histograma $H = \{H_m\}, m = 0,1, \dots, M-1$.

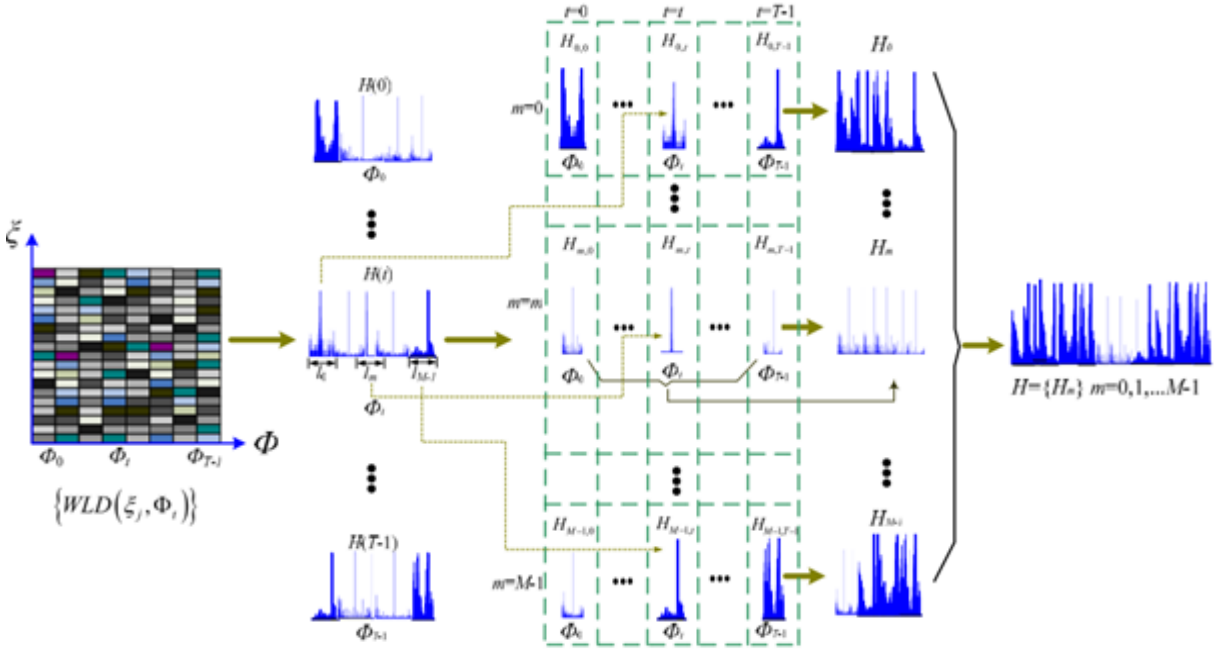


Figura 2.14: Ilustração de um histograma WLD [56]

2.3. Redução de Dimensionalidade

A redução de dimensionalidade, ou seleção de atributos, é definida como um método para selecionar um subconjunto de atributos a partir de uma lista de atributos candidatos que podem ser usados com melhor precisão por algoritmos de aprendizagem de máquina [69]. Os objetivos da seleção de atributos é evitar o *overfitting* e para melhorar a classificação, reduzir a dimensionalidade dos dados de entrada para o algoritmo de aprendizagem [70]. As abordagens para redução de dimensionalidade podem ser dividas em três categorias baseado em como a seleção de características é combinada com o modelo de classificação. As categorias são classificadas em *filter*, *wrapper* e *embedded*, sendo que neste trabalho apenas serão utilizadas as duas primeiras.

Na proposta deste trabalho a extração de características com o LBP e WLD produzem vetores com muitas características, no entanto, nem todas elas podem ser relevantes para discriminar as diferentes expressões faciais. Atributos irrelevantes podem prejudicar o desempenho da classificação e por isso algumas abordagens e técnicas são aplicadas para reduzir este problema. Como descrito na seção 3.3 grande parte dos trabalhos tem utilizado a

PCA para redução de atributos, com esta técnica os atributos são transformados para um novo espaço, enquanto que as técnicas para redução de dimensionalidade utilizadas neste trabalho, tal como IG (*Information Gain*), CFS (*Correlation-based Feature Selection*) e KW (*Kruskall Wallis*), mantém as variáveis originais, selecionando apenas um subconjunto destas. A seguir as estratégias *wrapper* e *filter* são descritas, assim como as técnicas IG, CFS e KW.

2.3.1. Wrapper

O *wrapper* é uma estratégia utilizada para seleção de características. Neste tipo de estratégia os subconjuntos de atributos selecionados são avaliados em um determinado algoritmo de aprendizagem de máquina, tal como NN, SVM e Redes Bayesianas[71]. Isso significa que o classificador é treinado repetidamente com os dados de treinamento para cada subconjunto selecionado (Figura 2.15). Este procedimento é computacionalmente muito caro. Diferentes estratégias de busca como *sequential backward elimination* (SBE) [72], *sequential forward elimination* (SFE) [72] ou buscas bidirecionais são utilizados para selecionar subconjuntos de atributos.

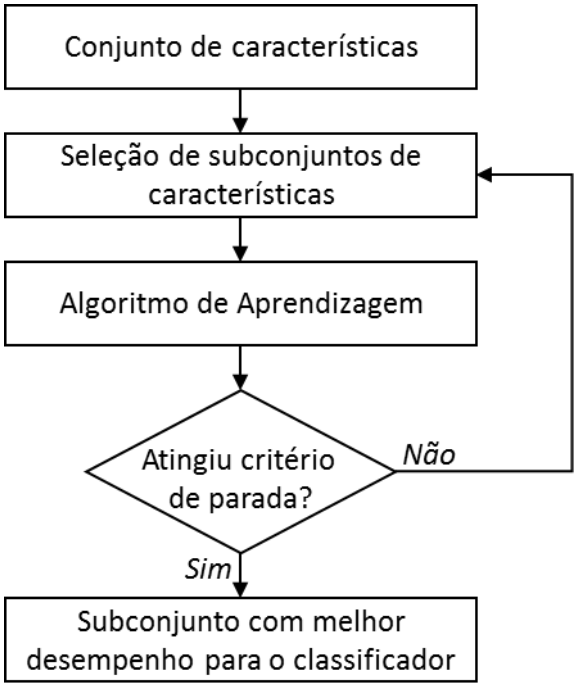


Figura 2.15: Fluxograma da abordagem *wrapper*

As estratégias de busca SBE e o SFE são gulosas, ou seja, sempre realizam escolhas que parecem ser a melhor no momento [73]. O algoritmo SFE sempre inicia com um conjunto vazio

de atributos. Na primeira iteração, o algoritmo considera todos os subconjuntos contendo apenas uma característica. O subconjunto com precisão mais elevada é usado como a base para a iteração seguinte. Em cada iteração, o SFE adiciona ao subconjunto um atributo não previamente selecionado e retém um subconjunto de atributos que resulta no desempenho máximo do classificador. A busca termina após a precisão do subconjunto selecionado não ser melhorada pela adição de qualquer outra característica. O SBE funciona de forma análoga, a partir de um subconjunto contendo todos os atributos o objetivo é eliminar as características que aumenta a precisão do classificador.

Os subconjuntos de atributos selecionados com a abordagem *wrapper* são avaliados pela acurácia preditiva do classificador treinado, portanto, são mais significativas do que a abordagem *filter*, que será descrito a seguir e mede somente a redundância ou relevância dos atributos.

2.3.2. Filter

O método *filter* consiste em avaliadores de atributos e métodos de busca para ranquear as características de um conjunto de dados. Todas as características são ranqueadas com base em seu peso (e.g. informação, correlação ou outra métrica). Os atributos com postos mais baixos são removidos a partir de uma condição ou um limiar e, em seguida, o algoritmo de aprendizagem de máquina é usado para testar a precisão do modelo [74]. A Figura 2.16 representa o fluxo do *filter*. Existem vários métodos para avaliar os atributos, entre eles podem ser citados a seleção de atributos baseada em correlação, ReliefF, Kruskal Wallis, os métodos baseados em entropia, ganho de informação, informação mútua, e incerteza simétrica. O tempo de processamento com *filter* é mais rápido do que usando *wrappers*, mas a precisão da classificação do *filter* não é tão boa como *wrapper* [71].

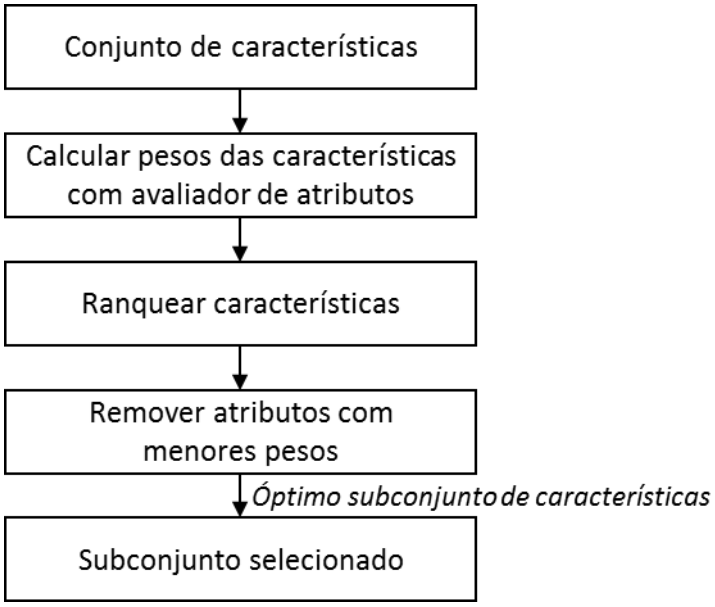


Figura 2.16: Fluxograma da abordagem *filter*

2.3.2.1. Ganho de Informação

Ganho de Informação é um método popular para seleção de características e normalmente consegue produzir bons resultados [75][76]. IG é usado como uma medida de significância baseada na entropia. Para seleção de atributos o IG procura por atributos que possuem mais informação da classe [76].

A entropia, é um conceito muito importante na Teoria da Informação, expressando a uniformidade de distribuição de qualquer tipo de energia no espaço. Quanto mais uniforme a distribuição de energia, maior a incerteza e maior a entropia. A entropia é usada no processamento de informação por Shannon, que propôs o conceito de Entropia da Informação. Este conceito é uma medida quantitativa de informação e pode medir o grau de incerteza de uma variável aleatória [77].

O valor do ganho de informação contém a contribuição do atributo no conjunto de características. Assim a seleção de atributos depende do valor de ganho de informação da característica. Normalmente é utilizado um limiar para escolher um atributo, em que somente as características que possuem ganho de informação acima do limiar estabelecido são selecionadas para compor o vetor de atributos final.

O ganho de informação para um dado atributo t_k em relação à classe c_i é a redução da incerteza sobre o valor de c_i quando se conhece o valor de t_k. O Ganho de Informação de um atributo t_k para a classe c_i é obtido pela Equação (2.10).

$$IG(t_k, c_i) = \sum_{C \in ci} \sum_{t \in t_k} p(t,c) \log \frac{p(t,c)}{p(t)p(c)} \quad (2.10)$$

Sendo $p(c)$ a probabilidade de ocorrer a classe c e $p(t,c)$ a probabilidade de instâncias de classe c conter o atributo t. A probabilidade de t ocorrer é dado por $p(t)$. Quanto maior o ganho de informação de um atributo, maior é a sua importância para a classificação.

2.3.2.2. Kruskal Wallis

Alguns trabalhos tem utilizado a seleção de atributos com o Método de Kruskal Wallis (KW) para reconhecer expressões faciais [2], faces e raça [78], o qual é muito simples de implementar e envolve baixo processamento.

O Método de KW é um teste não-paramétrico baseado em ANOVA (Análise de Variância) aplicado em duas ou mais classes. A hipótese nula do teste consiste em verificar se as amostras de dois ou mais grupos tem mediana iguais e retorna p (nível descritivo). Se p é próximo de 0, então as características são mais discriminantes. O valor de p pode ser calculado através da Equação (2.11) [16].

$$p = \frac{12}{N(N-1)} \sum_{j=1}^{k} \frac{R_j^2}{n_j} - 3(N+1) \quad (2.11)$$

Em que N é número total de observações em todos os grupos, n_j é o número de observações no grupo j, R_j é a soma do rank do grupo j e k é o número de observações para um grupo. As observações correspondem ao número de exemplos ou total de faces utilizadas, enquanto que os grupos são as classes ou as expressões faciais. A partir dos valores p de cada um dos atributos, são selecionadas as características mais discriminantes ou são descartadas as características com p maior que um limiar.

2.3.2.3. Seleção de Atributos baseada em Correlação

Um método bastante popular utilizado na literatura de aprendizagem de máquina para selecionar atributos é a Seleção baseada em Correlação (*Correlation-based Feature Selection*

– CFS) [79]. Esta técnica foi desenvolvida por Hall [80] e usa heurística baseada em correlação para avaliar os atributos que podem ser úteis. A hipótese da qual a heurística está baseada é que "*um bom subconjunto contém atributos altamente relacionados com a classe, e não correlacionados um com os outros*", assim é preferível um subconjunto tenha um alto valor de mérito heurístico. A hipótese é definida matematicamente pela Equação (2.12).

$$Merit_s = \frac{k\overline{r_{cf}}}{\sqrt{k + k(k-1)\overline{r_{ff}}}} \tag{2.12}$$

Em que $Merit_s$ é o mérito heurístico de um subconjunto S contendo k atributos, $\overline{r_{cf}}$ a média de correlação entre atributos e classe, e $\overline{r_{ff}}$ a média da intercorrelação entre os atributos.

O propósito da seleção de atributos é decidir quais das características iniciais podem ser incluídas e quais podem ser ignoradas. Se existem n características inicialmente, então existem 2^n possíveis subconjuntos. O caminho para encontrar o melhor subconjunto deveria tentar todos eles, mas dependendo do valor n isto é inviável.

Então, o CFS primeiramente calcula uma matriz de correlação de atributo-atributo e atributo-classe dos dados de treinamento e então utiliza o *best first search* para encontrar um espaço de características. O *best first search* é um algoritmo de busca que explora grafos, expandindo o nó mais promissor escolhido de acordo com a regra especificada. O algoritmo inicia com um conjunto de características vazio e explora por novos subconjuntos de atributos fazendo alterações no conjunto atual. Quando as escolhas locais parecem ser menos promissoras, a busca retorna e o próximo melhor subconjunto não explorado é selecionado para avaliação dando continuidade ao processo. O critério de parada é atingido quando a expansão consecutiva de um número preestabelecido de subconjuntos, normalmente 5 [81], não produz nenhuma melhoria

2.4. Classificação

Após extraídas as características da região de interesse da imagem e selecionadas as mais discriminantes, estas características são aplicadas em algoritmos de Aprendizagem de Máquina para aprender o modelo de classificação das expressões faciais e fazer predições. A seguir são abordados os princípios básicos dos classificadores que foram utilizados neste estudo, para maior profundidade dos algoritmos consultar Vapnik [82] e Fix e Hodges [83].

2.4.1. k-Nearest Neighbor

Este algoritmo consiste em classificar um novo exemplo atribuindo a ele o rótulo representado mais frequentemente dentre as k amostras mais próximas e utilizando um esquema baseado em votação [84].

Para determinar a classe de um elemento que não pertença ao conjunto de treinamento, o classificador kNN procura k elementos do conjunto de treinamento que estejam mais próximos deste elemento desconhecido, ou seja, que tenham a menor distância no espaço de atributos. Estes k elementos são chamados de *k*-vizinhos mais próximos. A classe mais predominante dos *k* vizinhos será atribuída ao elemento desconhecido.

O treinamento deste algoritmo apenas retém as instâncias de treinamento. Quando uma nova instância deve ser classificada, as medidas de distância são realizadas com os exemplos do treinamento. As k menores distâncias são utilizadas para fazer a predição, sendo que k é o único parâmetro de ajuste.

Existem várias métricas para determinar a distância entre duas instâncias [85]. Seja X uma instância aleatória descrita pelo vetor de características $X = [a_1(X), a_2(X), a_3(X), \dots, a_n(X)]$, em que a_r é o valor do *r-ésimo* atributo de X e n é a dimensão do vetor, as seguintes distâncias podem ser calculadas:

- *Distância Euclidiana*: é a métrica mais popular e calcula a raiz quadrada das somas de diferenças do vetor de atributos de um par de instâncias (Equação (2.13)).

$$Dist(X_i, X_j) = \sqrt{\sum_{r=1}^{n} \left(a_r(X_i) - a_r(X_j)\right)^2} \tag{2.13}$$

- *Distância Manhattan*: calcula as diferenças absolutas entre duas instâncias (2.14)).

$$Dist(X_i, X_j) = \sum_{r=1}^{n} \left|a_r(X_i) - a_r(X_j)\right| \tag{2.14}$$

- *Distância Chebychev*: é também conhecida como valor de distância máxima e é determinada como o máximo valor absoluto da diferença entre um par de instâncias (Equação (2.15))

$$Dist(X_i, X_j) = \max_{r=1}^{n} |a_r(X_i) - a_r(X_j)| \quad (2.15)$$

- *Distância Minkowski*: esta distância (Equação (2.16)) é a generalização das distâncias anteriores. Quando $P = 1$, esta distância representa a distância de Manhattan e quando $P = 2$, a distância Euclidiana. A distância Chebyshev também é uma variante quando $P = \infty$ (por limite). Esta distância pode ser usada para variáveis ordinais ou quantitativas.

$$Dist(X_i, X_j) = \left(\sum_{r=1}^{n} |a_r(X_i) - a_r(X_j)|^{\frac{1}{P}} \right)^P \quad (2.16)$$

Na Figura 2.17 é apresentado um exemplo de classificação, sendo considerado que existem duas classes e $k = 3$. O novo elemento a ser classificado é simbolizado pelo triângulo.

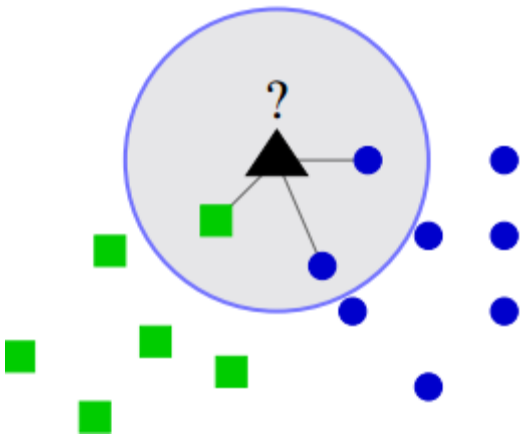


Figura 2.17: Exemplo de classificação com kNN

Pode-se observar que os 3 vizinhos mais próximos ($k = 3$) são dois círculos e um quadrado. Como o número de círculos são mais predominantes que quadrados, o novo elemento será rotulado como círculo.

Devido à possibilidade de haver um grande número de exemplos de treinamento para calcular a distância, esse algoritmo pode apresentar um tempo elevado para classificação [86].

2.4.2. Support Vector Machine

Outro classificador bastante utilizado na literatura é a *Support Vector Machine* (SVM), que consiste em um método de classificação de padrões baseado na teoria de aprendizagem estatística. Em problemas de classificação binária, a SVM não só distingue ambas as classes, mas também encontra a melhor linha de separação para fazer a maior margem entre duas classes [41].

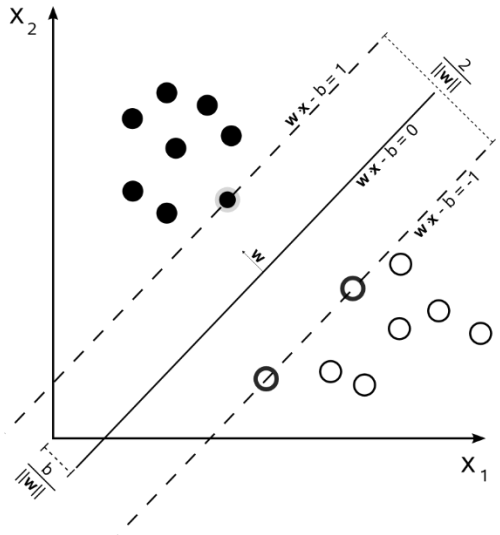


Figura 2.18: SVM com margens rígidas[1].

A SVM com margens rígidas (Figura 2.18) é o modelo mais simples de SVM e somente pode ser utilizada em dados linearmente separáveis. A equação de um hiperplano é apresentada na Equação (2.17), em que $w \cdot x$ é o produto escalar entre os vetores w e x, w é o vetor normal ao hiperplano descrito, o qual deve ser ajustado, x é um vetor de entrada e $\frac{b}{\|w\|}$ corresponde à distância do hiperplano em relação à origem, com $b \in \Re$ [87].

$$f(x) = w \cdot x + b = 0 \qquad 2.17$$

[1] Imagem retirada de http://en.wikipedia.org/wiki/File:Svm_max_sep_hyperplane_with_margin.png

A margem que maximiza a separação de duas classes é representada por $\frac{2}{\|w\|}$ e o hiperplano óptimo $w \cdot x + b = 0$ pode ser obtido pela minimização da Equação (2.18), em que $y_i = \{+1, -1\}$ representa a classe do respectivo padrão x_i, em que i é o *i*-ésimo exemplo do conjunto de treinamento.

$$\begin{matrix} minimizar \\ w, b \end{matrix} \quad \frac{1}{2}\|w\|^2 \qquad com\ as\ restrições: y_i(w \cdot x_i + b) \geq 0 \tag{2.18}$$

A otimização da Equação (2.18) é realizada com a introdução de uma função Lagrangiana, desta forma é obtido a forma dual apresentada na Equação (2.19) e que ser resolvida mais facilmente [88], sendo α_i os multiplicadores de Lagrange e n o número de exemplos no conjunto de treinamento.

$$\begin{matrix} maximizar \\ \propto \end{matrix} \quad \sum_{i=1}^{n} \alpha_i y_i - \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} \alpha_i \alpha_j y_i y_j (x_i \cdot x_j)$$
$$com\ restrições: \begin{cases} \alpha_i \geq 0, i = 1, \dots, n \\ \sum_{i=1}^{n} \alpha_i y_i = 0 \end{cases} \tag{2.19}$$

O vetor w pode ser determinado a partir dos subconjunto de treinamento com multiplicadores de Lagrange α_i (Equação (2.20)). De acordo com a condição de Karush-Kuhn-Tucker, o subconjunto com α_i diferente de zero corresponde aos Vetores de Suporte (*Vector Support* – SV), que são utilizados para gerar o hiperplano de separação. Com dados linearmente separáveis todos os SV estão na margem e então o número de SV é muito pequeno. Consequentemente o hiperplano é obtido por um pequeno subconjunto dos dados de treinamento, e os outros pontos podem ser removidos que o resultado é o mesmo.

$$w = \sum_{i=1}^{n} \alpha_i x_i y_i \tag{2.20}$$

Então a função de decisão do classificador é obtido obtido a partir da Equação (2.21)

$$f(x) = sgn\left(\sum_{i=1}^{n} y_i \alpha_i (x \cdot x_i) + b\right) \tag{2.21}$$

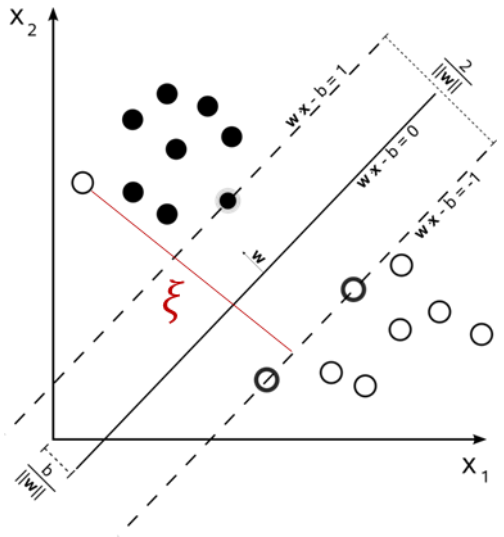


Figura 2.19: SVM com margens suaves.

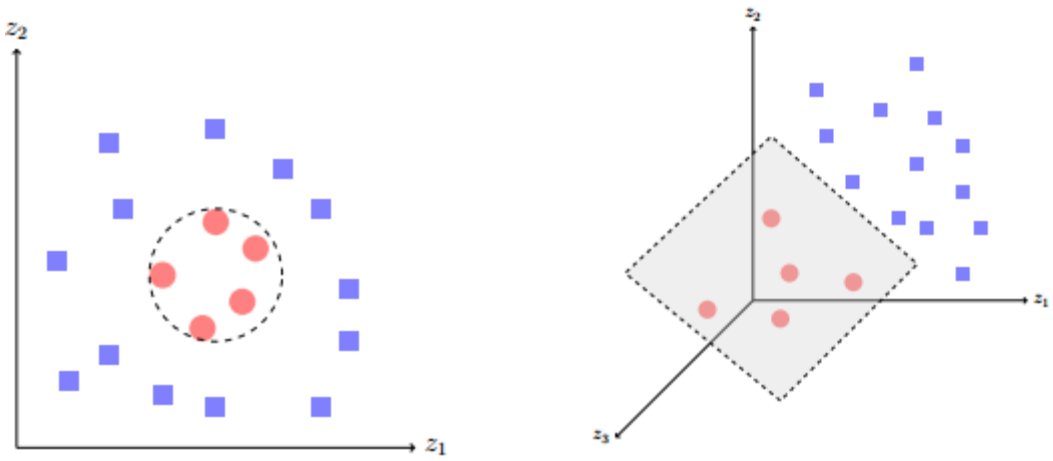
Para dados não são linearmente separáveis ou que apresentam ruídos é utilizada a SVM com margens suaves (Figura 2.19), que é uma adaptação da SVM com margens rígidas introduzindo as variáveis de folga ξ_i. Isso permite que alguns dados possam violar a restrição da Equação (2.18). Assim a nova função a ser minimizada para obter a margem de separação passa a ser definida pela Equação (2.22). Esta equação é resolvida através de sua forma dual da Equação(2.23)

$$\begin{aligned} &\underset{w,b,\xi}{minimizar} \quad \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{n} \xi_i \\ &com\ as\ restrições: y_i(w \cdot x_i + b) \geq 1 - \xi_i \end{aligned} \tag{2.22}$$

$$\begin{aligned} &\underset{\propto}{maximizar} \quad \sum_{i=1}^{n} \alpha_i y_i - \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} \alpha_i \alpha_j y_i y_j (x_i \cdot x_j) \\ &com\ restrições: \begin{cases} 0 \leq \alpha_i \leq C, i = 1, \dots, n \\ \sum_{i=1}^{n} \alpha_i y_i = 0 \end{cases} \end{aligned} \tag{2.23}$$

O parâmetro C é definido pelo usuário e atua como uma função de penalidade prevenindo que ruídos afetem o hiperplano óptimo. Um C maior corresponde a assumir uma penalidade maior para os erros.

As SVMs são eficazes na classificação de conjuntos de dados linearmente separáveis ou que possuam uma distribuição aproximadamente linear, como apresentado na Figura 2.18 e na Figura 2.19. No entanto, há muitos casos em que não é possível dividir satisfatoriamente os dados por um hiperplano (Figura 2.20(a)). Quando isso ocorre é mapeado o conjunto de treinamento de seu espaço original, referenciado como entrada, para um novo espaço de maior dimensão, denominado de espaço de características, facilitando a separação dos dados por meio de uma SVM linear (Figura 2.20(b)) [84]. A motivação deste procedimento é dada pelo teorema de Cover.



(a) Conjunto não linearmente separável

(b) Transformação em um espaço linearmente separável

Figura 2.20: Exemplo de conjuntos não linearmente separáveis

Para um conjunto de dados não linear no espaço de entradas X, o teorema de Cover estabelece que X pode ser transformado em um espaço de características F, no qual com alta probabilidade os dados são linearmente separáveis. Para isso duas condições devem ser satisfeitas. A primeira é que a transformação seja não linear, enquanto a segunda é que a dimensão do espaço de características seja suficientemente alta. No entanto esse procedimento não garante que os dados se tornem linearmente separáveis, sendo necessário utilizar a SVM com margens suaves para classificar os dados.

O mapeamento para um novo espaço de características não linear e de alta dimensão é feito com o uso de uma função de *kernel*. O desempenho do classificador SVM é dependente da escolha de uma função de *kernel* apropriada, e diferentes funções têm sido empregadas para diferentes tarefas de classificação. As funções mais comuns são apresentadas na Tabela 2.1, em que x_i e x_j são os vetores de dados para dois padrões.

Tabela 2.1: Funções de *kernel* mais comuns [84]

Kernel	**Função**	**Parâmetros**
Polinomial	$\left(\delta(x_i \cdot x_{j)} + k\right)^d$	δ: escala k: deslocamento d: grau do polinômio
Radial Basis Function (RBF)	$e^{-\sigma\lvert x_i - x_j\rvert^2}$	σ: largura do raio
Sigmoidal	$\tanh(\delta(x_i \cdot x_j) + k)$	δ: escala k: deslocamento

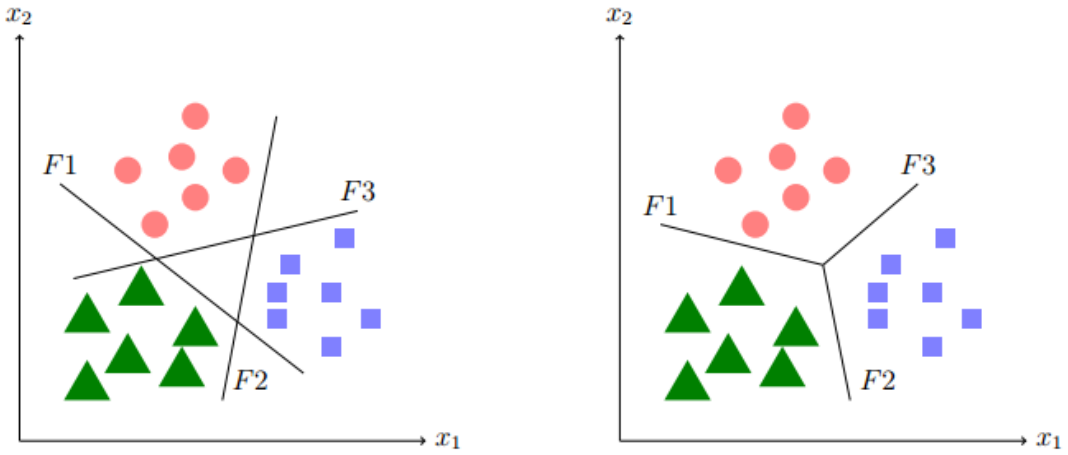
Com a utilização de funções de *kernel*, o problema dual para encontrar o hiperplano que maximiza a margem de duas classes é modificado para a Equação (2.24), em que *K* é a uma função de *kernel* que mapeia o espaço de entrada para uma dimensão maior. Logo a função de decisão é reescrita para a Equação (2.25)

$$\underset{\propto}{maximizar} \quad \sum_{i=1}^{n} \alpha_i y_i - \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} \alpha_i \alpha_j y_i y_j K(x_i, x_j)$$
$$com\ restrições: \begin{cases} 0 \le \alpha_i \le C, i = 1, \dots, n \\ \sum_{i=1}^{n} \alpha_i y_i = 0 \end{cases} \qquad 2.24$$

$$f(x) = sgn\left(\sum_{i=1}^{n} y_i \alpha_i K(x \cdot x_i) + b\right) \qquad 2.25$$

Inicialmente a SVM foi desenvolvida para o problema de classificação binária, ou seja, para classificação com apenas duas classes. Problemas de classificação com várias classes

podem ser resolvidos através da criação de um novo modelo de classificação múltipla. No entanto, múltiplos modelos de classificação são complexos no cálculo e difíceis de implementar. Por isso, é melhor resolver o problema com mais de uma classe usando vários classificadores binários. Um-Contra-Um (Figura 2.21(b)) e Um-Contra-Todos (Figura 2.21(b)) são duas estratégias regulares para formar um classificador multiclasse a partir de uma série de classificadores binários [41].



(a) Estratégia Um-Contra-Todos (b) Estratégia Um-Contra-Um

Figura 2.21: Estratégias para SVM Multiclasses

Para problemas com N classes, a estratégia Um-Contra-Um deve criar $\frac{N(N-1)}{2}$ classificadores binários, enquanto que Um-Contra-Todos necessita de apenas N classificadores binários.

A Figura 2.21 apresenta um exemplo para classificar um problema com 3 classes. Quando as amostras estão na região do triângulo formado pelo cruzamento de duas fronteiras de decisão, denotado na figura por Fn, Um-Contra-Todos é difícil dizer a qual dos casos a classe pertence, no entanto necessita menos custo computacional [41].

Em Um-Contra-Todos, quando se treina uma classe, todas as demais são utilizadas como amostras negativas. Em Um-Contra-Um, é realizado o treinamento de uma classe considerando cada uma das demais.

2.5. Considerações Finais

Este capítulo descreveu as principais técnicas utilizadas para a implementação no reconhecimento de expressões faciais. Foram abordadas as etapas necessárias para conseguir

reconhecer uma expressão, assim como as técnicas envolvidas. Para detecção facial foi apresentado o algoritmo proposto por Viola-Jones e a extração de características foi descrita com LBP e WLD. Para redução da dimensionalidade foram explicadas as estratégias *wrapper* e *filter*, e as técnicas IG, KW e CFS. Por fim, para classificação foram apresentados os algoritmos kNN e SVM. No capítulo seguinte são descritos e detalhados recentes trabalhos desenvolvidos para determinar expressões faciais.

Capítulo 3

Estado da Arte

O cérebro humano tem grande facilidade em compreender e reconhecer expressões faciais, é rápido e não requer nenhum esforço. No caso de um sistema computacional, este processo envolve uma série de restrições e, consequentemente, implica o uso de um conjunto de técnicas e algoritmos relativamente complexos [30]. Neste capítulo serão abordados alguns dos recentes trabalhos da literatura e as técnicas e estratégias empregadas para reconhecer expressões faciais. Normalmente os métodos estão constituídos por detecção facial, extração de características, redução de dimensionalidade e classificação (Figura 3.1) [16][31].

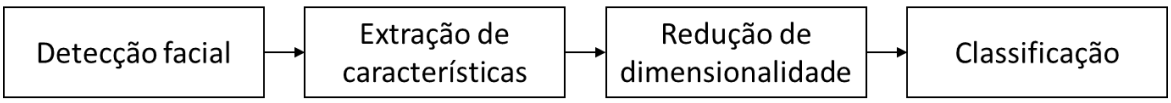


Figura 3.1: Estruturas básicas de um sistema para reconhecimento de expressões faciais

Na detecção facial é extraída a face da imagem em que será aplicado o reconhecimento de expressão facial (REF), isso permite remover dados desnecessários assim como elementos de fundo. Em alguns trabalhos [32] nesta etapa também é aplicado algum tipo de pré-processamento nas imagens, como melhoria do contraste e redimensionamento, sendo que o processo para ajustar a face para um tamanho padrão também pode ser realizado na etapa de extração de características dependendo da proposta do método e dos algoritmos utilizados.

Após a face ser detectada é aplicado a extração de características para obter informações relevantes da face e posteriormente ser utilizada por um algoritmo de Aprendizagem de Máquina. A extração de características deve fornecer um conjunto de dados discriminantes entre as expressões e preferencialmente com baixa dimensionalidade. Conforme observado na literatura, grande parte dos trabalhos focam em melhorar esta etapa, evidenciando que a

extração é o ponto chave para atingir bons desempenhos [22]. De acordo com Zavaschi et al. [17] e Bashar et al.[15], as duas principais categorias para extração de características utilizadas para REF estão baseadas em:

- Geometria: esta estratégia está baseada na obtenção de informações como distância, posições e ângulos entre diferentes componentes faciais, como olhos, sobrancelhas, nariz e boca (Figura 3.2);

Figura 3.2: Pontos utilizados para a extração de características baseado em geometria [22]

- Textura: esta abordagem explora vincos, rugas e dobras da face como informações para classificar uma expressão (Figura 3.3).

Figura 3.3: Extração de características baseada em aparência [14]

Por fim, a última etapa do processo é a classificação. Nesta fase o conjunto de informações fornecido pela extração de características deve ser avaliado e classificado fornecendo uma expressão facial. Alguns trabalhos antes da classificação aplicam algumas técnicas de seleção de atributos para reduzir a dimensão do vetor de características.

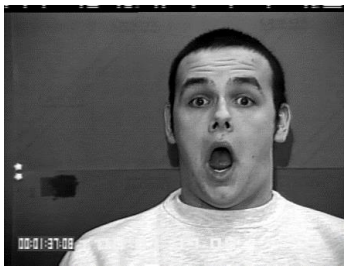
3.1. Detecção Facial e Pré-Processamento

Esta etapa é muito importante para reduzir a área de interesse da imagem focando o processamento todo na face, que é onde contém as informações relevantes para o REF. Existem diversos trabalhos [14][16][33][34][35][36] que tem utilizado o algoritmo de Viola-Jones [37], por ser um método rápido e eficiente para detectar objetos. Testes demonstram que o algoritmo

de detecção facial proposto por Viola-Jones é capaz de processar uma imagem 384×288 em 0,067 segundos, sendo no mínimo 15 vezes mais rápido que outras abordagens e ainda obter desempenho semelhantes [37].

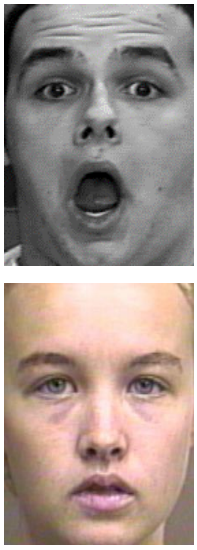
Abordagens que realizam a detecção facial em imagens através da cor da pele também vem sendo utilizadas [38] [23]. Segundo Sobia et al. [38], as componentes RGB são inadequadas para detecção de pele, pois além da cor também representam a luminosidade, que pode variar dependendo do ambiente. Para resolver essa questão, o espaço RGB é convertido para YCbCr, em que *Y* é a luminosidade, *Cb* é a diferença cromática de azul e *Cr* é a diferença cromática de vermelho. Assim a pele é detectada com o uso das componentes Cb e Cr, e um limiar θ que estabelece a detecção de pele. Por considerar componentes RGB, o método de detecção facial baseado em YCbCr é incapaz de funcionar corretamente em imagens do conjunto JAFFE e algumas do conjunto TFEID, em que são representadas em nível de cinza.

Um ponto muito importante da detecção facial é fornecer uma base de alinhamento para as faces, isso faz com que as respectivas características extraídas das face sejam sempre extraídas de regiões próximas [35]. A face em uma imagem pode estar localizada em qualquer lugar e a Figura 3.4 ilustra um exemplo de duas imagens do conjunto CK. Quando não é utilizado a detecção facial, ao extrair as características que representam a face e produzem o vetor de características utilizado para classificação, a região dos olhos e boca da face da Figura 3.4(a) corresponde ao plano de fundo da face presente na Figura 3.4(b). Desta forma o aprendizado do modelo não é baseado na expressão facial, e sim na diferença de imagens, consequentemente o aprendizado e a predição realizada por um classificador podem ser prejudicados. No entanto quando é realizada a detecção facial e eliminada toda a região desnecessária (Figura 3.4(c)), as características extraídas de uma face correspondem a regiões próximas da outra face, assim as variações decorrentes de cada tipo de expressão podem ser aprendidas por um classificador. O redimensionamento das faces para um tamanho comum também é utilizado para melhorar o alinhamento e padronizar as regiões das faces.



(a) Imagem com face localizada a direita.

(b) Imagem com face localizada a esquerda



(c) Faces obtidas com a detecção facial

Figura 3.4: Exemplo de alinhamento de faces.

Segundo Sadegui et al. [39], as variações de uma expressão facial são mais perceptíveis pelo formato da boca e dos olhos, no entanto, as rugas e sulcos gerados por uma expressão têm menor variação do que características geométricas. Assim o formato do rosto é normalizado para modelo geométrico fixo (Figura 3.5) e, em seguida, é aplicado um extrator de características baseado em textura para representar a face. Seu método consiste em utilizar pontos faciais para ajustar um modelo geométrico, então através da Triangulação de *Delaunay* a face é segmentada e com o *Piecewise Linear Warp* os segmentos são normalizados para um padrão pré-definido. O método utilizado pelo autor é bem sofisticado, no entanto o trabalho não aborda como são encontrados os pontos fiduciais.

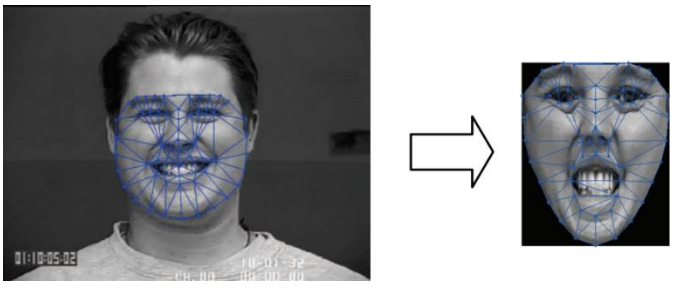


Figura 3.5: Normalização da face em Sadegui et al. [39]

As diferenças de iluminações entre as faces tendem a afetar o desempenho da extração de características e da classificação. Para diminuir a influência da iluminação diversos trabalhos aplicam técnicas de melhoria de contraste nas imagens das face [25][36][38]. Em um estudo comparativo utilizando *Principal Component Analisys* (PCA), na etapa de classificação foi possível elevar a taxa de reconhecimento de expressões faciais de 83% para 87% somente com a equalização do histograma [40]. Também para reconhecer expressões faciais, Zilu e Guoyi1 [41] utilizaram o histograma como função de distribuição de probabilidade para melhorar o contraste. Ainda existem métodos mais complexos compostos por dois estágios em que são introduzidas variáveis de ajuste para regular o ganho de luminosidade e impor limite na contagem de níveis de cinza do histograma [42].

Quanto maior o tamanho de uma imagem, maior tende a ser o custo computacional para o processamento, por este motivo alguns trabalhos procuram reduzir o tamanho das faces. Ainda o redimensionamento das faces para um tamanho padrão possibilita que o extrator de características obtenha informações de regiões próximas para diferentes faces, como quando utilizado extrator de característica Gabor Filter que avalia a face de forma global. Em Wang et al. [16] as faces do conjunto JAFFE e CK são redimensionadas para 128×128 pixels. Como o algoritmo utilizado para extração de características da face é baseado em micro padrões de textura, a imagem é divida em 8×8 sub-regiões e cada sub-região produz 156 atributos. Desta forma uma imagem com tamanho que necessite gerar mais divisões, o número de atributos também seria maior, para 8×8 sub-regiões são obtidos 9984 atributos e um aumento para 9×9 sub-regiões produziria 12636 atributos, enquanto que para 7×7 sub-regiões são gerados 7644 atributos, aproximadamente 40% menor que 9×9 sub-regiões . O objetivo é gerar um vetor de características com atributos reduzidos e também reduzir tempo de processamento.

Em Tian [35] foi avaliado a taxa de reconhecimento de expressões faciais em diferentes resoluções com 3 diferentes técnicas para extrair as características em 5 diferentes tipos de abordagens. A extração de características foi realizada com o “G1” *features tracking* [43], “G2” *features dectection* [44] e “AP” Gabor Filter, sendo os dois primeiros baseados em geometira e o último baseado em textura. Os resultados obtidos (Tabela 3.1) mostram que a redução da imagem em 50% em média não produz perda, no entanto ao reduzir a face em 3 vezes a taxa de acerto média cai de 89% para 88.5%.

Tabela 3.1 : Desempenho (%) para reconhecimento de expressões faciais obtidos por [35]

	288×384 (original)	144×192	72×96	36×48	18×24
G1	92,5	91,8	91,6	N/D	N/D
G2	74,0	73,8	72,9	61,3	N/D
AP	91,7	92,2	91,6	77,6	68,2
G1+AP	93,8	94,0	93,5	N/D	N/D
G2+AP	93,2	93,0	92,8	89	N/D
Média	89,0	89,0	88,5	76,0	68,2

Baseado no trabalho de Tian [35], Shan et al [45] avaliou o impacto do reconhecimento de expressões faciais em faces com baixas resoluções. Além das técnicas para extração de características apresentadas em Tian [35], foi incluído o LBP e uma variação do Gabor Filter. Uma rede neural com 3 camadas foi utilizada para classificar as características. A Tabela 3.2 apresenta os resultados obtidos por Shan et al [45] e é possível verificar que a medida que a resolução diminui, o reconhecimento de expressões faciais também degrada. Em média quando se reduz uma imagem em 50% a taxa de reconhecimento diminui de 88,0% para 87,2%, e ao diminuir a face em 3 vezes a taxa de acerto decaí para 86,9%.

Tabela 3.2: Desempenho para reconhecimento de expressões faciais obtidos por [45]

	110×150	55×75	36×48	27×37	18×24	14×19
LBP	92,6	89,9	87,3	84,3	79,6	76,9
Gabor	89,8	89,9	86,4	83	78,2	75,1
AP	92,2	91,6	N/D	N/D	N/D	68,2
G1	91,8	91,6	N/D	N/D	N/D	N/D
G2	73,8	72,9	N/D	61,3	N/D	N/D
Média	88,0	87,2	86,9	76,2	78,9	73,4

3.2. Extração de Características

A extração de características é considerada a etapa mais importante no REF. Nesta fase são obtidas as informações que serão utilizadas na classificação, ou seja, quanto melhor a qualidade dos dados obtidos nesta fase, melhor será o desempenho do método para reconhecer expressões faciais [22]. A prova da relevância desta etapa pode ser verificada na literatura, pois é a etapa em que os autores têm focado grande parte da atenção.

Existe uma grande variação de técnicas utilizadas para extrair características das faces e seguem a mesma divisão das abordagens para extração de características, ou seja, as técnicas

são baseadas em características geométricas ou são baseadas em textura. Os algoritmos baseados em geometria como o *Active Appearance Model* (AAM) e *Template Matching* são aplicados para determinar pontos fiduciais e representar uma face de acordo com o formato geométrico dos olhos, boca ou sobrancelha. Os algoritmos baseados em textura são utilizados para obter informação de uma face como a textura da pele, incluindo rugas e sulcos. As técnicas com *Principal Component* Analysis (PCA), Gabor Filter, *Discrete Cosine Transform* (DCT) e Eigenfaces são empregadas para obter informações de textura da face como um todo. Enquanto que *Local Binary Pattern* (LBP), *Local Ternary Pattern* (LTP), *Weber Local Descriptor* (WLD) e *Median Ternary Pattern* (MTP) são técnicas que extraem micro padrões das faces [46][47].

A PCA é uma abordagem estatística utilizada para encontrar as principais componentes de cada imagem do conjunto de treinamento, sendo representada como uma combinação linear de vetores [38]. Utilizando a PCA para obter informações de imagens, o trabalho de Deng et al. [40] obteve 90% de reconhecimento de expressões separando o conjunto JAFFE em 138 imagens para treinamento e 75 para teste. A partir da PCA foi criada a Eigenfaces, que é uma adaptação com custo computacional reduzido. Em Sobia et al. [38] foi usado a Eigenfaces para extrair informações de faces e obteve-se uma taxa de acerto de 97% para um pequeno conjunto de dados privados.

O LBP é um operador para extrair informação de textura invariante de escala que rotula os pixels da vizinhança de um valor central e produz o resultado como um padrão binário [14]. O LBP tem sido utilizado por Verma e Dabbagh [14] proporcionando uma taxa de acerto de 87% e no trabalho de Zavaschi et al. [17] utilizando fusão de classificadores foi possível atingir 99% de reconhecimento. No estudo de reconhecimento de facial em que também se faz necessário obter características da face, a PCA demorou 220 milissegundos para realizar a tarefa com 65% de taxa de acerto, enquanto o LBP demorou 5.23 segundos com 95% de taxa de acerto [33], ou seja, o LBP consegue obter características mais discriminantes da face do que a PCA, no entanto seu custo computacional é superior.

O LBP é um algoritmo que extrai informações de imagens a partir das texturas. As expressões faciais produzem pequenos padrões que são representados pelas mudanças na textura da face, como rugas e sulcos. As pequenas variações que ocorrem em função de cada expressão podem ser obtidas aplicando o LBP.

No estudo de Shan et al. [45] é possível verificar as regiões da face em que expressões faciais são mais discriminantes. Para extrair as características com o LBP foi utilizada uma janela de varredura para percorrer uma face. A janela de varredura inicia com tamanho 10×10 pixels e a cada interação a janela é incrementada em 5 pixels até atingir 25×25 pixels. Então o AdaBoost foi utilizado para selecionar as janelas de varredura que forcem mais discriminantes para cada tipo de expressão. Na Figura 3.6 são apresentados os pontos centrais para as 50 zonas mais discriminantes. É possível identificar que as expressões têm diferentes características LBP discriminantes, e as características discriminantes são distribuídas principalmente nas regiões dos olhos e da boca. A Figura 3.7 ilustrada as sub-regiões com histogramas mais discriminante selecionadas pelo AdaBoost a partir das janelas de varredura mais promissoras.

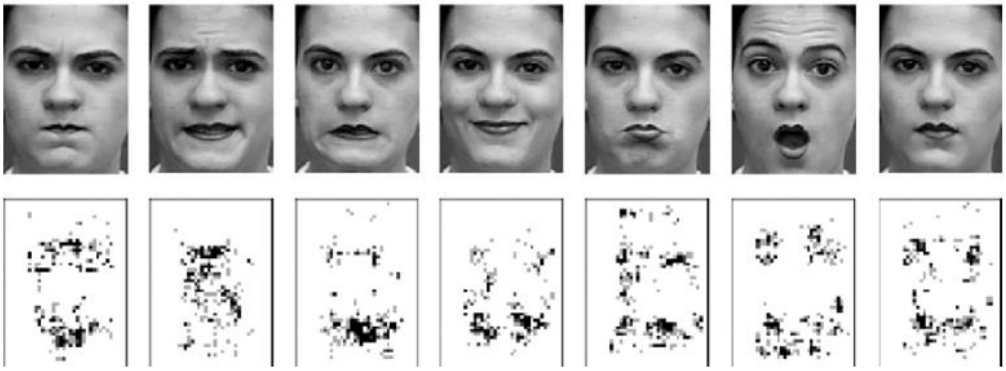


Figura 3.6: Representação das 50 características mais discriminantes selecionadas pelo AdaBoost para cada expressão facial [45]. Da esquerda para direita: raiva, desgosto, medo, felicidade, tristeza, surpresa e neutro

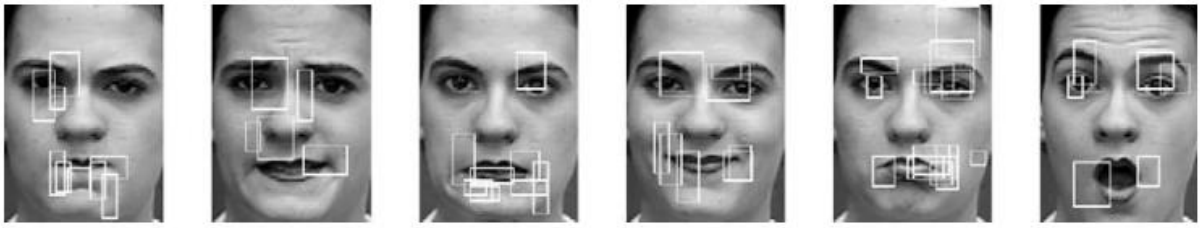


Figura 3.7: Sub-regiões selecionadas pelo AdaBoost para cada expressão. Da esquerda para direita: raiva, nojo, medo, felicidade, tristeza e surpresa.

Assim como o LBP, outro extrator popular baseado em textura é o Gabor Filter. Ao contrário do LBP, este extrator é usado para obter características globais de uma imagem em diferentes escalas e frequências [47]. Esta técnica também tem sido utilizada para encontrar pontos faciais [48]. Nos estudos de Ramireddy e Kishore [47] e Zavaschi et al. [17] o Gabor Filter em conjunto com outas técnicas foi possível classificar corretamente até 99% das expressões faciais do conjunto CK. Em Shan et al. [45] foi realizado uma comparação de tempo e o uso de memória entre o Gabor Filter e o LBP para extrair características da face. Os resultados obtidos da Tabela 3.3 mostram que o Gabor Filter possui alto custo computacional. O fato também pode ser verificado em Wang et al. [16] em que compara o tempo de 4 técnicas para extração de características (Figura 3.8). Devido ao processamento necessário para extrair as características, torna-se difícil aplicar o Gabor Filter em sistemas que necessitam respostas em um curto intervalo de tempo [49][46].

Tabela 3.3: Comparativo de tempo e uso de memória entre LBP e Gabor Filter

	LBP	**GABOR**
Memória (dimensão de características)	2478	42650
Tempo (extração de características)	30 milissegundos	30 segundos

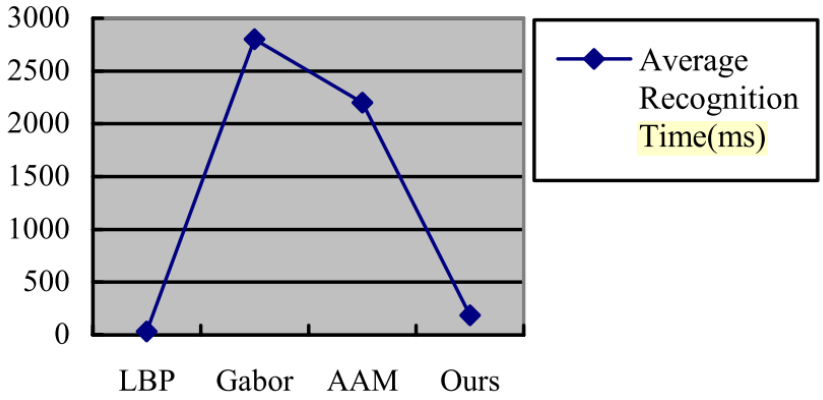


Figura 3.8: Comparativo de tempo para reconhecimento de expressão facial entre LBP, Gabor Filter, AAM e método de [16]

O AAM é um método de extração de características baseado em geometria bastante eficaz para descrever expressões faciais e detectar pontos fiduciais. Seu processo é ilustrado na Figura 3.9 e consiste em gerar um novo modelo AAM minimizando a diferença entre uma imagem de entrada e uma instância modelo através da otimização de seus parâmetros de ajuste

para reduzir o erro entre as imagens [46]. São realizadas interações com um algoritmo de ajuste para determinar os melhores parâmetros de forma e textura do novo modelo AAM [50]. Este algoritmo tem sido utilizado por Martin et al. [36] e Choi e Oh [51] para reconhecer expressões. Na abordagem seguida por Choi e Oh [51] utilizando o algoritmo de ajuste *Efficient Second Order Minimization* (ESOM) a taxa de acerto foi de 99% com um tempo de 180 milissegundos para processar uma face, enquanto que Martin et al. [36] utilizou *Inverse Compositional Algorithm* como algoritmo de ajuste, e atingiu 92% de acerto com tempo de reconhecimento de 24 milissegundos para 15 interações. Segundo Wang et al. [16], as desvantagens do uso da AAM incluem a complexidade dos cálculos e a dificuldade em obter os parâmetros iniciais.

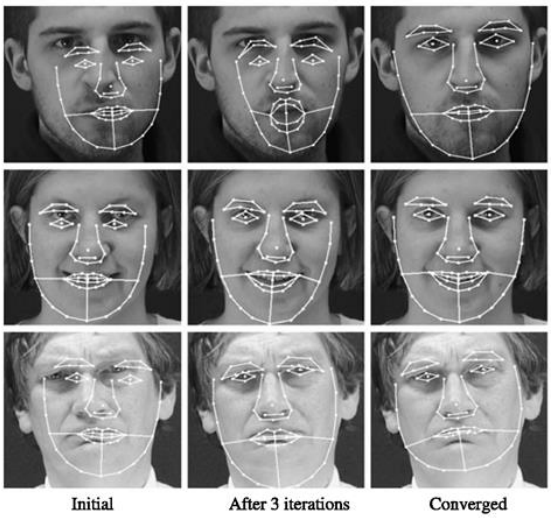


Figura 3.9: Exemplo de detecção com AAM[2]

O sucesso nos últimos tempos de técnicas baseadas em geometria, como pode ser verificado na literatura e é citado em Bashar et al. [15], começaram a ser questionados quanto a sua usabilidade em cenários reais. De acordo com Bashar et al. [15] a detecção dos pontos faciais é muito dependente do ambiente e apresenta maior custo computacional. A partir destes fatos, os métodos baseados em textura passaram a ser explorados mais intensivamente.

No estudo de Shan et al [45] foram avaliadas duas abordagens, a primeira que consiste em características geométricas extraídas a partir do método proposto por Cohen et al. [52] e classificação com *Tree-Augmented-Naive Bayes*. O método de Cohen et al. [52] realiza a

[2] Imagem retirada de http://what-when-how.com/face-recognition/face-alignment-models-face-image-modeling-and-representation-face-recognition-part-3/

detecção automática dos pontos fiduciais através de interações que ajustam um modelo de representação facial à uma face. O *Tree-Augmented-Naive Bayes* é um classificador baseado redes Baysianas que possibilita representar dependências entre pares de atributos. A segunda abordagem é baseada em textura, foi utilizado LBP para extrair informação das imagens e KNN para classificação, também foi utilizado T*emplate Matching* para medir a similaridade de um histograma de entrada com os *templates* de cada expressão facial. No conjunto de imagens utilizadas por Shan et al [45] não foram aplicados técnicas para melhoraria de contraste. Os resultados obtidos demonstram o método baseado em textura conseguiu reconhecer corretamente 79% das expressões faciais, enquanto que o método baseado em geometria atingiu 73%.

Os trabalhos mais recentes, têm se preocupado com o tempo envolvido no reconhecimento de expressões faciais, além do desempenho. Em Wang et al. [16] (Figura 3.8) pode ser verificado que os custos exigidos pelos extratores são elevados, em seu comparativo Gabor Filter levou 2.7 segundos, a AAM 2.2 segundos e o método proposto com *Histogram Oriented Gradient* (HOG) e WLD demorou 150 milissegundos para reconhecer uma expressão facial. Os trabalhos desenvolvidos por Shuaishi et al. [53] e Hussain et al. [2] para identificar expressões faciais também consideraram o tempo para avaliar o método proposto.

Diante deste cenário, as abordagens baseadas em textura vêm ganhando destaque e técnicas mais robustas estão sendo utilizadas no reconhecimento de expressão facial, como o WLD, e até melhorias de técnicas consolidadas. Um exemplo é o LBP, a partir deste algoritmo foram propostos o LTP e MTP que tem alcançado taxas de acerto de até 98% e 94% respectivamente [15].

O estudo de Bashar et al. [15] utilizando MTP para extrair características locais avaliou impacto em realizar do zoneamento em faces para reconhecer expressões faciais. Como ilustrado na Figura 3.10 o zoneamento é a divisão da face em sub-regiões, isso permite que técnicas baseadas em textura locais consigam representar melhor os micros padrões existentes. No seu estudo, a divisão de face em 3×3 zonas possibilitou uma taxa de acerto de 95% para MTP e 75.3% para PCA, enquanto que a divisão em 7×6 partes foi alcançado 98% para MTP e 89% para PCA. Conclui-se que a medida que são geradas mais divisões, a taxa de acerto aumenta, pois é possível extrair mais informações locais, no entanto a dimensionalidade também aumenta. Considerando extrator MTP que produz 512 dimensões, em um zoneamento com 7×6 divisões são gerados vetores com $512 \times 7 \times 6 = 21504$ atributos

Figura 3.10: Exemplo de zoneamento com 3 × 3 sub-regiões

Com o conjunto de técnicas WLD, SVM e zoneamento, foi possível obter até 91% de acerto considerando apenas uma parte da face [53]. Isso demonstra que as características baseadas em de textura são robustas e eficientes. Neste estudo com a oclusão da região dos olhos (Figura 3.11(a)) foi identificado corretamente 87% das expressões faciais, a oclusão da região da boca (Figura 3.11(b)) atingiu 89%, e lado esquerdo (Figura 3.11(c)) ou direito (Figura 3.11(d)) aproximadamente 90%. Considerando toda a face 96% das expressões faciais são classificadas corretamente.

(a) Oclusão dos olhos (b) Oclusão da boca (c) Oclusão da lateral esquerda (d) Oclusão da lateral direita

Figura 3.11: Exemplo de oclusões realizadas por [53]

A Figura 3.12 (coluna do meio) mostra que o WLD é capaz de extrair bordas com perfeição mesmo com ruído. Além disso, os resultados da análise de textura mostram que grande parte da informação de textura discriminante está contido em altas frequências espaciais como bordas [54]. Desta forma é possível verificar que o WLD é um extrator de texturas poderoso [55]. Ainda o estudo de Chen et al. [56] demonstra que a complexidade do WLD é baixa, sendo descrita por $O_{WLD} = C_1 mn$, em que C_1 é a computação para as operações de um pixel e as variáveis m e n representam a dimensão da imagem. A complexidade de tempo é

semelhante ao LBP, que é dada por $O_{LBP} = C_2 mn$, em que C_2 é a computação para as operações de um pixel com LBP.

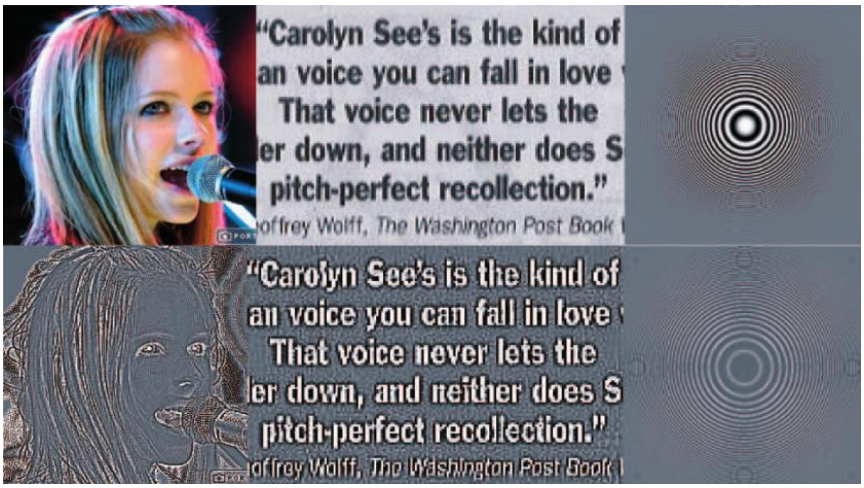


Figura 3.12: A linha superior contém imagens originais e na linha abaixo são as respectivas imagens filtradas com o WLD. A intensidade de cada pixel das imagens filtradas é determinada pelo *diferencial de excitação* escalados de 0 à 255 [55].

Em busca de novas alternativas para o REF, em Ramireddy e Kishore [47] foi avaliado o desempenho fundindo características globais e locais da face. No estudo foi utilizado *Discrete Cosine Transform* (DCT) para a extração global e Gabor Filter para extração local dos olhos, boca e nariz. Na validação foram gerados dois diferentes conjuntos de faces com ângulo de rotação de 2º e 5º do conjunto CK, e para os dois conjuntos gerados a partir da JAFFE o ângulo de rotação das faces é de 5º e 10º. O autor não deixa claro como é realizado a construção do algoritmo de classificação, ou seja, se são utilizadas as faces rotacionadas para gerar o modelo de classificação. Os resultados obtidos são apresentados na Tabela 3.4 e demonstram que a fusão de características locais e globais é cerca de 2% melhor que o uso características globais e 17% melhor que as características locais. Os resultados podem variar de acordo o conjunto de técnicas utilizadas, uma vez que o método foi avaliado com apenas um extrator de características locais e um global, e recentemente as representações baseadas em características locais tem se destacado por conseguir capturar grande parte das pequenas informações relevantes [39]. Deve-se avaliar que a utilização de dois extratores de características aumenta a complexidade do método e o custo computacional, pois é necessário obter as características de uma imagem utilizando dois algoritmos e consequentemente a etapa de redução de

características processa mais atributos do que utilizando apenas as informações fornecidas por um único extrator de características.

Tabela 3.4: Taxas de acerto obtidos por [47]

	JAFFE		**CK**		**MÉDIA**
Rotação da face em graus	**5º**	**10º**	**2º**	**5º**	**(%)**
DCT+PCA+RBF	77,5	51,25	90,5	66,94	71,55
Gabor+PCA+RBF	94,16	80,41	94,11	76,11	86,20
Gabor+DCT+PCA+RBF	96,67	77	98,33	81,66	88,42

De modo geral, avaliando a capacidade para representação de expressões faciais, o custo e a complexidade computacional, os algoritmos LBP e WLD são duas alternativas promissoras para a extração de características. Conforme descrito anteriormente os algoritmos são utilizados em trabalhos que conseguem as maiores taxas de acerto no reconhecimento de expressões faciais com tempo de processamento menor que outras abordagens. No entanto, assim como o MTP, o LBP e WLD extraem características de textura baseada em micro padrões sendo necessário zonear a face, produzindo um espaço de alta dimensionalidade. Como apresentado anteriormente uma face dividida em 6×7 zonas pode produzir mais de 21000 atributos dependendo da configuração do extrator de características, o que faz da etapa de redução de dimensionalidade um excelente recurso para reduzir o custo computacional do modelo de aprendizagem e o custo computacional envolvido (menos atributos a serem processados).

3.3. Redução de Dimensionalidade

Nesta etapa são utilizadas técnicas para remover as informações redundantes, pois funcionam como ruído para os classificadores, e também são eliminados os atributos que não contribuam para a distinção entre as classes, e como consequência é obtido a redução no custo computacional e melhora no desempenho de classificação. A redução de dimensionalidade é importante para diminuir o tempo e espaço, também os modelos simples são mais robustos com pequenos conjuntos de dados, além de evitar a Maldição da Dimensionalidade. É muito frequente o uso de PCA para a seleção de atributos [57][46][40]. Esta técnica é simples e eficiente, e reduz a dimensionalidade do vetor de atributos enquanto retêm a variação presente no conjunto de dados.

No trabalho de Ramireddy e Kishore [47] foi utilizado a *Kernel Principal Component Analisys* (KPCA) para reduzir o número de atributos gerados por dois extratores, o Gabor Filter e o DCT. Enquanto que em Deng et al. [40] é avaliado a extração de características com várias configurações do Gabor Filter e duas técnicas são aplicadas para reduzir a dimensionalidade, a PCA e a LDA. A classificação das expressões faciais é feita pela distância entre os exemplos de treinamento com a respectiva imagem de entrada. O conjunto JAFFE é utilizado para validação, separando 138 imagens para treinamento e 75 de testes. Os resultados mostram que para classificação com distância Euclidiana, em média a redução de características somente com PCA é possível atingir 79% de acerto, enquanto que com PCA e LDA o desempenho atinge os 96%, sendo que a PCA produz uma saída com 180 dimensões e a LDA com apenas 6. Ao contrário da PCA, a LDA produz um espaço que aumenta a distinção das classes demonstrando que a avaliação da classes para a redução de dimensionalidade influência no resultado final do método. O trabalho Deng et al. [40] deveria ser investigado sem a redução de características, além de outros conjuntos de dados e extratores de características para avaliar qual o ganho real de desempenho com o uso de PCA e LDA.

Recentemente Hussain et al. [2] tem utilizado Kruskal Wallis (KW) para reduzir a dimensionalidade do vetor de atributos constituído por dois extratores baseados em histograma, o WLD e o LTP. O método KW tem baixo custo computacional e consiste em selecionar os atributos baseado na mediana de duas ou mais classes. Utilizando apenas as características extraídas pelo WLD foi classificado corretamente 75% das expressões faciais do conjunto JAFFE com um vetor de atributos com 50 dimensões, enquanto que a taxa de acerto com LTP foi 72% e vetor de atributos de mesmo tamanho que WLD. Após fundir os histogramas (Figura 3.13) dos respectivos extratores de características e aplicar o método KW para reduzir a dimensionalidade, foi selecionado um vetor de atributos com 80 dimensões e a taxa de acerto alcançou 83%, ou seja, uma melhora de 8%. Assim como em Deng et al. [40], outros cenários e conjunto de técnicas poderiam ter sido investigados em Hussain et al. [2]

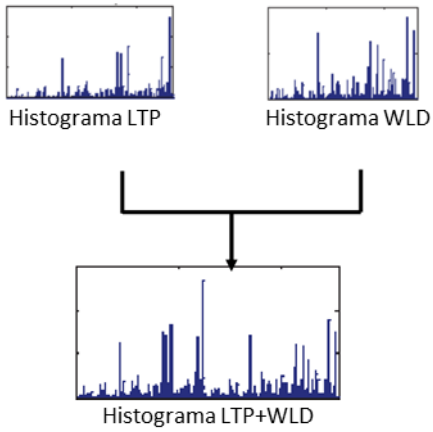


Figura 3.13: Fusão de características entre LTP e WLD [2]

Em Kyperountas et al. [18] é proposto uma abordagem para redução de dimensionalidade utilizado pares de expressões faciais, para isso as faces foram detectadas manualmente e então são extraídas as características da face com Gabor Filter. Como a redução de características é realizada por pares de expressões faciais, com 7 tipos de expressões faciais são produzidas 21 combinações, ou seja, existem 21 subconjuntos de atributos, sendo que cada um é resultante da seleção de atributos de duas expressões faciais. A LDA foi a técnicas utilizada para seleção de atributos, que ao final produz um espaço de 24 dimensões, apesar da baixa dimensionalidade o custo computacional com Gabor Filter ainda é elevado. Cada um dos subconjuntos de atributos é classificado por um algoritmo baseado em distância, ao final a predição é dada para a expressão com mais indicações. A taxa de acerto com *leave-one-sample-out* foi de 95% para o conjunto JAFFE. Os resultados obtidos demonstram que a seleção de atributos em pares pode ser uma alternativa para reduzir a dimensionalidade, no entanto, deve-se avaliar que a abordagem foi conduzida em um único conjunto de técnicas e também foi avaliado somente no conjunto JAFFE impossibilitando obter uma conclusão mais precisa do desempenho da abordagem.

A etapa de redução de dimensionalidade tem sido menos explorada no reconhecimento de expressões faciais, ou seja, a seleção de atributos normalmente é realizada com a PCA sob todo o conjunto de atributos disponíveis, e os resultados obtidos por Kyperountas et al. [18], demonstram que é importante buscar novas alternativas. A PCA apesar de ser frequentemente utilizada não garante de que o as principais componentes consigam uma maior discriminação entre as classes, podendo prejudicar no desempenho de algoritmos de aprendizagem de máquina. Por outro lado, a LDA reduz a dimensionalidade ao mesmo tempo que mantêm a

discriminação entre as classes, no entanto esta técnica também possui desvantagens como alta sensibilidade a *outlier* [59]. Isso evidência a necessidade de explorar novas técnicas como seleção de atributos baseado em ganho de informação, seleção de atributos baseada em correlação, Krusakal Wallis, Informação mútua e outras.

3.4. Classificação

Para fazer o reconhecimento de expressões faciais a partir do conjunto de informações obtidas da face, algoritmos como vizinho mais próximo (*k-Nearest Neighbor* – KNN), SVM, Redes Neurais (Neural Networks – NN) e AdaBoost tem sido utilizados para classificação. Abdulrahman et al. [57] conseguiram classificar corretamente mais de 90% das expressões faciais utilizando a Distância Euclidiana, no entanto o método não utiliza validação cruzada e separa o conjunto JAFFE em 137 imagens para treinamento e 76 imagens para teste. Em uma abordagem baseada em geometria [46], o KNN com medida Chi Quadrado obteve uma taxa de acerto de 93% para um pequeno conjunto de imagens da CK. O KNN tem um problema de desempenho, todas as instâncias de treinamento devem ser armazenadas, o que exige espaço e tempo para calcular a distância dos atributos de cada exemplo do conjunto de treinamento.

A NN tem produzido bons resultados em Juanjuan et al. [32] e Ramireddy e Kishore [47] no reconhecimento de expressões faciais atingindo 94% e 98% de acerto respectivamente para o conjunto JAFFE. Com o objetivo de melhorar os resultados, em Hussain et al. [2] são utilizadas 6 redes neurais, cada uma especializada em uma emoção. A abordagem alcançou 83% de taxa de acerto, no entanto não há nenhuma comparação com o uso de somente uma NN e não deixa claro o protocolo de validação.

Devido ao bom desempenho da SVM, este é o classificador que mais tem sido utilizado na identificação de expressões. Vários trabalhos [17][39][14][53] tem aplicado SVM para classificação atingindo até 96% de acerto [17] com validação cruzada de 10 partições. A SVM tem conseguido superar a *Multilayer Perceptron* (MLP) em alguns casos. Uma comparação [36] entre MLP e SVM, foi obtido uma taxa de acerto de 75% e 92% respectivamente.

O AdaBoost é um algoritmo com desempenho superior a SVM, mas tem sido menos utilizado no REF devido ao seu custo computacional. No trabalho de Verma e Dabbagh [14] o AdaBoost reconheceu corretamente 87% das expressões faciais e a SVM com *kernel* RBF alcançou 83%. Entretanto o tempo exigido pelo AdaBoost tende a ser mais elevado, o tempo médio para classificar uma face utilizando SVM foi de aproximadamente 41 milissegundos, enquanto que o Adaboost foi de aproximadamente 900 milissegundos.

Uma abordagem pouco utilizada e mais complexa é o uso de Algoritmos Genéticos proposto por Zavaschi et al. [17] o qual mostrou ser eficaz alcançando uma taxa de acerto de 99% para CK. Em seu estudo foram utilizados pontos fiduciais, os quais foram detectados manualmente.

Existem trabalhos na literatura que propõem abordagens alternativas para reconhecer expressões faciais. Em: Shuaishi et al. [53] a face é dividida em 2×3 sub-regiões para serem classificadas separadamente. Em cada uma das divisões é realizada a extração de características com o WLD e classificação com SVM. Desta forma, cada divisão gera como saída uma expressão facial e o resultado da predição é atribuída a expressão mais predominante. Em Hussain et al. [2] a etapa de classificação é dividida em partes, de tal modo que uma rede neural é especializada em reconhecer uma única expressão facial para cada uma das 6 emoções utilizadas. Cada rede neural fornece como saída uma pontuação do exemplo avaliado pertencer a expressão do classificador. A predição final é feita para a expressão com maior pontuação.

Uma das principais diferenças encontradas entre os trabalhos de REF é a validação. Cada autor utiliza sua própria avaliação, alguns utilizam Validação Cruzada, enquanto outros separam os conjuntos em treinamento e teste. Também existem casos em que são utilizadas somente bases privadas [51] para a validação e não são consideradas bases públicas como a JAFFE e a CK, que são muito populares e difundidas na literatura. A falta de um método comum para validação dos trabalhos é um fator que impede uma comparação mais coerente entre os trabalhos. Ainda existem trabalhos que validam seus métodos separando os sujeitos em teste e treinamento produzindo taxas de acerto menores. Em Zavaschi et al. [17] com os mesmos sujeitos no treinamento e no teste foi obtido 99% de acerto e quando os sujeitos de treinamento são diferentes do conjunto de teste a taxa de acerto é de 89%, o mesmo ocorre para Juanjuan et al. [32], que consegue taxas de acerto de 94% e 77% para as respectivas abordagens.

Na Tabela 3.5 é apresentado um resumo do conjunto de técnicas e taxas de acertos de trabalhos relacionados ao REF. Conforme apresentado anteriormente os extratores baseados em textura com maior uso são LBP e Gabor Filter. O AAM que é baseado em geometria vem perdendo espaço devido as suas limitações e propiciando na investigação de novas técnicas baseadas em textura, como WLD e MTP. Enquanto para classificação os trabalhos têm focado mais na SVM, pela sua eficiência e baixo custo computacional. As divergências que existem nos resultados devem-se pela maneira que foi realizada a avaliação, alguns trabalhos utilizam

Validação Cruzada, outros separam a conjunto em treinamento e teste. Existe também uma variação quanto ao número de expressões avaliadas e imagens de exemplos utilizadas.

Tabela 3.5: Resumo dos trabalhos desenvolvidos para REF

TÉCNICAS	RESUMO	COMENTÁRIOS	EXATIDÃO
LBP + SVM RBF – 2013 [14]	Extração com LBP e SVM para classificação	Não foi avaliado a expressão de medo mas foi considerado neutro em um total de 6 expressões. A extração de características foi aplicada na região dos olhos e boca. Foi utilizada Validação Cruzada com 10 partições.	83% (JAFFE)
LBP + AdaBoost – 2013 [14]	Extração com LBP e AdaBoost para classificação		86% (JAFFE)
Delaunay Triangulation + Piecewise Linear Warp + LBP + SVM) – 2013 [39]	Normaliza a face para um formato padrão utilizando Delaunay Triangulation e Piecewise Linear Warp. Extração de características com LBP e classificação com SVM	Os pontos faciais não são determinados pelo método. Validação Cruzada com 10 partições utilizando 6 expressões.	94 % (CK+)
LBP + Gabor Filter + Conjutno de SVM's- 2013 [17]	Utiliza LBP e Gabor Filter com pontos fiduciais para extração de características. Cada SVM é treinada por um vetor de características diferente. Ao final são selecionadas o menor grupo de SVM que consiga maior taxa de acerto para classificação.	Pontos fiduciais determinados manualmente. Validação Cruzada com 10 partições utilizando 7 expressões.	96% (JAFFE) 99% (CK)
Gabor Filter + PCA + LBP + kNN - 2014 [57]	Gabor Filter para extração de características, PCA e LBP para redução da dimensionalidade e kNN para classificação	Validação é feita separando o conjunto de dados em 64% (137 imagens) para treino e 36% (76 teste) para teste. Utiliza 7 expressões.	90 % (JAFFE)
AAM + *Efficient Second Order Minimization* (ESOM) + MLP - 2007 [51]	Utiliza ESOM para melhorar a AAM. Classificação é feita por uma MLP.	Os dados são separados 100 para treino e 200 para teste para cada uma das 5 expressões (1000 imagens de teste).	99% (Privada)

AAM + SIFT + kNN (Chi Quadrado) - 2013 [46]	AAM para extrair pontos da face, *Scale-invariant Feature Transform* (SIFT) para representação do gradiente de cada ponto facial e kNN para classificação.	Para validação com CK foram escolhidas 70 imagens para treinamento e 70 para teste, enquanto que para JAFFE foram escolhidas 1-2 imagens por expressão de cada das 10 pessoas. Foram consideradas 7 expressões.	93% (JAFFE) 87% (CK)
Differential AAM +Manifold Learning + kNN - 2008 [60]	Extração de características com AAM, Manifold Learning para seleção de características e kNN para classificação	Validação Cruzada com 2 folds (5x) com 4 expressões.	96% (POSTECH)
Gabor Filter + DCT + KPCA + WFT + NN - 2013 [47]	As características locais são extraídas com Gabor Filter e as características globais com DCT. Em cada uma das saídas geradas é aplicada a KPCA para reduzir a dimensionalidade. As saídas filtradas pela KPCA são concatenadas através da *Wavelet Fusion Technique* (WFT) e então aplicadas em uma rede neural	A validação é realizada separando o conjunto em treinamento e teste. Os resultados são apresentados somente para as faces rotacionadas. A extração local é realizada somente nos olhos, boca e nariz.	98 % (JAFFE) 99% (CK)
SNE + SVM - 2011 [61]	Extração de características com *Stochastic Neighbor Embedding* (SNE) e classificação com SVM	Todas as imagens são reduzidas para 32x32. É utilizada Validação Cruzada com 10 partições. São consideradas com somente algumas faces da JAFFE, sendo avaliadas 7 expressões.	66% (JAFFE)

Sobel Filter + Active Contour + Sipina - 2013 [62]	Depois de detectada a face é aplicado Sobel Filter para encontrar pontos de interesse nos olhos, sobrancelhas e boca. Em seguida é utilizado *Active Contour* para extrair o contorno dos elementos faciais e assim determinar os pontos fiduciais. A classificação é realizada com o algoritmo Sipina com base na medida entre os pontos encontrados.	Não é apresentado como a validação foi realizada e nem apresenta detalhes do classificador. Além das 7 expressões, o autor inclui uma oitava para quando não é possível encontrar os elementos faciais.	70 % (JAFFE)
WLD + LTP + Kruskal Wallis + NN - 2014 [2]	WLD e LTP são usadas para extração de características, em seguida as características são fundidas e selecionadas através da técnica de Kruskal Wallis . Para classificação são utilizadas 6 redes neurais (NN), uma para cada expressão, fornecendo como saída a probabilidade da face pertencer a expressão do classificador. A saída máxima é atribuída ao exemplo.	A forma que foi realizada a validação não é informada. O método de fundir características de dois extratores conseguiu melhorar a taxa de acerto em 8%.	83% (JAFFE)

3.5. Considerações Finais

Neste capítulo foram apresentados alguns dos recentes trabalhos propostos para REF procurando discutir a abordagem seguida por cada método além das técnicas utilizadas. Os trabalhos demonstram uma grande preocupação com o custo computacional necessário para fazer o reconhecimento das emoções, com isso técnicas mais eficientes têm sido exploradas. A alta dimensionalidade produzida pelos extratores de características aumenta o custo computacional e prejudica a aprendizagem dos classificadores, sendo necessário aplicar técnicas para obter um subconjunto menor de atributos. Existe ainda uma diferença nos trabalhos encontrados na literatura por parte da validação, ou seja, não é seguido um padrão para avaliar o desempenho do método, permitindo com que métodos menos eficazes atinjam

altas taxas de acertos. Com base nos pontos levantados neste capítulo, a seguir é proposto uma abordagem para reconhecer expressões faciais, em que é avaliado um novo método para seleção de atributos para elevar o desempenho e reduzir o custo computacional.

Capítulo 4

Método Proposto

Como descrito anteriormente (seção 3.2), na literatura de REF é possível verificar que o desempenho dos métodos que utilizam características geométricas é dependente da precisão dos pontos fiduciais. Em determinados ambientes a localização dos pontos tende a ser mais afetada por fatores como a iluminação. Outra dificuldade encontrada é que as técnicas para localização dos pontos fiduciais tendem a produzir alto custo computacional e os parâmetros iniciais são difíceis de serem determinados. Diante disto, os extratores de características baseados em textura tem sido a opção dos autores, no entanto, uma consequência do uso desta abordagem é a geração de um número muito grande de atributos. A alta dimensionalidade é um fator que afeta diretamente na taxa de acerto do método, pois o modelo gerado para a aprendizagem pode conter muitas características desnecessárias impactando no resultado final. Com base nesta dificuldade, o presente capítulo aborda uma estratégia para amenizar estes impactos da alta dimensão.

Como o método será avaliado em imagens contendo uma expressão facial, é necessário implementar as etapas desde a detecção da face até a classificação. Um caminho alternativo é a detecção manual da face, entretanto, a utilização de um algoritmo para esta tarefa permite uma melhor avalição do método em cenários próximos do mundo real e não necessita da intervenção do usuário. A Figura 4.1 ilustra a estrutura do método, sendo composto pelas etapas de detecção facial, extração de características, redução da dimensionalidade e classificação. Para validar o método serão utilizados 3 conjuntos de dados, 2 extratores de características, 3 técnicas para redução de dimensionalidade e 2 classificadores. Isso permitirá uma avaliação mais precisa da proposta e também será avaliado sem o uso da redução de atributos.

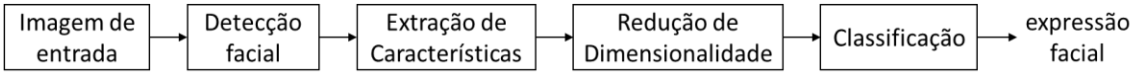


Figura 4.1: Estrutura proposta para reconhecimento de expressões faciais

Para este trabalho serão consideradas as expressões de raiva, medo, alegria, nojo, surpresa, tristeza e neutro. Os resultados obtidos serão comparados com recentes trabalhos da literatura [17][53][15][89][2] e que foram selecionados de acordo com a semelhança deste estudo em relação ao uso de conjunto de dados e protocolo de validação.

Normalmente a redução de dimensionalidade, ou seleção de atributos, é executada com o uso da PCA [57][22][46]. É um método simples e poderoso de eliminar características desnecessárias. No entanto o cálculo das componentes principais não considera a classe dos exemplos (seção 3.3). Por este motivo o presente trabalho explora técnicas mais robusta para selecionar características e com o objetivo de melhorar a distinção entre os tipos expressões. Nas próximas seções serão descritas as etapas e os resultados esperados para cada uma.

4.1. Detecção Facial

A primeira etapa no REF automático é encontrar a face em uma imagem e pelo motivos descritos na seção 3.1 foi utilizado o algoritmo de Viola-Jones [37]. Esse procedimento é importante para extratores de características baseados em textura, pois como se faz necessário dividir a face em sub-regiões, com a detecção facial é possível melhorar o alinhamento entre as diferentes faces, permitindo que cada sub-região corresponda sempre a mesma posição em uma face (seção 3.1). O algoritmo de Viola-Jones [37] foi inicialmente utilizado para detectar faces, mas pode ser modificado para localizar olhos, boca, nariz, além de outros objetos, para isso deve-se treinar o detector com imagens relativas ao que se deseja encontrar. Este detector é largamente utilizado na literatura [14][16][33][34][35] devido a sua precisão e baixo custo computacional em relação as outras técnicas da literatura (seção 3.1).

Após encontrada uma face, são utilizadas instâncias do algoritmo de Viola-Jones [37] para detectar os olhos. Como ilustrado na Figura 4.2, a borda inferior da região da face detectada e as laterais dos olhos são utilizados de base para o recorte da região facial [15][3]. Depois do recorte da face é aplicado a equalização do histograma para melhorar o contraste da imagem. As faces não detectadas são descartadas do processo.

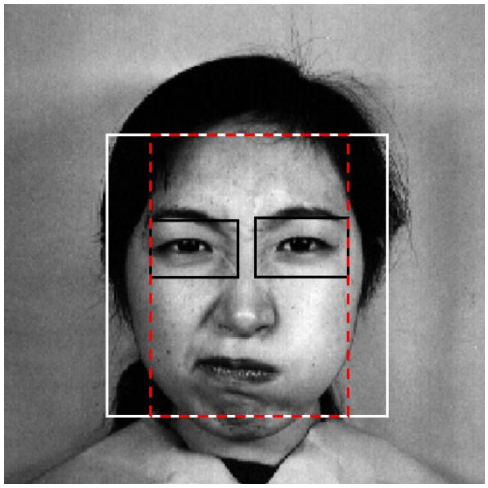


Figura 4.2: Detecção de face.

As regiões dos olhos que são buscados na face definem a região de recorte e também são utilizados para validação. Desta forma, o olho esquerdo deve estar posicionado acima do meio da face e na metade esquerda, enquanto que o olho direito deve estar posicionado acima do meio da face e na metade direta. Quando algum destes critérios não são obtidos a face é descartada, caso contrário é realizado o recorte.

4.2. Extração de Características

Após as faces serem recortadas da imagem original é realizada a extração de características, a qual é responsável por obter informações capazes de descrever os diferentes tipos de expressões faciais. Neste trabalho são utilizados extratores de características baseados em texturas locais. Conforme apresentado na seção 3.2, os extratores baseados em geometria dependem muito da precisão dos pontos faciais e ainda não são robustos o suficiente. Atualmente os métodos baseados em textura tem atraído mais atenção dos pesquisadores por não ser necessário mapear os pontos fiduciais, em vez disso, utiliza um banco de filtros em toda a imagem para extrair as características faciais [15].

Os algoritmos LBP e WLD foram selecionados para extração de características devido ao bom desempenho obtido em trabalhos REF e também pelo custo computacional reduzido (seção 3.2). Ambos os algoritmos são baseados em extração de micro padrões de textura e com isso são capazes de obter detalhes finos da pele, como rugas, vincos e manchas, portanto, a face deve ser zoneada para obter o padrão de cada sub-região. A Figura 4.3 e a Figura 4.4 ilustram as expressões de neutro e raiva de dois sujeitos do conjunto CK divididas em 3×3 sub-regiões,

em que cada linha representa a face de uma expressão facial e cada coluna é uma sub-região. Neste exemplo é possível verificar que as mudanças de textura da face gerada por uma expressão facial são visualmente mais notáveis na área da boca (sub-região 8) e olhos (sub-regiões 1 à 3). E como descrito na seção 2.2 essas variações de texturas podem ser obtidas pelos algoritmos LBP ou WLD. O exemplo também deixa claro a importância da detecção facial, como cada sub-região se torna uma parte do vetor de atributos, um erro na posição ou no alinhamento da face pode fornecer informações irrelevantes (seção 3.1).

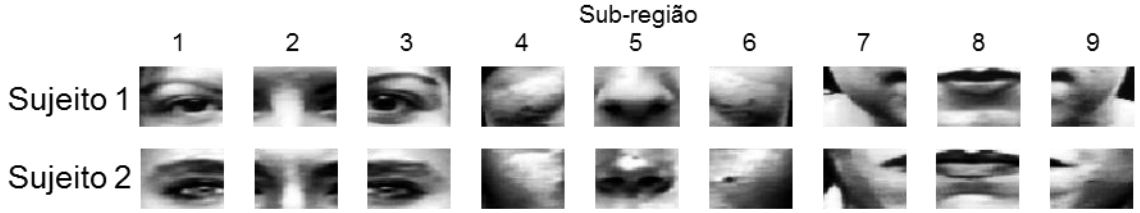


Figura 4.3: Sub-regiões da expressão neutro de dois sujeitos

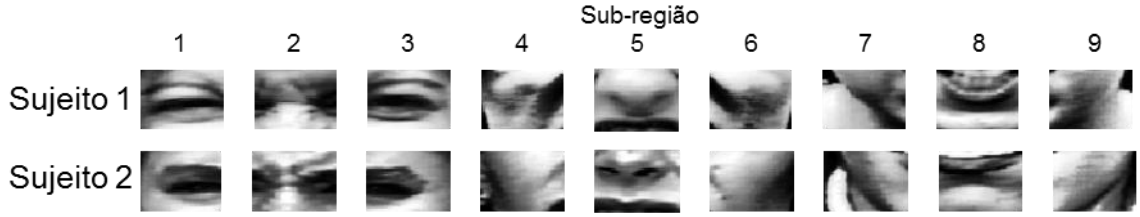


Figura 4.4: Sub-regiões da expressão raiva de dois sujeitos

Com relação as sub-regiões ilustradas na Figura 4.3 e Figura 4.4 é possível fazer uma breve análise das características produzidas pelos algoritmos utilizados na extração de características. As representações obtidas pelo WLD podem ser verificadas na Figura 4.5. É possível visualizar que as sub-regiões 2 (região entre os olhos) e 7 (região à esquerda da boca) identificam melhor as expressões faciais, independentemente do sujeito. Na sub-região 2 é verificado a predominância de dois picos para raiva, enquanto que para a expressão neutro o histograma tende a produzir um comportamento mais uniforme do que a expressão raiva, ainda existe uma diferença de tamanho no primeiro pico destacado, sendo maior para neutro e menor para raiva. Na sub-região 7 as relações entre as expressões faciais podem ser encontradas nas duas colunas à direita, sendo que para neutro a diferença entre as colunas é maior e uma apresenta escala alta, enquanto que para a expressão raiva as duas colunas apresentam escala menores e tamanhos similares.

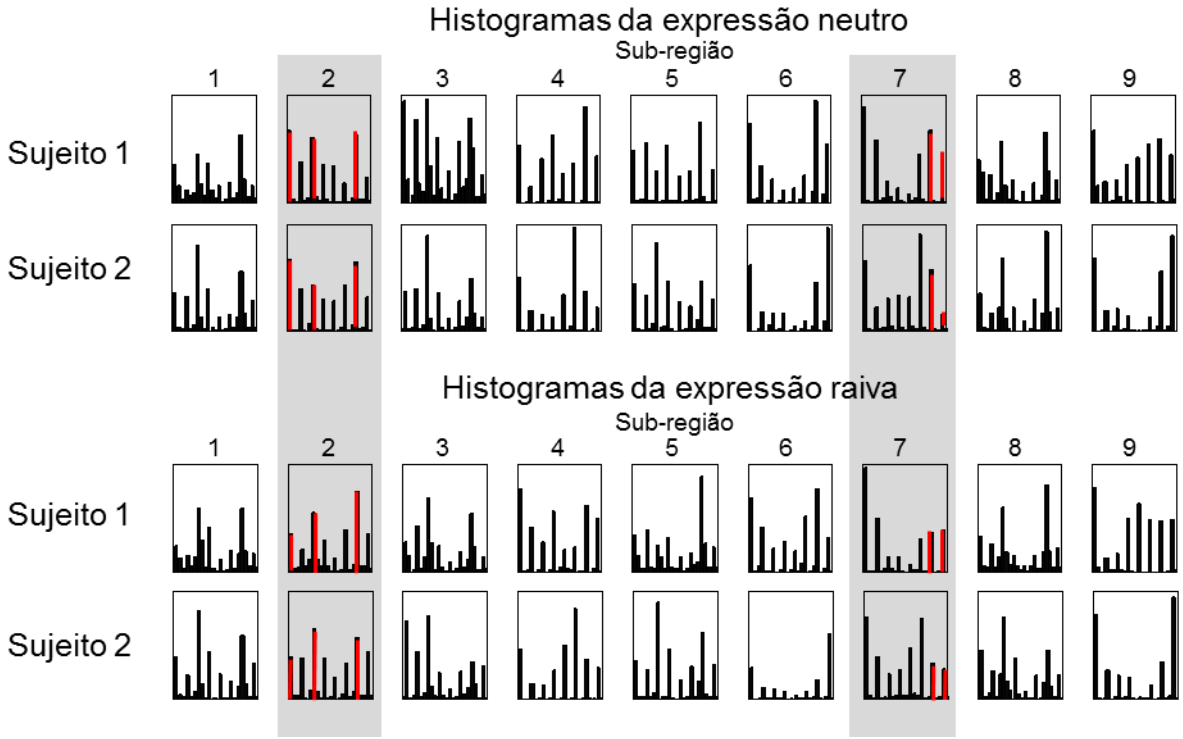


Figura 4.5: Histogramas WLD de dois sujeitos para as expressões de neutro e raiva.

As características obtidas com LBP a partir das expressões faciais da Figura 4.3 e Figura 4.4 são ilustradas na Figura 4.6. As sub-regiões 2 e 6 são mais perceptíveis visualmente para identificar as expressões faciais. Analisando a sub-região 2 é possível verificar que raiva pode ser caracterizada pelos três picos destacados no histograma e que diferem dos picos encontrados na expressão neutro, ainda os respectivos picos destacados à direita resultam em tamanhos similares para neutro, enquanto que para raiva existe uma diferença notável. Também na sub-região 6 é observado a existência de três picos para raiva, o que a diferencia de neutro.

Figura 4.6: Histogramas LBP de dois sujeitos para as expressões de neutro e raiva

O número de zonas ou sub-regiões é difícil de determinar, pois deve equilibrar a representação da face e o consumo de memória e tempo, o que leva à diferentes esquemas de divisão para comparação [90]. Tanto o número de sub-regiões, quanto os parâmetros dos algoritmos para extração de características devem ser estabelecidos a partir do tamanho das imagens que serão utilizadas. Para extração de características em faces de 110×150 pixels alguns trabalhos tem utilizado 6×7 divisões e $LBP_{8,2}^{u2}$ (uniforme, $P = 8, r = 1$), enquanto que imagens de baixa resolução, tamanho inferior à 55×75 pixels, o número de divisões é definido de maneira que cada sub-região possua 10×10 pixels e $LBP_{4,1}$ [45][24]. Assim, o zoneamento deve ser capaz de gerar sub-regiões de tamanho compatível com a configuração dos algoritmos utilizados na extração de características, ou seja, para o LBP e o WLD, o raio de operação e a dimensão do espaço para representação de padrões devem ser adequados ao tamanho de cada sub-região (seção 2.2.1).

A extração de características é aplicada separadamente em cada uma das sub-regiões produzidas pela divisão da face, assim são produzidos $N = Div_{horz} \times Div_{vert}$ subvetores de características f_n (Figura 4.7), em que Div_{horz} e Div_{vert} é o número de divisões horizontais e verticais respectivamente, e $1 \leq n \leq N$ corresponde a n-ésima zona da face. Ao final todos os

subvetores f_n são concatenados para compor o vetor $F = \{f_1, f_2, f_3, \dots, f_N\}$ que representará a face. Desta forma, se o extrator de características utilizado produz k dimensões para cada subvetor f_n, o número total de características para representar uma expressão é de $k \times N$. Assim, a dimensão dos dados é dependente do número de divisões da face e dos parâmetros dos extratores de características.

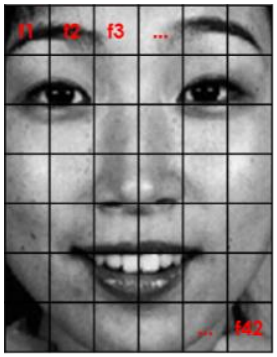


Figura 4.7: Exemplo de divisão facial com 6×7 sub-regiões, são gerados 42 subvetores de atributos

4.3. Redução de Dimensionalidade

Depois de obtidos os vetores de características, é realizada a redução de dimensionalidade visando encontrar os atributos de maior relevância para a aprendizagem dos classificadores. Vários trabalhos da literatura tem realizado a redução de dimensionalidade utilizando todas as expressões faciais [2][25][47], e diante deste fato, busca-se por uma alternativa mais eficiente com a hipótese de que é possível obter um subconjunto de atributos mais discriminantes utilizando a seleção de atributos em pares de expressões faciais e classificação Um-Contra-Um. A motivação é inspirada em recentes trabalhos da literatura [53][2][18] (seção 3.4), em que etapas para reconhecer expressões faciais são divididas em subproblemas menores, para então resolver cada um e produzir a resposta final, semelhante a estratégia dividir em conquistar [91]. Com o uso de pares de expressões faciais para seleção de atributos espera-se que ao final o número de atributos processados seja inferior à abordagem tradicional.

Assim como em Kyperountas et al [18], neste trabalho é avaliada a redução de características utilizando pares de expressões. O número de atributos selecionados é comparado com a redução de atributos utilizando todas as classes (expressões faciais) e sem a redução de atributos, além da avaliação em diferentes técnicas de seleção de atributos. A comparação do

número de atributos selecionados em diferentes abordagens e técnicas não é avaliada por Kyperountas et al [18], também um único conjunto de dados é utilizado para realização dos experimentos. Esses fatos impossibilitam uma conclusão mais precisa a respeito do método.

Na Figura 4.8 o método de seleção em pares de expressão é ilustrado. Dado o vetor de atributos F obtido a partir da extração de características, é produzido um subconjunto de atributos F_p' para cada par de classe p, ou seja, a combinação das 7 possíveis expressões faciais de alegria (A), tristeza (T), raiva (R), medo (M), nojo (D), neutro (N) e surpresa (S), geram 21 subconjuntos $F_p{}'$ em que $p \in \{A \cup T,\ A \cup R,\ A \cup M,\ \ldots, N \cup S\}$. A seleção de características é realizada por uma técnica R, sendo assim $F_p' = R(I_p)$ em que I_p é o conjunto de instâncias contendo somente as expressões do par p e F_p' é o vetor com as características mais discriminantes para as classes de p.

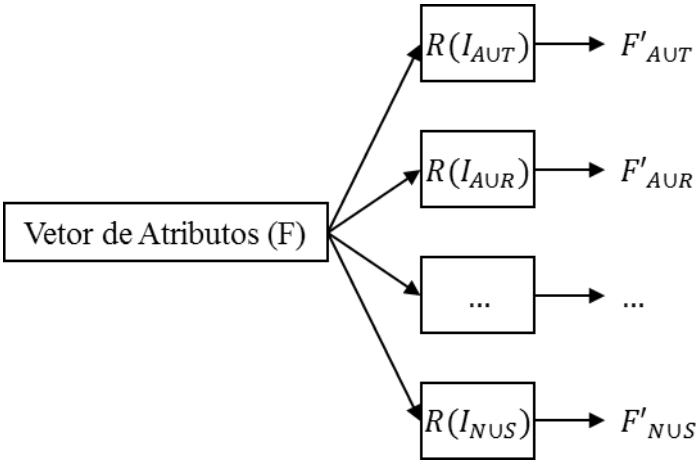


Figura 4.8: Redução de características com pares de expressão

Os resultados obtidos pela seleção em pares são comparados com a seleção tradicional, a qual utiliza todas as 7 expressões faciais ($I_{AUTURUMUDUNUS}$) para determinar o subconjunto F' com atributos mais relevantes (Figura 4.9). Neste caso é produzido um único vetor de dimensão de tamanho reduzido e contendo os atributos mais importantes de acordo com a técnica R. Os resultados também são comparados com o desempenho sem a seleção de atributos (Figura 4.10).

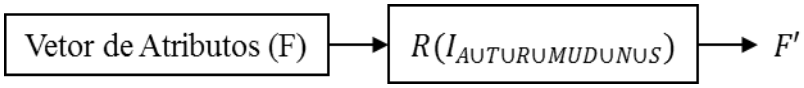


Figura 4.9: Redução de características com todo conjunto

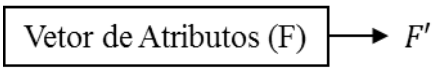


Figura 4.10: Reconhecimento de expressões sem redução de dimensionalidade.

Na literatura pode ser encontrado vários trabalhos que utilizam a PCA para redução da dimensionalidade [25][66][57]. A técnica seleciona os atributos através da correlação entre eles e não considera as classes das amostras, podendo selecionar um subconjunto de atributos que não aumenta a discriminação entre as classes (seção 3.3). Ao contrário da PCA, as técnicas descritas na seção 2.3.2 (CFS, IG e KW) identificam atributos discriminantes através da relação com as classes do problema. Atributos capazes de obter maior distinção entre as classes melhoram o aprendizado do classificador e, consequentemente, o desempenho do método. Com base nos fatos apresentados, para validar as abordagens de seleção de atributos foram utilizadas as técnicas com $R \in \{CFS, IG, KW\}$ em diferentes estratégias para seleção de atributos, quando denominado seleção de atributos com CFS considera-se a estratégia do tipo *filter* e para IG e KW a seleção de atributos é com estratégia do tipo híbrida.

A primeira estratégia é a seleção do tipo *filter* com CFS. Este algoritmo é popular no campo de aprendizagem de máquina e pesquisas tem mostrado ser capaz de selecionar atributos altamente relacionados com a classe [79], além de ser pouco explorado no reconhecimento de expressões faciais [13]. Nesta abordagem os novos subconjuntos de atributos são selecionados de acordo com o objetivo heurístico do CFS. Como explicado anteriormente, para a seleção baseada em pares de expressões são obtidos 21 subconjuntos, ou seja, o CFS é executado uma vez para cada F'_p, enquanto que para seleção considerando todas as classes é obtido apenas um subconjunto F' com os atributos mais relevantes, ou seja, o CFS é executado uma única vez.

A segunda abordagem é considerada um modelo híbrido, pois utiliza as características de *filter* e *wrapper*, e as técnicas avaliadas são IG e KW. A seleção de atributos baseada em IG tem sido aplicada com sucesso no reconhecimento facial e foi extensivamente utilizada no reconhecimento de texto [92][76][75]. O KW recentemente foi aplicado no reconhecimento de expressões faciais [2] e possui baixo custo computacional. As técnicas IG e KW baseada em *filter* são utilizadas para classificar os atributos conforme a relevância para distinção das classes,

mas não fornecem um subconjunto de atributos óptimos. Por outro lado, o método *wrapper* seleciona um subconjunto de atributos óptimos avaliando exaustivamente cada possível atributo em um classificador c', sempre verificando se houve um ganho de desempenho, no entanto requer várias interações para avaliar todo o espaço de atributos. Assim, para obter um subconjunto mais discriminante e com custo computacional menor, neste trabalho foi considerado o modelo híbrido. Nesta estratégia um *filter* é aplicado para ranquear os atributos de um conjunto de características e então são testadas faixas de atributos no classificador desejado de modo a escolher o subconjunto mais discriminante, desta forma é possível avaliar um número menor de interações com o *wrapper*, pois já são utilizados os atributos mais promissores. Como na etapa de classificação serão utilizados os algoritmos SVM e KNN, então $c' \in \{SVM, KNN\}$

De forma mais detalhada, no modelo híbrido os atributos de F são ranqueados de acordo com a heurísticas da técnica R, considerando as respectivas classes caso seja realizada em pares. Com base no ranqueamento são testadas faixas de $min_{atrib} \leq N \leq max_{atrib}$ com os atributos considerados mais discriminantes, em que min_{atrib} é o número mínimo de atributos a serem avaliados e max_{atrib} é o máximo, a cada teste são adicionados inc_{atrib} atributos para avaliação. O subconjunto de cada faixa selecionada é classificado por c'. Ao final é escolhido o subconjunto que produz melhor resultado para c'. Este procedimento é aplicado na seleção em pares para gerar cada um dos 21 subconjuntos F'_p (Figura 4.8) e também para a seleção com todas as expressões (Figura 4.9). O pseudocódigo do Quadro 4.1 apresenta o procedimento do modelo híbrido. Quando utilizado a seleção em pares a entrada do algoritmo receberá somente exemplos contendo duas expressões faciais.

ALGORITMO *seleçãoModeloHíbrido*(I[1..n][1..T], E[1..n], R, C, min_{atrib} , max_{atrib}, inc_{atrib})

// Entrada: instâncias de expressões faciais I com n exemplos e T atributos. Classes das instâncias E. Técnica para redução de dimensionalidade R e classificador C para avaliar o subconjunto de características. Número mínimo min_{atrib} e máximo max_{atrib} de atributos a serem avaliados com incremento inc_{atrib}
// Saída: atributos selecionados

$acerto\_maior \leftarrow 0$;
$atributos\_selecionados \leftarrow []$;

// Ordena os atributos de acordo com a heurística da técnica de seleção, IG ou KW
$atributos\_ordenados \leftarrow seleção\_atributos(R, I, E)$;

for $N \leftarrow min_{atrib}$ **to** max_{atrib}
// Seleciona os N atributos mais discriminantes
$atributos\_discriminantes \leftarrow atributos\_ordenados[1..N]$;

// Gera instâncias somente com os atributos selecionados
$instancias\_atributos \leftarrow I[1..n][atributos\_discriminantes]$;

// Validação cruzada com os atributos mais importantes
$acerto\_maior \leftarrow classifica(C, instancias\_atributos)$;

// Verifica desempenho
if $acerto > acerto\_maior$ **then**
$acerto\_maior \leftarrow acerto$;
$atributos\_selecionados \leftarrow atributos\_discriminantes$;
end if
$N \leftarrow N + inc_{atrib}$;
end for

return $atributos\_selecionados$;

Quadro 4.1: Algoritmo de seleção de atributos com modelo híbrido

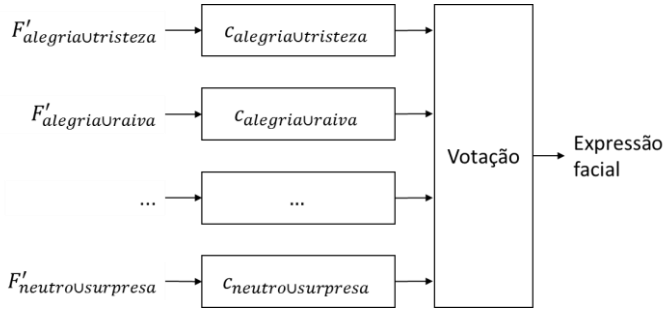
Ao final os resultados obtidos com seleção em pares, seleção com todas as classes e sem seleção são comparados de modo a verificar o número de atributos selecionados e a taxa de acerto produzido por cada estratégia.

4.4. Classificação

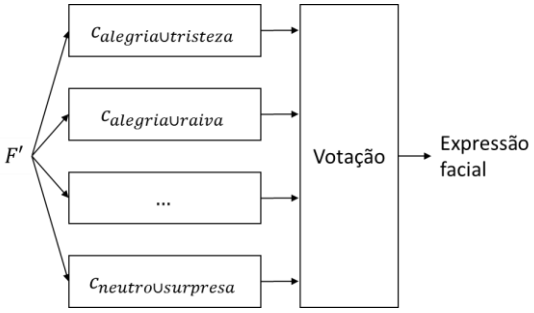
Para reconhecer a expressão de uma face foi utilizada a estratégia de classificação Um-Contra-Um. Nesta abordagem são utilizados 21 classificadores c_p gerados pela combinação das 7 expressões em pares. Cada classificador sempre é treinado com exemplos pertencentes as classes do par p e ao final a predição é dada para classe com mais votos. Quando utilizado

seleção de características em pares, cada classificador c_p recebe como entrada o respectivo subconjunto F_p' (Figura 4.11(a)). Para quando não for utilizada a seleção de atributos (Figura 4.11(c)) ou a seleção de atributos for realizada com todas as classes (Figura 4.11(b)), os classificadores c_p devem respeitar os atributos definidos por F e F' respectivamente.

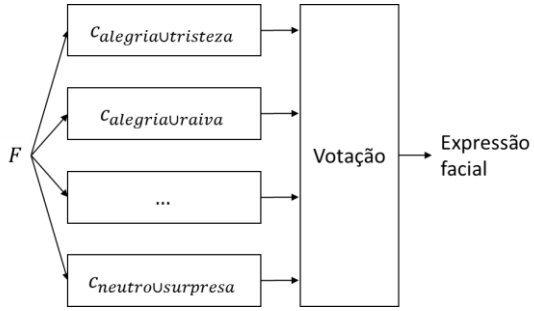
Os atributos provenientes de cada estratégia de seleção de atributos utilizados para o reconhecimento de expressões faciais são avaliados em dois classificadores, a SVM, que possui maior destaque devido ao seu bom desempenho, e o KNN, um algoritmo menos complexo e que também tem sido empregado para reconhecer expressões faciais (seção 3.4).



(a) Classificação utilizando redução de características em pares de expressões faciais.



(b) Classificação utilizando redução de características com todas as expressões faciais



(c) Classificação sem redução de características

Figura 4.11: Estruturas para classificação para as diferentes abordagens de redução de características.

4.5. Considerações Finais

Este capítulo abordou um método com o propósito de tornar mais robusto a escolha das características e classificação Um-Contra-Um contemplando todas as fases necessárias de um sistema para reconhecimento de expressões faciais. Além de utilizar técnicas mais robustas para seleção de atributos, são avaliados dois métodos para redução de características, um do tipo *filter* com CFS e outro do tipo do híbrido com IG e KW, em que são combinados os métodos

filter e *wrapper*. Diante dos métodos discutidos neste capítulo, a seguir é abordado o protocolo experimental para validação da proposta e das hipóteses.

Capítulo 5

Protocolo Experimental

A partir dos métodos apresentados no capítulo anterior para reconhecer expressões faciais, o presente capítulo descreve os procedimentos experimentais, tal como os ajustes dos parâmetros das técnicas utilizadas, seguidos para obtenção dos resultados, que posteriormente são utilizados com o intuito de validar as hipóteses. A Figura 5.1 ilustra um mapa geral dos experimentos executados.

A validação do método proposto é realizada separadamente em 3 conjuntos de imagens, a JAFFE, CK e TFEID. Para as imagens dos conjuntos JAFFE e CK é aplicada a detecção facial para obter a face dos indivíduos presentes nas imagens e remover o fundo, esta etapa não é necessária no conjunto TFEID por já possuir as faces extraídas e prontas para serem processadas. Para todas as faces detectadas e o conjunto TFEID é aplicado a equalização do histograma visando melhorar o contraste. Em seguida é realizada a extração de características, portanto, as imagens das faces são divididas sub-regiões e técnicas como LBP ou WLD são aplicadas para produzir o vetor de características F.

A etapa seguinte é a redução de dimensionalidade, em que são aplicadas as técnicas IG, KW e CFS em diferentes estratégias nos vetores de atributos F produzidos pelo LBP e WLD. Na redução de dimensionalidade em pares de expressões faciais, a partir do conjunto de atributos F fornecido pela extração de características são derivados 21 subconjuntos $F_p{}'$ de dimensão menor, para isso as faces obtidas de cada conjunto de imagens são agrupadas conforme a expressão. Desta forma para obter um conjunto F_p' são escolhidos os grupos referentes ao par da classe p e então é realizada a seleção de atributos. Para melhor compreensão do número de atributos selecionados, na etapa de redução em pares de expressões, será considerado a média aritmética do tamanho dos subconjuntos F_p'.

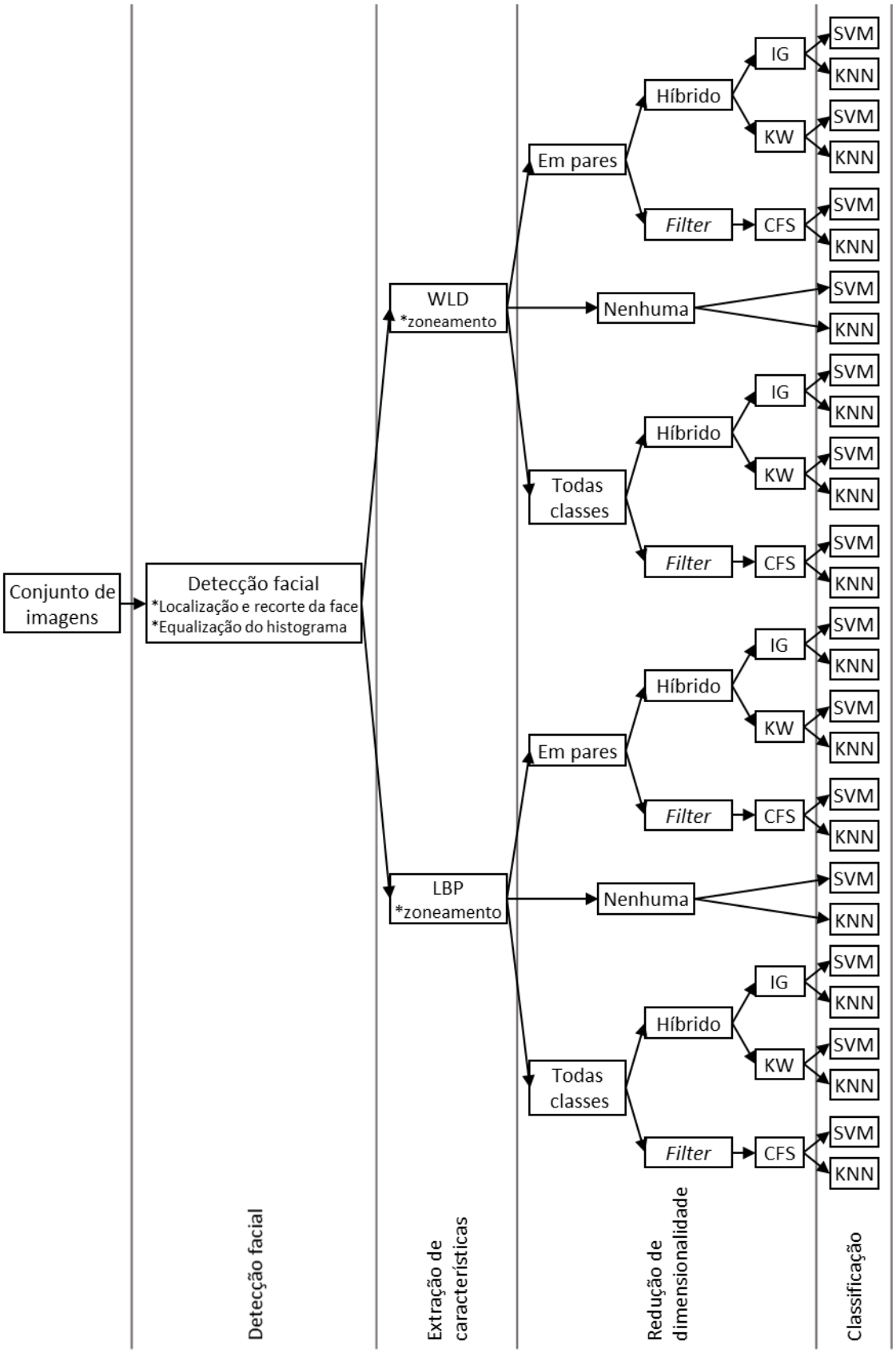


Figura 5.1: Visão geral dos experimentos realizados

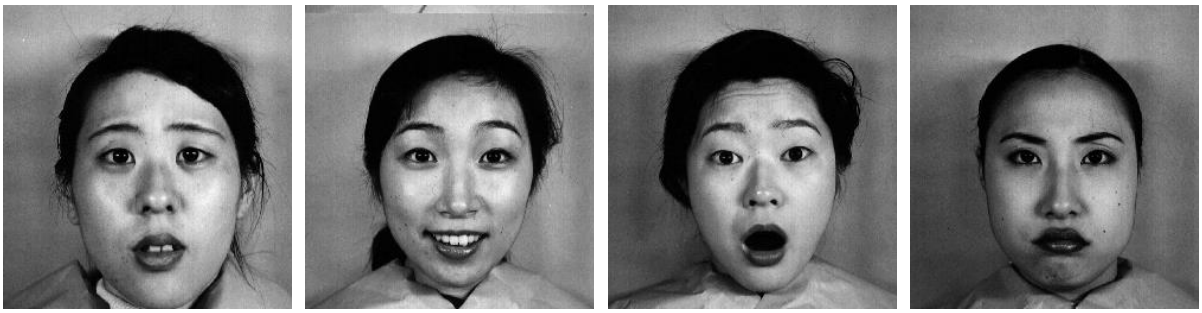
Na última etapa os atributos das faces são classificados com os algoritmos KNN e SVM em estratégia Um-Contra-Um utilizando validação cruzada. Desta forma, como são utilizadas 7 expressões faciais existem 21 combinações de pares de expressões faciais, assim são gerados 21 classificadores especializados em 2 expressões e a predição final é dada para a classe com mais indicações.

Cada uma das técnicas e fases abordadas na proposta possuem parâmetros que devem ser ajustados para reconhecer de expressões faciais. As subseções seguintes apresentam os ajustes realizados para a execução deste experimento, assim como a descrição dos conjuntos de imagens utilizados.

5.1. Conjunto de Dados

Existem poucos trabalhos de reconhecimento de expressões faciais que avaliam o método em diferentes cenários [17][16], a maioria utiliza apenas um conjunto de imagens, geralmente a JAFFE ou a CK [2][14][57]. Para obter uma avaliação mais precisa do desempenho dos métodos implementados, foram considerados os conjuntos de imagens JAFFE, CK e TFEID, sendo os dois primeiros muito populares para o reconhecimento de expressões faciais e o terceiro tem sido mais utilizado para reconhecimento facial.

A *Japanese Female Facial Expression* (JAFFE) [27] (Figura 5.2) é um conjunto composto por 10 mulheres japonesas com 3 à 4 imagens em escala de cinza para cada uma das 6 expressões consideradas universais (alegria, tristeza, raiva, surpresa, nojo e medo) mais a expressão neutro, somando um total de 213 imagens. A resolução das imagens é de 256×256 pixels.



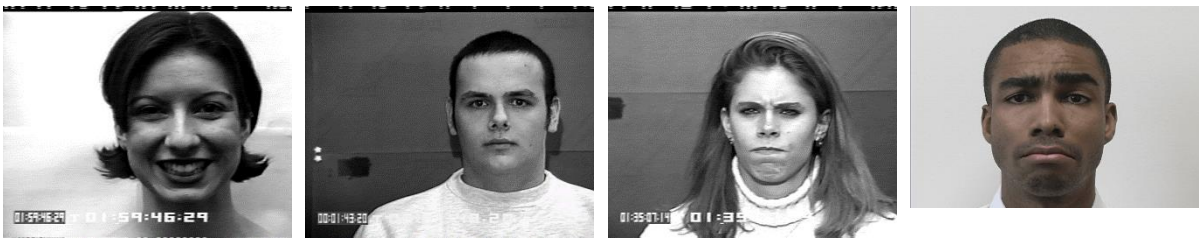
Medo Alegria Surpresa Raiva

Figura 5.2: Exemplos de sujeitos do conjunto JAFFE

O conjunto *The Extended Cohn-Kanade Database* (CK) [93] (Figura 5.3) possui imagens de 123 pessoas entre 18 e 50 anos, sendo que 69% são mulheres, 81% são euro-americano, 13% são afro-americano e 6% pertencem a outros grupos. As imagens do conjunto possuem tamanho de 640×490 pixels e 640×480 pixels com 256 níveis de cinza e 24 bits de tons coloridos respectivamente. As expressões faciais de um sujeito são compostas por sequências de imagens, que partem do neutro até atingir a respectiva expressão de maior intensidade. Neste conjunto somente será utilizada a última imagem de cada sequência e com rótulo fornecido pelo autor do conjunto, uma vez que somente algumas sequências de imagens possuem classe conhecida (Tabela 5.1). Para compor o conjunto de neutro será considerada a primeira amostra da sequência de cada indivíduo que possuir pelo menos uma sequência rotulada. Assim, para este conjunto de dados serão utilizadas 427 imagens.

Tabela 5.1: Composição do conjunto CK

Expressão facial	**Sequências rotuladas**	**Proporção (%)**
Raiva	45	11
Nojo	59	14
Medo	25	6
Alegria	69	16
Tristeza	28	7
Surpresa	83	19
[3]Neutro	118	28



Alegria Neutro Raiva Tristeza

Figura 5.3: Exemplos de sujeitos do conjunto CK

[3] Número de sujeitos que possuem pelo menos uma sequência com classe conhecida

Por fim, o conjunto de dados *Taiwanese Facial Expression Image Database* (TFEID) [28] contendo 40 indivíduos, sendo 20 mulheres, com expressões de alegria, tristeza, raiva, nojo, medo, surpresa, desprezo e neutro. As imagens deste conjunto apresentam 480×600 pixels em tons coloridos de 24 bits. Assim como vários trabalhos da literatura, o escopo deste trabalho compreende somente as expressões universais e a expressão neutro, portanto a expressão de desprezo não foi incluída na validação. Neste conjunto as faces já estão localizadas e alinhadas (Figura 5.4), não necessitando da etapa de detecção facial, apenas de transformação para escala de cinza e equalização do histograma.

Nojo Alegria Medo Raiva

Figura 5.4: Exemplos de sujeitos do conjunto TFEID

5.2. Detecção Facial

Para detecção da face e dos olhos foi utilizada a implementação do algoritmo Viola-Jones disponível no MATLAB, sendo definida pelos seguintes parâmetros de ajustes:

T_{min}: tamanho mínimo do objeto a ser detectado;

T_{max}: tamanho máximo do objeto a ser detectado;

F_e: fator de escala para redimensionamento da janela de varredura. Padrão 1.1;

D_s: mínimo de detecções sobrepostas necessárias para definir a detecção de um objeto. Padrão 4.

Quando procurado por uma face em uma imagem de altura I_{alt} e largura I_{larg}, $T_{max} = I_{larg} \times I_{alt}$ e $T_{min} = \frac{T_{max}}{3}$, os valores foram assim definidos, pois nos conjuntos de imagens utilizados a face é predominante. Ao se buscar pelos olhos em uma face com $F_{larg} \times F_{alt}$

pixels, o tamanho máximo da janela de varredura é dado por $T_{max} = \frac{F_{larg}}{2} \times \frac{F_{alt}}{2}$, enquanto que T_{min} foi mantido o tamanho mínimo permitido de 12×18 pixels. O valor de F_e foi mantido padrão.

Como o detector pode retornar mais de uma região por imagem devido aos falsos positivos, o parâmetro D_s é ajustado automaticamente de modo a produzir o menor número possível de saídas, ou seja, enquanto houver mais de uma face detectada por imagem o valor de D_s é incrementado em 1, e para os olhos é seguido o mesmo protocolo. Os valores de incremento D_s foram determinados empiricamente. Um incremento baixo pode demorar para detectar o objeto desejado, enquanto que um valor de incremento alto pode deixar o detector com sensibilidade muito baixa reduzindo o número de detecção. Uma busca mais eficiente poderia ser empregada, mas isso não está dentro do escopo do trabalho.

O treinamento dos classificadores fracos do AdaBoost utilizados para detecção facial é realizado com imagens da própria *framework* com 24×24 pixels e são utilizadas características extraídas com o LBP, que são robustas a mudança de luminosidade [54]. Os classificadores para detecção dos olhos são treinados com características Haar e imagens de tamanho 12×18 pixels.

5.3. Extração de Características

Os trabalhos para reconhecer expressões faciais e que utilizam características baseadas em textura normalmente reduzem as imagens dos conjuntos de dados para um tamanho padrão de modo a reduzir o custo computacional. Uma comparação prática pode ser avaliada reduzindo uma imagem em 50% do tamanho original. Desta forma a largura e a altura são reduzidas pela metade, assim em termos de tamanho é possível dizer que a extração de características é aplicada em somente 25% da imagem original. Para validação do método, Yan et al. [94] utilizou três conjuntos de dados com imagens de tamanhos 256×256, 640×490 e 320×240 pixels, sendo que para a extração de características com PCA e LDA todas as imagens são ajustadas para 64×64 pixels. No entanto, trabalhos que utilizam LBP e WLD tem redimensionado as faces dos conjuntos JAFFE e CK para tamanhos como 128×128 [16], 126×104 [53] e 150×110 [45][15] pixels. Portanto, neste experimento foi considerado um valor médio de 134×114 pixels para ajuste do tamanho das imagens. A utilização de um tamanho aproximado permite ajustar os extratores de características com os mesmos parâmetros.

No trabalho de Bashar et al. [15] é avaliado a influência do número de sub-regiões em diferentes algoritmos de extração de características, entre eles o $LBP_{8,2}^{u2}$, para reconhecer expressões faciais. E utilizando faces ajustadas para 110×150 pixels foi constatado melhor desempenho para divisão em 6×7 zonas, os resultados dos experimentos são apresentados na Tabela 5.2 e Tabela 5.3. Deve-se avaliar que a medida que o número de divisões aumenta, o número de atributos também eleva. Considerando um exemplo com $LBP_{8,2}^{u2}$ que produz 59 dimensões, ao dividir uma imagem em 3×3 são produzidos $3 \times 3 \times 59 = 531$ atributos, aumentando a divisão para 6×7 são obtidos 2478 atributos. Neste experimento, para a extração de características com o LBP, a face foi zoneada em 6×7 sub-regiões e também foi utilizado o $LBP_{8,2}^{u2}$, logo o vetor de atributos que descreve uma expressão facial possui $6 \times 7 \times 59 = 2478$ dimensões.

Tabela 5.2: Taxa de reconhecimento para 6 expressões faciais obtidos por [15] com diferentes divisões faciais

Extrator de características	$\mathbf{3 \times 3}$	$\mathbf{5 \times 5}$	$\mathbf{6 \times 7}$
MTP	95,2	97,5	98,1
MBP	80,2	87,3	93,1
LTP	91,3	92,3	94,6
LBP	79,1	89,7	90,1

Tabela 5.3: Taxa de reconhecimento para 7 expressões faciais obtidos por [15] com diferentes divisões faciais

Extrator de características	$\mathbf{3 \times 3}$	$\mathbf{5 \times 5}$	$\mathbf{6 \times 7}$
MTP	89,1	92,4	94,2
MBP	77,2	85,3	90,1
LTP	87,3	89,3	91,6
LBP	75,1	84,7	86,1

Em Hussain et al. [2] foi utilizado o WLD com diferentes parâmetros, $5 \leq T \leq 8$ e $4 \leq N \leq 6$, em imagens de tamanho 256×256 pixels, no entanto não é apresentado o desempenho por cada conjunto de parâmetros e nem o número de divisões da face. No estudo Shuaishi et al. [53] é apresentado a taxa de reconhecimento de expressões faciais com $T = 8$ e diferentes

valores de N. As características extraídas são classificadas com SVM e KNN. Para validação, as imagens do conjunto JAFFE são redimensionadas para 126×104 pixels e é utilizado um esquema de divisão facial com 6×9 sub-regiões. Os resultados ilustrados na Figura 5.5 demonstram uma maior precisão com $N = 5$ e $T = 8$. Portanto, para a extração de características com o WLD foi seguida a mesma divisão de face de Shuaishi et al. [53] e também os parâmetros do extrator de características foram ajustados para $N = 5$ e $T = 8$, resultando em um vetor de atributos com $6 \times 9 \times 5 \times 8 = 2160$ dimensões.

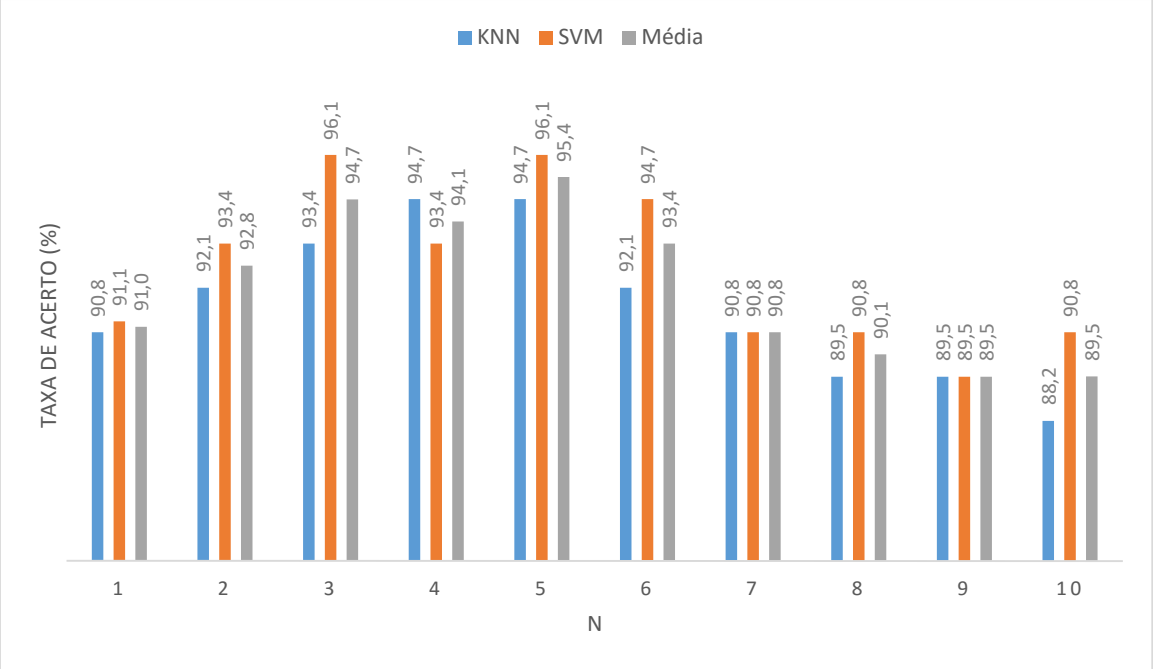


Figura 5.5: Taxa de reconhecimento facial com WLD para diferentes valores de N

Após obter as características das faces, os atributos são normalizados entre 0 e 1 por recomendação das bibliotecas utilizadas na etapa de classificação.

5.4. Redução de Dimensionalidade

Para redução de dimensionalidade utilizando a estratégia híbrida é necessário estabelecer a faixa de atributos a ser explorada (seção 4.3), tal como a configuração dos classificadores para avaliação dos subconjuntos. Considerar todo o conjunto de dados pode demandar muito tempo e um número pequeno pode não fornecer informações necessárias para aprender o modelo de classificação. Portanto, trabalhos relacionados ao reconhecimento de

expressões faciais e que utilizam características baseadas em textura foram utilizados para definir uma faixa adequada.

Em Hussain et al [2] foi utilizado WLD e LTP em imagens com tamanho 256×256 pixels, sendo que cada um dos extratores de características foram ajustados para gerar vetores com 30, 40 e 50 atributos. Uma comparação (Figura 5.6) classificando cada um dos conjuntos de atributos separadamente demonstrou que quanto maior o número de atributos melhor é o desempenho, ou seja, o conjunto de 40 atributos produz resultados superiores a 30 atributos, mas inferior a 50 atributos, no entanto a diferença entre 40 e 50 atributos é menor. Em seguida as características extraídas com o WLD e o LTP são concatenadas produzindo conjuntos com 60, 80 e 100 atributos. Então para obter as características mais discriminantes é utilizado o KW e ao final são obtidos subconjuntos com 40, 60 e 80 atributos que permitem obter maior reconhecimento de expressões faciais.

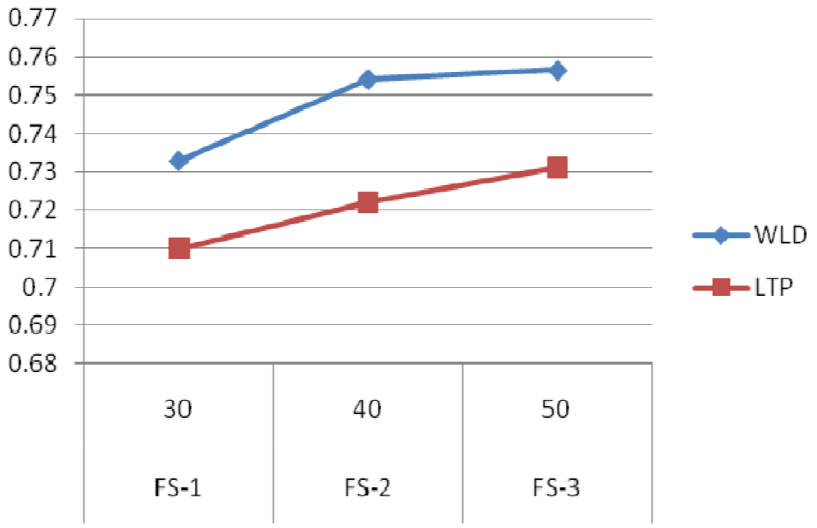


Figura 5.6: Comparação de desempenho para subconjuntos de atributos com diferentes tamanhos [2].

No trabalho de Zhang et al. [13] são utilizados LBP, Gabor Filter e SIFT para extração de características produzindo 2597, 2120 e 6784 atributos respectivamente. As características extraídas são avaliadas em diferentes estratégias para classificação. A redução de dimensionalidade é avaliada em uma faixa de 1 à 300 atributos. Os resultados (Figura 5.7) demonstram que a medida que o número de atributos aumenta a taxa de erro diminui, no entanto, a diferença de desempenho entre o uso de 150 e 300 atributos é menor que 3%. Avaliando a mesma faixa de atributos, no estudo de Huang et al. [61] é realizada a redução de

dimensionalidade de 1024 atributos com SNE em imagens de 32×32 pixels, os resultados mostram que os melhores resultados para o reconhecimento de expressões faciais são obtidos com 70 à 200 atributos.

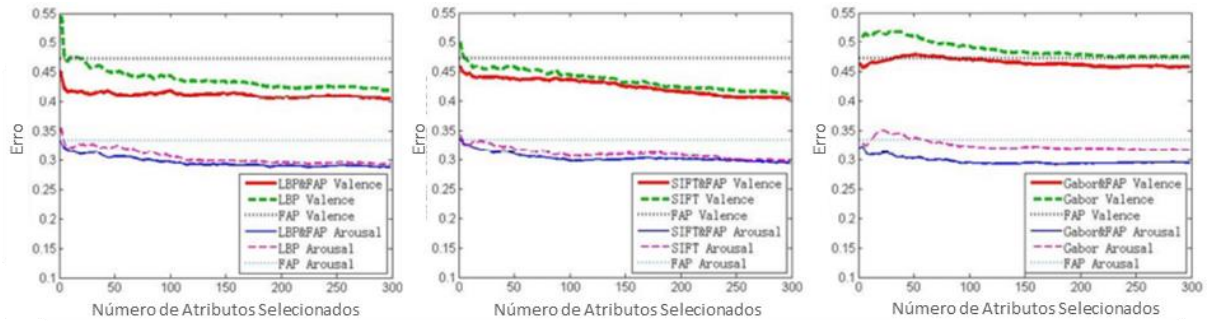


Figura 5.7: Erro no reconhecimento de expressões faciais para subconjuntos de atributos com diferentes dimensões [13]

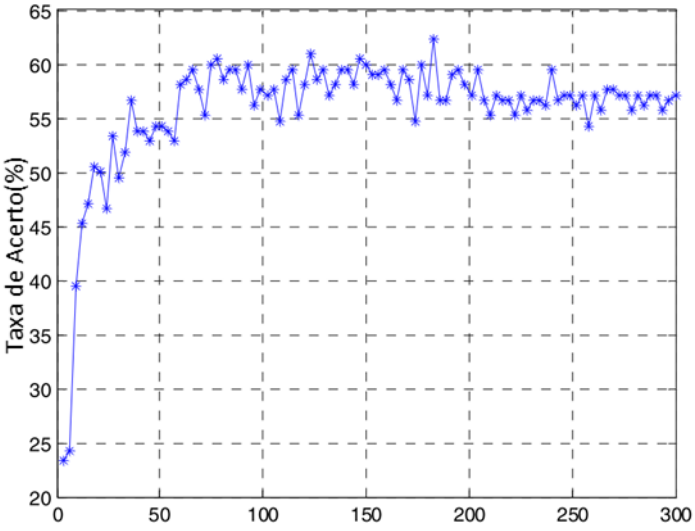


Figura 5.8: Taxa de reconhecimento de expressão obtido por [61]

Zhang et al. [3] conduziram um estudo para reconhecer expressões faciais utilizando faces de 110×150 pixels. Para extração de características foi utilizado o Gabor Filter, produzindo 660000 atributos. A seleção de atributos foi avaliada em uma faixa de 10 à 100 atributos com 10 intervalos e verificou-se que as maiores taxa de acerto são obtidas com 50 à 70 atributos.

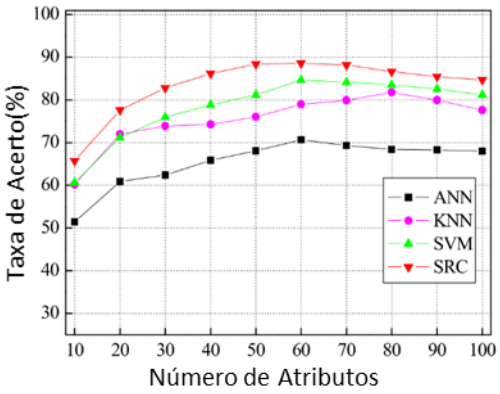


Figura 5.9: Desempenho obtido por [3] para vetores de atributos com diferentes dimensões

Conforme pode ser verificado na literatura (Tabela 5.4), o número de atributos avaliados não segue um padrão ou um critério em relação ao tamanho original dos dados. No entanto, o trabalho de Zhang et al. [13] (Figura 5.7) fornece uma base mais sólida para a escolha da faixa de atributos, pois seus experimentos são conduzidos em características obtidas de diferentes extratores e estratégias. Ainda Zhang et al. [13] utiliza técnicas e vetores de atributos com tamanho semelhantes a este experimento. Portanto, foi explorado uma faixa de $20 \leq N \leq 300$, com $N = N + 20$ a cada interação, assim $min_{atrib} = 20$ e $max_{atrib} = 300$. Para as etapas de treinamento e testes foram utilizadas todas as instâncias disponíveis para selecionar os atributos mais promissores.

Tabela 5.4: Resumo de trabalhos com as respectivas faixas de atributos avaliados para redução de dimensionalidade.

Trabalho	**Extração de características**	**Total de Atributos**	**Faixa de atributos avaliada**	**Atributos selecionados**
Hussain et al [2]	WLD e LTP	60,80,100	60,80,100	40,60,80
Zhang et al. [13]	LBP	2597	1 à 300	300
	Gabor	2120		
	SIFT	6784		
Huang et al. [61]	SNE	1024	1 à 300	70 à 200
Zhang et al. [3]	Gabor	660000	10 à 100	Entre 60 e 80

Ainda na seleção de atributos com estratégia híbrida são utilizados os classificadores SVM e KNN para avaliar a capacidade de classificação de cada subconjunto de atributos selecionado dentro da faixa *N*. Para isso é utilizada a validação cruzada com 10 partições, similar ao trabalho de Zavaschi et al [17] e Bashar et al. [15], e também é foi usado o *grid search* para encontrar os melhores valores dos parâmetros da SVM, sendo adotado $2^{-5} < C < 2^{15}$ e $2^{-15} < \gamma < 2^{3}$. Para o KNN foi explorado $1 < k < 131$ utilizando a distância Euclidiana.

Para a abordagem do tipo *filter* com CFS, ao contrário de Zhang et al. [13] que utilizou como critério de parada a seleção de 300 atributos, neste estudo assim como outros que utilizaram CFS [95][81], o *best first search forward* foi aplicado para buscar por subconjuntos mais discriminantes com critério de parada de 5 atributos que não melhoram o mérito do subconjunto.

5.5. Classificação

A partir dos atributos obtidos na extração de características ou selecionados na redução de dimensionalidade são utilizados a SVM e o KNN para classificar os exemplos. Como descrito na seção 4.4, são gerados 21 classificadores binários especializados em duas expressões faciais. Para validação dos classificadores foram utilizadas as mesmas configurações dos classificadores descrito na seção 5.4, ou seja, foi utilizado validação cruzada com 10 partições, *grid search* para encontrar os melhores parâmetros da SVM com $2^{-5} < C < 2^{15}$ e $2^{-15} < \gamma < 2^{3}$ e o KNN foi avaliado com $1 < k < 131$.

5.6. Considerações Finais

Com base na proposta descrita anteriormente, este capítulo abordou como os experimentos foram executados para validação dos métodos e hipóteses, na sequência são apresentados e discutidos os resultados alcançados.

Capítulo 6

Resultados e Discussão

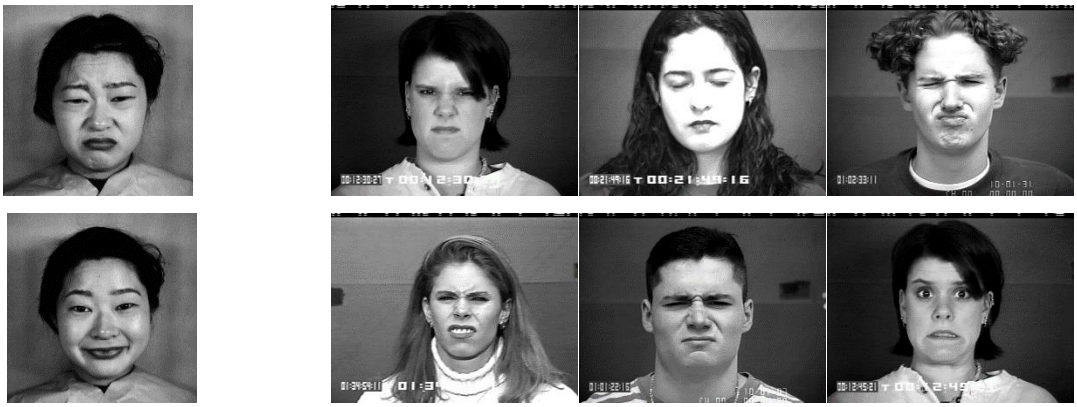
No capítulo anterior foram descritos como os experimentos foram executados e neste capítulo são apresentados e discutidos os resultados obtidos pela fase de detecção facial, extração de características, redução de dimensionalidade e classificação. Para comparar os resultados obtidos foram utilizados teste estatísticos. Primeiramente é aplicado o teste de Shapiro-Wilk de modo a verificar a normalidade. Quando os dados pertencem a uma distribuição normal a comparação é realizada com o teste paramétrico t-Student, caso contrário, será utilizado o teste não paramétrico U de Mann-Whitney. Por ser um valor comum na literatura, para todos os testes será utilizado o nível de significância com $\alpha = 0{,}05$. O software SPSS da IBM foi utilizado para computar as estatísticas.

6.1. Detecção facial

Com a detecção facial implementada foi possível encontrar 99% das faces no conjunto JAFFE e para 97% para CK (Tabela 6.1). Em ambos os conjuntos de dados a detecção falhou na localização de algum dos olhos. Para JAFFE somente 2 faces não foram encontradas, e como mostrado na Figura 6.1 (a) os fortes traços orientais do sujeito podem ter prejudicado na detecção. No conjunto CK existem 12 faces que não foram detectadas e como ilustrado na Figura 6.1 (b) alguns fatores podem ter impedido a localização, como cabelos próximos aos olhos e o estado dos olhos (fechados ou semifechados).

Tabela 6.1: Desempenho da detecção de face

Conjunto de dados	**Total**	**Detectadas**	**Taxa de faces detectadas**
JAFFE	213	211	99%
CK	427	415	97%

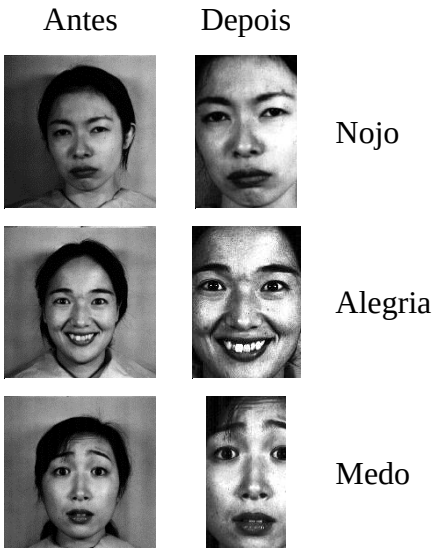


(a) JAFFE (b) CK

Figura 6.1: Exemplo de faces não detectadas

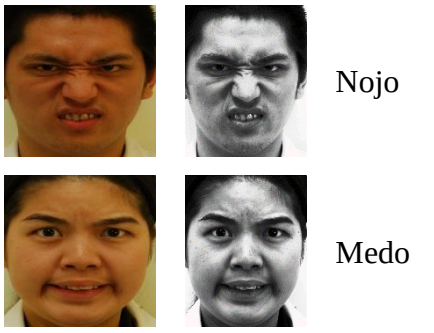
Na literatura dificilmente é apresentado o método de detecção facial, assim como o número de faces encontradas. Também não é especificado se as taxas de acerto obtidas pelos método consideram todas as faces do conjunto de dados, ou somente aquelas que tenham sido detectadas corretamente [14][36][16].

A Figura 6.2 ilustra alguns exemplos de imagens antes e após a detecção da face. Pode-se verificar que grande parte da informação desnecessária foi removida dando foco na região da face. Devido a variação no formato do rosto de cada indivíduo, algumas faces apresentam grandes variações no enquadramento do recorte, como a face de medo do conjunto JAFFE (Figura 6.2(a)), onde o corte é realizado na metade do olho direito, e a face de nojo do conjunto CK (Figura 6.2(b)), em que a região do olho esquerdo está mais afastada da margem em relação as demais faces.



(a) JAFFE

(b) CK



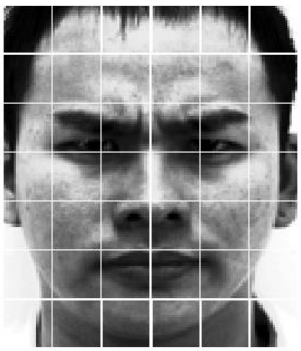
(c) TFEID

Figura 6.2: Faces antes e após a extração com equalização do histograma

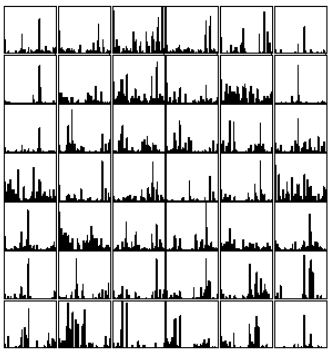
6.2. Extração de características

Nesta etapa são obtidas as características das faces. Conforme o protocolo experimental descrito na seção 5.3, para a extração de características as faces são ajustadas para 134×114 pixels e então divididas em sub-regiões de acordo com o algoritmo de extração de utilizado. Para o LBP a face é dividida em 6×7 sub-regiões e produz 2478 atributos, enquanto que para o WLD a face é dividida em 6×9 regiões e produz 2160 atributos.

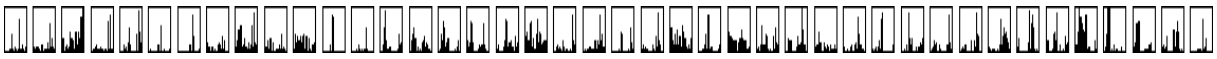
A Figura 6.3 e a Figura 6.4 apresentam as características geradas pela extração de características utilizando o LBP e WLD respectivamente. Depois de aplicado o zoneamento da face (Figura 6.3(a) e Figura 6.4(a)) são obtidas as características de cada sub-região (Figura 6.3(b) e Figura 6.4(b)), que ao final são concatenados para formar o vetor de características da imagem, como é ilustrado na Figura 6.3(c) e Figura 6.4(c).



(a) Face com zoneamento 7×6

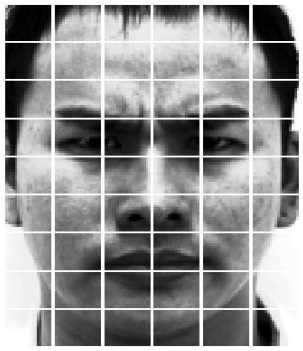


(b) Histograma de cada zona

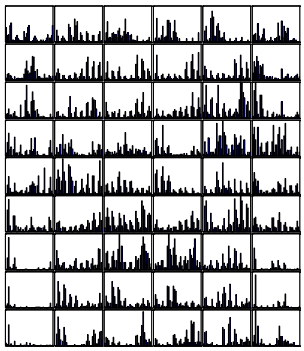


(c) Histogramas concatenados

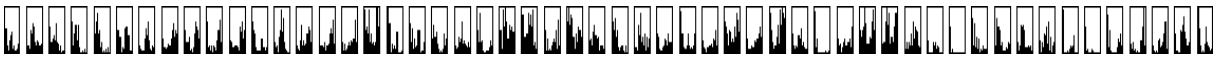
Figura 6.3: Extração de características com LBP



(a) Face com zoneamento 9 × 6



(b) Histograma de cada zona



(c) Histogramas concatenados

Figura 6.4: Extração de características com WLD

Na Figura 6.5 são ilustradas as faces de dois sujeitos do conjunto TFEID, cada um expressando surpresa, raiva e alegria. Os respectivos histogramas, ou vetor de características, escalados entre 0 e 1 de cada sujeito obtido com o LBP são ilustradas na Figura 6.6 e com o WLD na Figura 6.7. Algumas variações que podem ser visualmente encontradas entre as diferentes expressões faciais são representadas pelas regiões I, II, III e IV.

Figura 6.5: Expressão de surpresa, raiva e alegria de sujeitos da TFEID

Para a extração de características com o LBP, na região I é verificado uma menor ocorrência de padrões na expressão alegria, enquanto que para as demais expressões a contagem de padrões é mais perceptível. Na região II, o atributo destacado ocorre com maior escala para

a surpresa e alegria, para raiva o atributo é de menor valor. Semelhante à região II, na região III também existe um atributo de maior valor, mas para alegria. Por fim, pode ser encontrada uma ampla faixa com atributos de alta escala na região IV para alegria, enquanto que para a expressão de raiva e surpresa os atributos estão concentrados em uma faixa mais estreita.

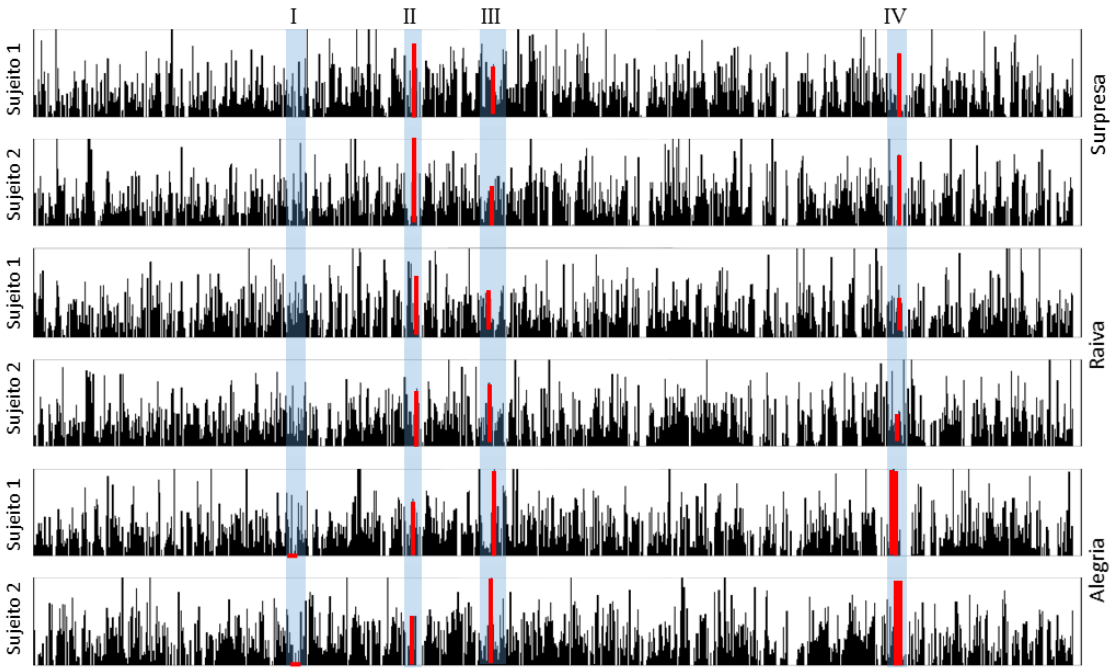


Figura 6.6: Vetor de características obtidos com LBP

As características obtidas com WLD apresentam regiões discriminantes em posições diferentes da extração com LBP. Na região I as maiores escalas podem ser encontradas na expressão surpresa, enquanto que para raiva e alegria os valores são visivelmente menores. A expressão de raiva e alegria na região II possuem atributos com valores semelhantes, mas para surpresa os atributos são de escala menor. A região III fornece informações capaz de diferenciar os três tipos de expressões, em que surpresa possui atributos com valores máximo, alegria é definido por atributos menores que 1,0 e maiores que 0,5, e os atributos de raiva são inferiores à 0,5. Por fim, na região IV é possível verificar maiores valores para alegria e menores para raiva, para surpresa os valores são intermediários com relação as demais expressões.

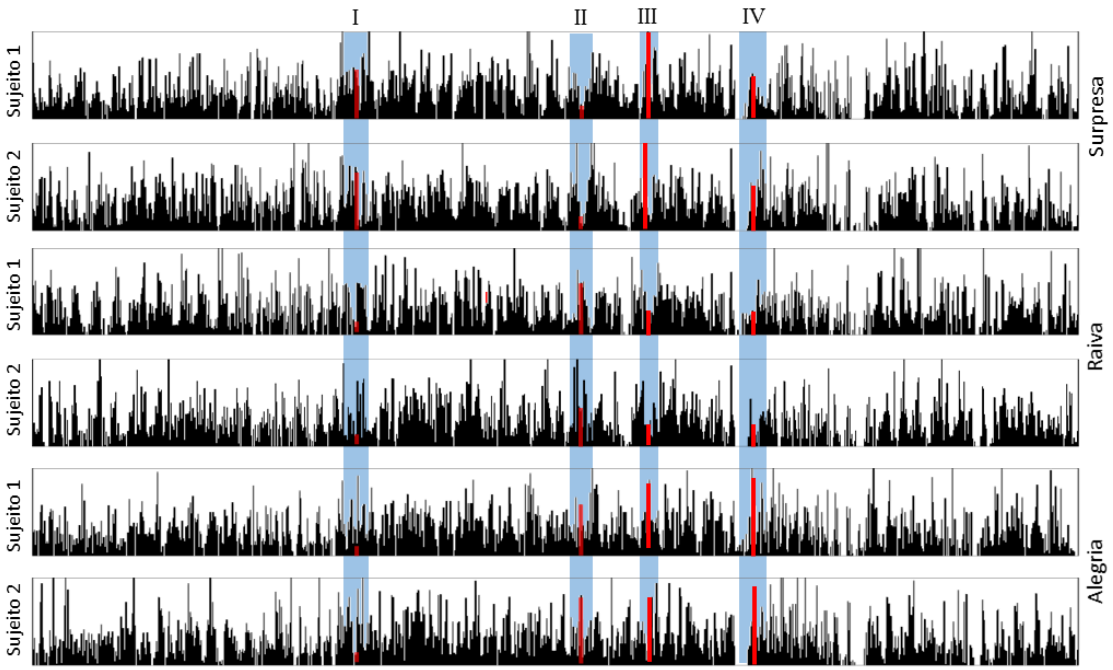


Figura 6.7: Vetor de características obtidos com WLD

As diferenças apresentadas nos histogramas da Figura 6.6 e da Figura 6.7 e destacadas pelas regiões I, II, III e IV, demonstram que apesar da dificuldade em identifica-las visualmente, a extração de características com o LBP e o WLD fornecem informações para distinção entre os diferentes tipos de expressões faciais.

6.3. Redução de Dimensionalidade

Com as características das faces extraídas o passo seguinte é reduzir a dimensionalidade retendo os atributos mais discriminantes. Sem a seleção de características a dimensionalidade é alta, com o LBP são produzidos 2478 atributos e 2160 com o WLD.

Conforme descrito na seção 4.3 foram seguidas duas estratégias para redução de atributos, uma utilizando todas as classes e outra com pares de expressões faciais, sendo a primeira baseada em *filter* e a segunda uma combinação híbrida de *wrapper* e *filter*. Na estratégia híbrida os atributos são classificados por um *filter* (IG ou KW) em ordem de importância, então subconjuntos com os atributos mais discriminantes são avaliados por um classificador, neste caso por uma SVM ou um KNN. O subconjunto de melhor desempenho é selecionado.

Os subconjuntos de atributos selecionados são utilizados e processados na etapa de classificação (seção 4.4). Desta forma, o subconjunto obtido através da seleção de atributos com

todas as classes é utilizado por cada um dos 21 classificadores binários. Para a redução de atributos em pares, cada um dos 21 subconjuntos obtidos são aplicados em um respectivo algoritmo de aprendizagem de máquina da estratégia de classificação Um-Contra-Um.

Na Tabela 6.2 é apresentado o número de características obtidos através da redução de dimensionalidade com o uso de todas as expressões faciais e, corresponde a dimensão do subconjunto F' utilizado como entrada para cada um dos 21 algoritmos da etapa de classificação (Figura 4.11(b)). Na estratégia híbrida com IG e KW são apresentados o número de atributos que produzem melhor desempenho pelos classificadores avaliados, que como descrito na seção 4.3 os atributos que produzem maior taxa de acerto devem ser retidos. O número médio de atributos selecionados utilizando todas as classes foi de 199, sendo 300 o número máximo de atributos retidos e 58 o mínimo. Para a redução de características com *filter* foram selecionados em média 113 atributos, enquanto com a redução de características híbrida a média é 221 atributos, sendo 193 atributos selecionados com IG e 248 com KW.

	Seleção de Atributos			**Conjunto de dados**		
Extrator	**Abordagem**	**Avaliador**	**Classificador**	**JAFFE**	**CK**	**TFEID**
WLD	*Filter*	CFS	Não aplicável	58	145	135
	Híbrida	IG	SVM	80	300	120
			KNN	80	300	140
		KW	SVM	280	260	300
			KNN	300	240	200
LBP	*Filter*	CFS	Não aplicável	64	137	136
	Híbrida	IG	SVM	120	280	280
			KNN	120	220	280
		KW	SVM	280	220	260
			KNN	180	200	260

Tabela 6.2: Número de atributos obtidos com redução de dimensionalidade utilizando todas as expressões faciais

O número médio de atributos selecionados por cada técnica na redução de dimensionalidade com todas as classes (Tabela 6.3) foram comparados estatisticamente. Após

a constatação de não normalidade, foi aplicado o teste não paramétrico U de Mann-Whitney. As seguintes conclusões foram obtidas:

- O número de atributos selecionados com abordagem híbrida é significativamente maior que o *filter*;

- O CFS seleciona um número menor de atributos que o método híbrido com KW;

- Apesar de CFS selecionar em média menos atributos que IG, a diferença não é significativa, o mesmo é verificado entre IG e KW;

Tabela 6.3: Dados estatísticos quanto ao número de atributos selecionados na redução de atributos com todas as classes

	Técnica			**Abordagem**	
	CFS	*KW*	*IG*	*Filter*	*Híbrido*
Média	112,50	248,33	193,33	112,50	220,83
Desvio Padrão	40,09	40,41	90,79	40,09	74,24

Um dos fatores que induz a estratégia *filter* obter subconjuntos de dimensões menores que a estratégia híbrida corresponde ao tamanho da faixa dos subconjuntos avaliados. No modelo híbrido sempre é avaliada uma faixa que vai de 20 á 300 de atributos, enquanto que com o CFS são explorados no máximo 5 alternativas para encontrar um atributo mais promissor a cada nó e então o processo é concluído. Também deve-se considerar que as métricas de avaliação das estratégias são diferentes, no modelo híbrido é utilizado a taxa de acerto de um classificador e na estratégia *filter* é utilizado uma heurística.

Na Figura 6.8 são ilustrados os desempenhos dos classificadores c' para cada faixa de atributos avaliadas na redução de características com todas as expressões faciais e estratégia híbrida. De maneira geral, o KW consegue melhor desempenho à medida que mais atributos são utilizados, até que em determinado momento o desempenho permanece constante. No entanto para IG, como pode ser observado mais nitidamente no conjunto JAFFE, o desempenho aumenta até determinado valor à medida que mais atributos são selecionados, e então, o

desempenho começa a degradar. Este comportamento explica o fato de em média o IG selecionar um número menor de atributos que KW.

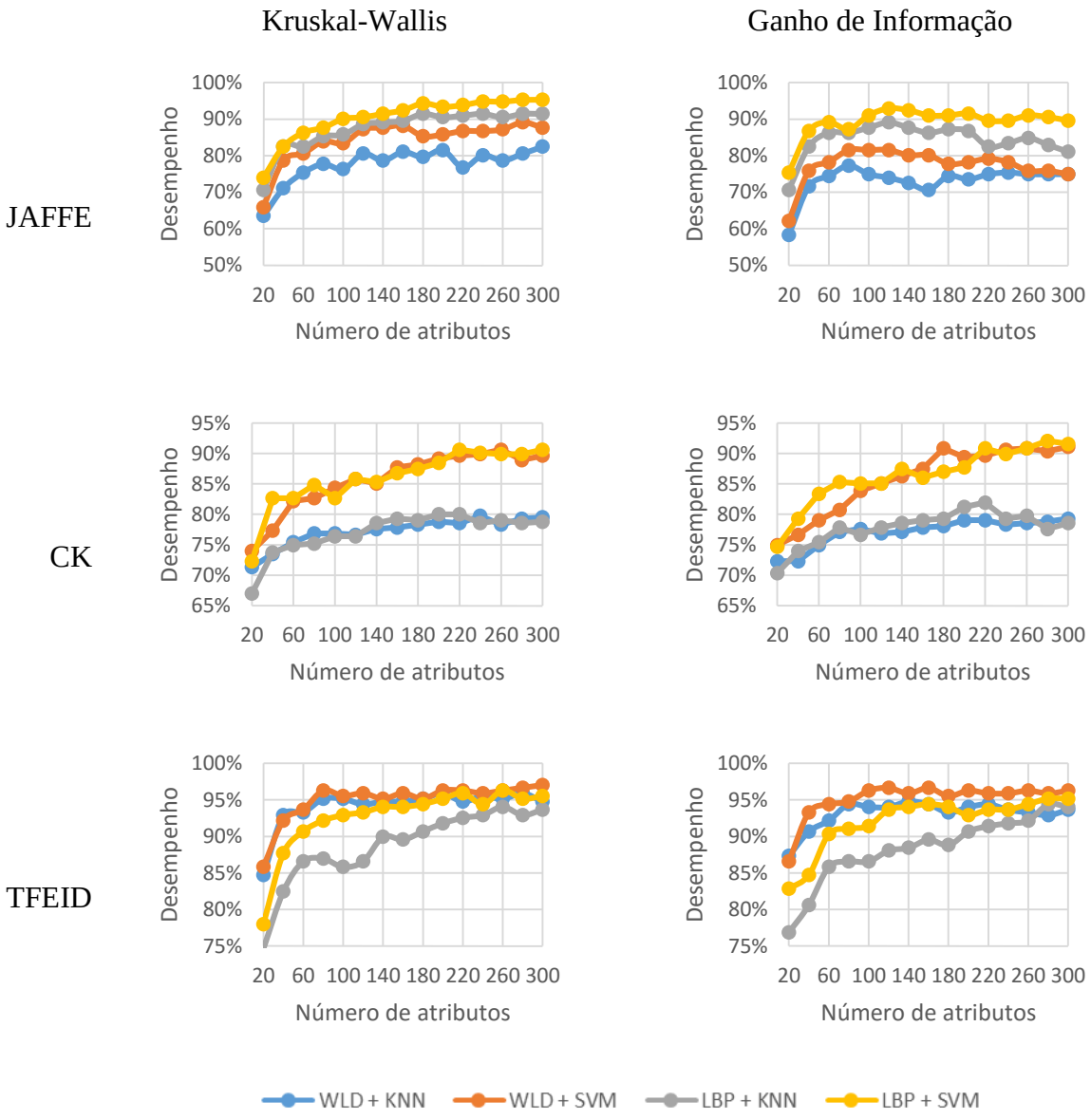


Figura 6.8: Desempenho de classificação em relação ao número de atributos utilizados durante a redução de características com todas expressões faciais

O número de atributos selecionados para cada classificador da estratégia Um-Contra-Um com redução de dimensionalidade em pares é apresentado Tabela 6.4. Como cada subconjunto F_p' pode conter um número de atributos diferentes, para melhor compreensão dos resultados são apresentados o número médio de atributos para cada conjunto de técnicas e corresponde a dimensão média dos dados utilizado como entrada para cada um dos 21

algoritmos da etapa de classificação (Figura 4.11(a)). Em média são utilizados 72 atributos por classificador com a redução de características em pares. A maior dimensão obtida é de 108 atributos e a menor é de 33 atributos, ambas para a abordagem híbrida. Com o CFS são manipulados em média 78 atributos por classificador, e com o IG e KW a dimensão média dos subconjuntos é de 73 e 69 respectivamente, assim, para a abordagem híbrida são processados em média 71 atributos por classificador.

Tabela 6.4: Número de atributos médios obtidos com seleção em pares

	Seleção de Atributos			**Conjunto de dados**		
Extrator	**Abordagem**	**Avaliador**	**Classificador**	**JAFFE**	**CK**	**TFEID**
WLD	*Filter*	CFS	Não aplicável	60	101	81
	Híbrida	IG	SVM	72	83	53
			KNN	103	95	41
		KW	SVM	74	77	78
			KNN	89	90	40
LBP	*Filter*	CFS	Não aplicável	59	91	78
	Híbrida	IG	SVM	43	108	34
			KNN	86	96	56
		KW	SVM	48	98	33
			KNN	69	92	34

Assim como na redução de dimensionalidade com todas as classes, também foi realizada a comparação estatística do número de atributos selecionados entre as diferentes abordagens e técnicas para a redução em pares (Tabela 6.5), e seguindo o mesmo protocolo de verificação não foi constatado nenhuma diferença significativa entre as abordagens *filter* e híbrida, e nem entre as técnicas IG, KW e CFS.

Tabela 6.5: Dados estatísticos quanto ao número de atributos selecionados na redução de dimensionalidade em pares de expressões

	Técnica			**Abordagem**	
	CFS	*KW*	*IG*	*Filter*	*Híbrido*
Média	78,33	69,00	72,50	78,33	70,75
Desvio Padrão	16,68	24,99	26,18	16,68	25,10

Considerando a quantidade de atributos utilizados por cada classificador binário (Figura 6.9), na seleção com todas as expressões em média são utilizados 199 atributos, enquanto que a redução em pares utiliza em média 72 atributos, ou seja, a seleção em pares consegue fornecer um espaço de atributos em cada classificador com uma dimensão aproximadamente 60% menor do que considerando todas as classes. De modo a verificar se a diferença é significativa, comparações estatísticas foram realizadas. Depois de avaliar a não normalidade das amostras, o teste não-paramétrico U de Mann-Whitney foi utilizado para comparar estatisticamente os valores médios do número de atributos utilizados por cada abordagem. Com o teste foi constatado que a dimensão obtida pela redução de atributos com todas as classes é significativamente maior que a redução com pares de expressão.

Tabela 6.6: Dados estatísticos quanto a dimensão dos dados obtidos pela redução com todas as classes e com a redução em pares

	Todas expressões	**Em pares**
Média	199,17	72,27
Desvio Padrão	81,19	23,60

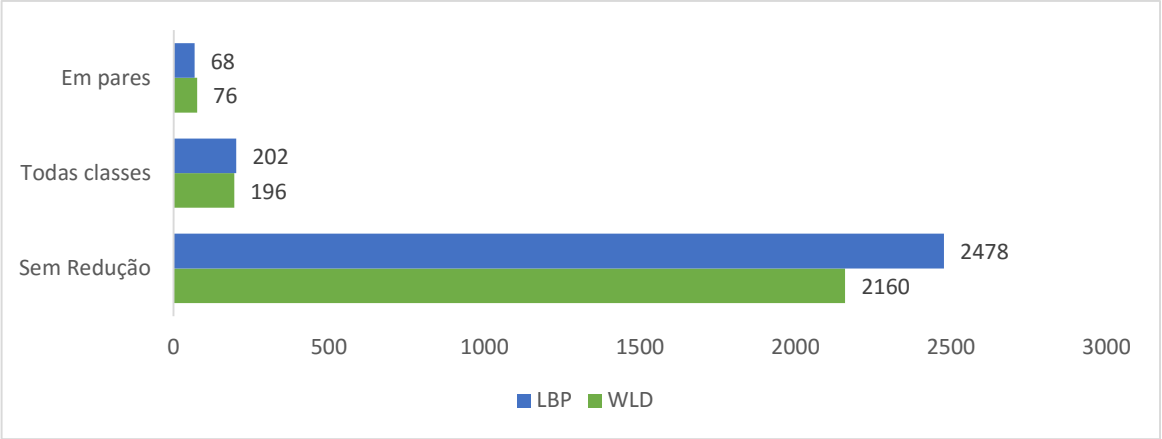


Figura 6.9: Número de atributos selecionados por cada estratégia de seleção

6.4. Classificação

Depois de reduzir a dimensionalidade do vetor de atributos a etapa final é a classificação das expressões faciais. Anteriormente foram utilizados classificadores para encontrar o melhor subconjunto de características devido ao uso do *wrapper*, nesta fase a classificação é realizada com o propósito de reconhecer uma expressão facial a partir de um conjunto de características.

Os resultados obtidos para o reconhecimento de expressões são apresentados na Tabela 6.7. Em média, o melhor desempenho obtido foi de 98,52% para extração de características com LBP, redução de características em pares com KW e classificação com SVM. Para o conjunto JAFFE, o melhor desempenho alcançado foi de 99,05%, sendo obtido pela extração de características com LBP e WLD, redução de dimensionalidade em pares com KW e classificação com SVM. A maior taxa de acerto para o conjunto CK foi obtida com LBP, IG em pares e SVM, em que 98,07% dos exemplos foram classificadas corretamente. Por fim, com o conjunto TFEID foi conseguido a maior taxa de acerto dos experimentos, 99,63%, e semelhante a JAFFE, o melhor resultado foi obtido com LBP, KW em pares e SVM. Desta forma, é possível verificar que os melhores resultados são obtidos com a redução de dimensionalidade em pares.

Tabela 6.7: Resultados de classificação

				Conjunto de dados			
Extrator	**Seleção de Atrib.**		**Classificador**	**JAFFE**	**CK**	**TFEID**	**Média**
WLD	Nenhum		SVM	90.05	89.40	94.78	91,41
			KNN	85.31	72.29	83.96	80,52
	Todas classes	CFS	SVM	81.99	91.08	97.39	90,15
			KNN	75.83	81.93	96.64	84,80
		IG	SVM	81.52	91.08	96.64	89,75
			KNN	77.25	79.28	94.78	83,77
		KW	SVM	89.10	90.60	97.01	92,24
			KNN	82.46	79.76	95.90	86,04
	Em pares	CFS	SVM	98.10	96.14	98.88	97,71
			KNN	93.84	89.16	97.76	93,59
		IG	SVM	97.63	96.14	98.51	97,43
			KNN	98.58	90.36	97.76	95,57
		KW	SVM	**99.05**	96.87	98.88	98,27
			KNN	97.63	91.33	98.51	95,82
LBP	Nenhum		SVM	91.94	91.81	93.28	92,34
			KNN	92.42	73.49	82.84	82,92
	Todas classes	CFS	SVM	94.31	92.05	95.90	94,09
			KNN	87.68	81.45	92.91	87,35
		IG	SVM	92.89	92.05	95.15	93,36
			KNN	89.10	81.93	94.40	88,48
		KW	SVM	95.26	90.60	96.27	94,04
			KNN	91.47	80.00	94.03	88,50
	Em pares	CFS	SVM	98.10	97.83	99.25	98,39
			KNN	95.26	93.25	98.51	95,67
		IG	SVM	98.10	**98.07**	99.25	98,47
			KNN	97.63	94.76	98.88	97,09
		KW	SVM	**99.05**	96.87	**99.63**	**98,52**
			KNN	98.58	93.01	98.51	96,70

Como ilustrado na Figura 6.10, a taxa de acerto média para classificação utilizado seleção de atributos com todas as classes tem se mostrado melhor do que utilizar todo o espaço de características. Os resultados mostram que em média a redução de dimensionalidade em pares é capaz de produzir taxas de acertos maiores que a abordagem sem seleção de atributos e seleção de atributos com todas expressões. No entanto para verificar se a diferença é significativa e também validar a hipótese "*Na classificação com estratégia Um-Contra-Um, a seleção de atributos em pares de expressões faciais consegue selecionar atributos mais discriminantes levando a uma maior taxa de acerto do que a abordagem considerando todas as expressões?"*, os resultados das classificações foram comparados estatisticamente (Tabela 6.8). Após verificar a normalidade dos dados com o teste de Shapiro-Wilk, foi aplicado o teste não paramétrico U de Mann-Whitney e as seguintes conclusões foram obtidas:

- Apesar de em média a taxa de acerto sem a redução de atributos ser inferior à abordagem de redução de características com todas as classes, a diferença não é significativa;

- Os desempenhos obtidos sem redução de atributos e com redução de atributos utilizando todas as classes são significativamente inferiores ao desempenho com a redução de atributos em pares. Isso prova de que a hipótese levantada deve ser aceita;

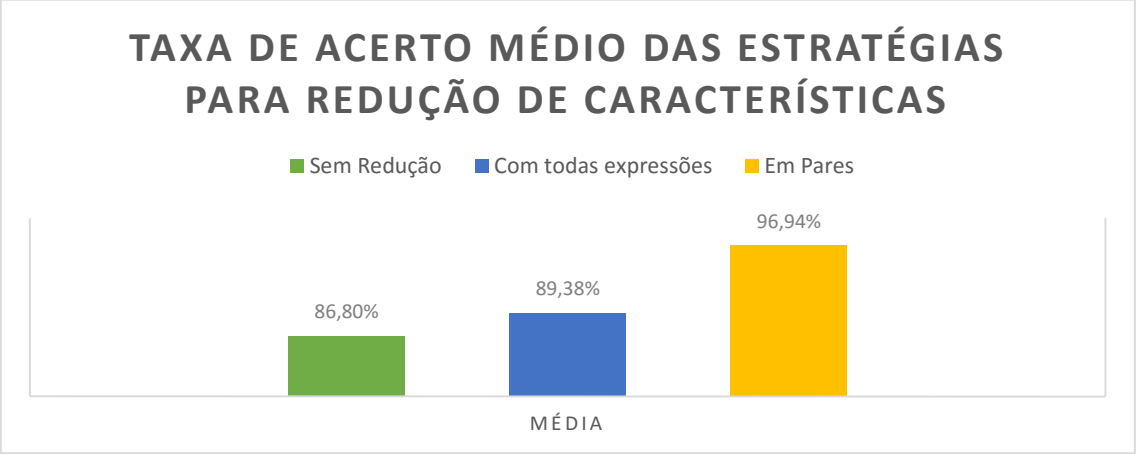


Figura 6.10: Comparativo das abordagens de redução de dimensionalidade

Tabela 6.8: Dados estatísticos do percentual de expressões classificadas corretamente para os três tipos de abordagens de redução de dimensionalidade

	Sem redução	**Todas classes**	**Em Pares**
Média	86,80	89,38	96,94
Desvio padrão	7,51	6,62	2,65

Os resultados de classificação mostram que a redução da dimensionalidade com pares além de produzir uma maior taxa de acerto que as demais abordagens também possui menos atributos (seção 6.3). Os resultados com cada técnica de redução em pares são ilustrados na Figura 6.11. Em média a abordagem híbrida com KW atinge maiores taxas de acertos que o CFS (Tabela 6.9), no entanto após comparar as médias com o teste paramétrico t-Student, não foram encontradas diferenças estatísticas nos resultados. Assim, o CFS pode ser utilizado como uma alternativa ao KW com a vantagem de necessitar menos processamento e também não haver preocupação com o *overfitting* resultante do *wrapper*. Também como descrito na seção 6.3, é constado que na redução de atributos em pares as diferentes técnicas utilizadas não apresentam tamanho significativo na dimensão dos subconjuntos selecionados, ou seja, o número de atributos processados pela estratégia com CFS não é significativamente maior de KW.

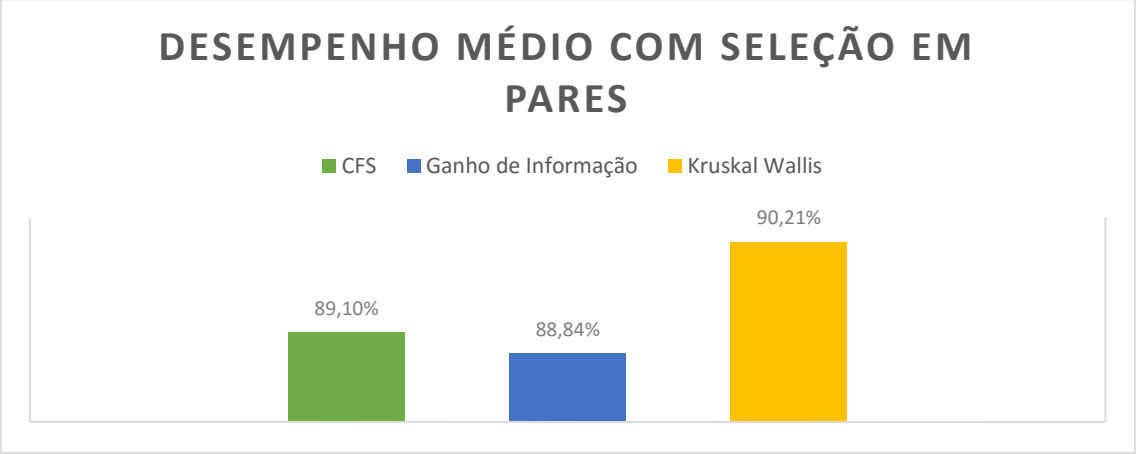


Figura 6.11: Gráfico do desempenho médio para seleção de atributos em pares

Tabela 6.9: Dados estatísticos da taxa de acerto para as técnicas utilizadas na redução em pares de expressões

	CFS	**IG**	**KW**
Média	89,10	88,84	90,21
Desvio padrão	7,16	6,91	6,27

Com relação ao desempenho dos classificadores, o gráfico da Figura 6.12 apresenta a taxa de reconhecimento facial médio da SVM e do KNN para os conjuntos JAFFE, CK e TFEID. Os resultados mostram que a SVM em média consegue identificar corretamente 94,73% das expressões faciais e o KNN 89,77%, uma diferença percentual de aproximadamente 5%. A maior diferença de desempenho dos classificadores pode ser verificada no conjunto CK, em que a SVM alcança 93,61% e o KNN 84,43%. Para verificar se as diferenças são significativas foram realizadas analises estatísticas (Tabela 6.10). Após verificar a não normalidade dos dados, o teste não paramétrico de Mann-Whiney foi aplicado e contatou-se que a taxa de acerto do KNN é significativamente inferior à SVM

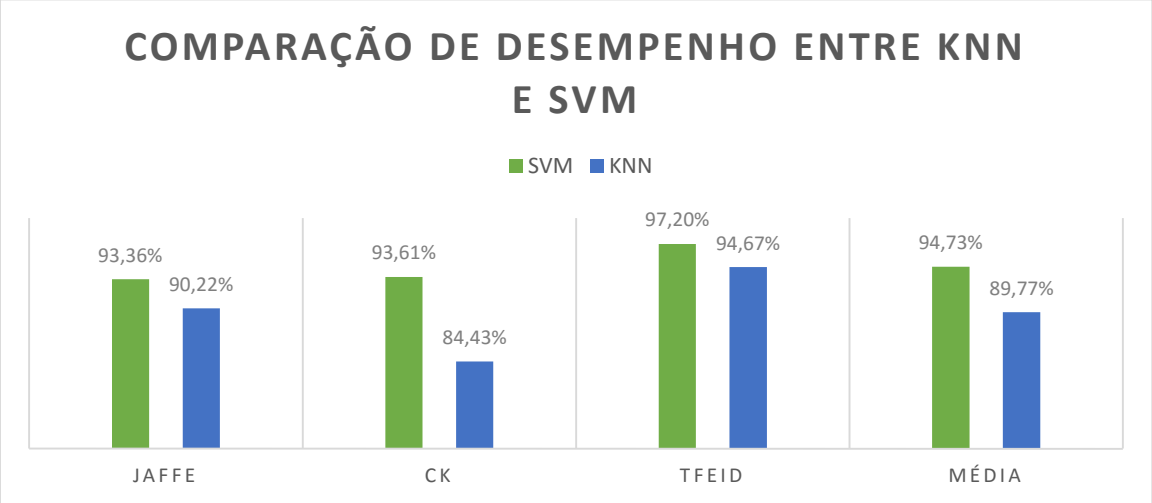


Figura 6.12: Comparação de desempenho entre KNN e SVM

Tabela 6.10: Média e desvio padrão para classificação com SVM e KNN de todos experimentos realizados

	SVM	**KNN**
Média	94,73	89,77
Desvio Padrão	4,31	7,91

A partir das matrizes de confusão pode ser verificado a taxa de acerto para cada expressão. Foram comparadas as matrizes de confusão para os dois extratores de características utilizados, duas técnicas de seleção de atributos baseados em pares que produzem maior desempenho (Figura 6.11), o CFS e o KW, e classificação com SVM. É possível verificar que as taxas de acertos para os diferentes métodos e conjuntos de dados produzem resultados semelhantes.

As Tabela 6.11 e Tabela 6.12 mostram que a expressão mais fácil de ser reconhecida pelo método implementado é alegria, esta emoção em quase todas as abordagens apresentou menos erro permanecendo quase sempre com 100% de acerto. Outras expressões, como desgosto e medo, apresentaram taxa de acerto de aproximadamente 99%. A tristeza foi a expressão com maior dificuldade de ser classificada e em ambas as tabelas esse fato fica evidente. Em Zavaschi et al. [17], para o conjunto JAFFE a expressão com menor taxa de acerto também foi tristeza, enquanto as mais fáceis foram neutro e medo. Em Bashar et al. [15] para as 7 expressões, neutro obteve menor taxa de acerto, enquanto raiva e medo conseguiram os melhores resultados. No método proposto por Hussain et al. [2] utilizando JAFFE, a expressão mais difícil de ser reconhecida foi medo e a mais fácil foi surpresa. Os trabalhos de Liu et al. [53] e Zhang et al. [89] não apresentam a matriz de confusão. Desta forma é possível concluir que a confusão que ocorre na classificação não é devido a composição das imagens ou semelhança entre as expressões, mas sim de acordo com o conjunto de técnicas utilizadas.

Tabela 6.11: Desempenho individual de cada expressão com CFS

	LBP + CFS + SVM				WLD + CFS + SVM			
	JAFFE	***CK***	***TFEID***	***Média***	***JAFFE***	***CK***	***TFEID***	***Média***
Raiva	100	91	100	97	100	98	97	98
Nojo	96	98	98	97	100	100	98	99
Medo	97	100	100	99	97	100	100	99
Alegria	100	100	100	100	100	100	100	100
Tristeza	97	89	97	94	97	86	100	94
Surpresa	100	99	100	100	97	99	100	99
Neutro	97	94	97	96	97	97	100	98

Tabela 6.12: Desempenho individual de cada expressão com KW

	LBP + KW EM PAR + SVM				WLD + KW EM PAR + SVM			
	JAFFE	***CK***	***TFEID***	***Média***	***JAFFE***	***CK***	***TFEID***	***Média***
Raiva	100	93	97	97	100	95	100	98
Nojo	100	98	98	99	100	100	98	99
Medo	97	100	100	99	100	100	100	100
Alegria	100	99	100	100	100	99	100	100
Tristeza	100	89	97	95	97	86	100	94
Surpresa	97	98	100	98	97	98	100	98
Neutro	100	97	100	99	100	97	100	99

O gráfico da Figura 6.13 apresenta um comparativo do desempenho obtido com o método aqui desenvolvido e outros trabalhos da literatura. Em relação à Liu et al. [53] que utiliza características baseadas em textura, o método deste estudo se mostrou 3% superior para o conjunto JAFFE. No entanto, pode existir uma diferença gerada pelo protocolo de validação, em Liu et al. [53] o conjunto é separado em 137 imagens para treinamento e 76 para testes repetindo o processo 3 vezes, enquanto que o presente trabalho utiliza validação cruzada com 10 partições. Também no estudo de Liu et al. [53] são utilizadas características baseadas em textura obtidas com WLD e demonstra que o método não necessita utilizar toda a face para fazer o reconhecimento da expressão, ou seja, é possível classificar uma face com apenas metade dela, ou sem a região dos olhos ou boca.

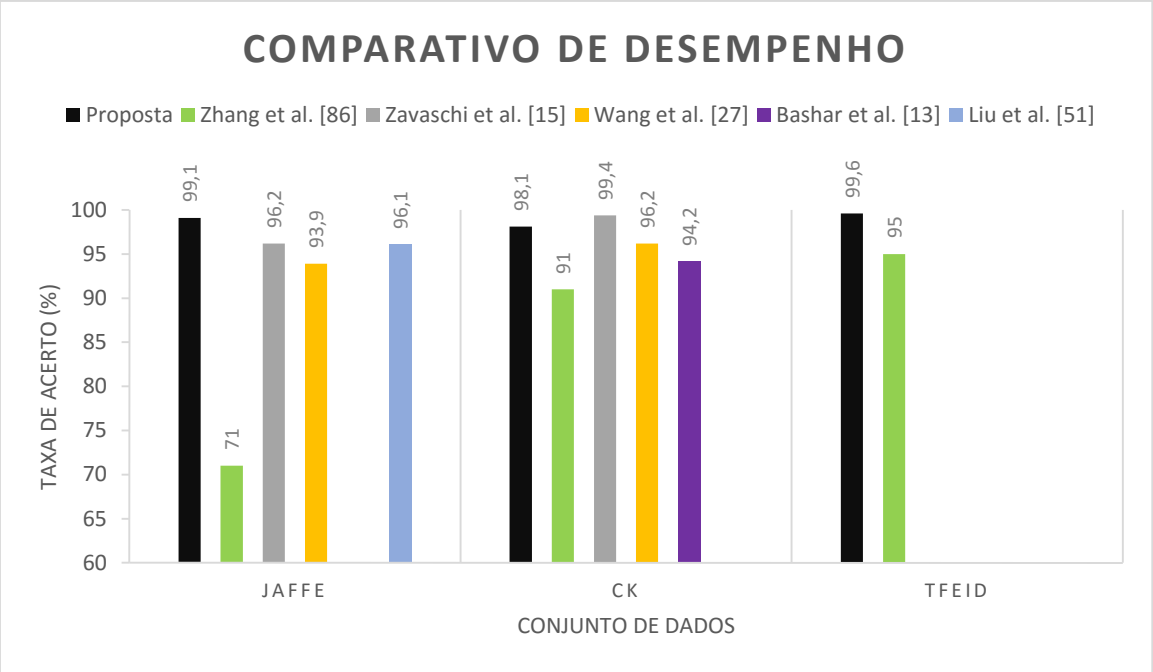


Figura 6.13: Comparação de desempenho.

Em Bashar et al. [15] é utilizado o extrator de características MTP baseado em textura e uma SVM para classificar as expressões faciais. A validação do método é feita através da validação cruzada com 10 partições sobre o conjunto CK. Com o método de Bashar et al. [15] foi obtido 94.2% de acerto para 7 expressões faciais permanecendo abaixo desta abordagem (98,1%). Ainda, há um elevado custo computacional em que a extração de características com MTP produz um vetor de 21504 dimensões e neste estudo para CK são utilizados 108 atributos. Também no experimento de Bashar et al [15] são utilizadas 1632 imagens sendo que a grande maioria teve que ser rotulada pelo autor e não está claro se o processo de recorte da face é realizado automaticamente.

No trabalho de Wang et al. [16] utilizando de fusão de características do algoritmos WLD e HOG foi obtido uma taxa de acerto de 93,9% para a JAFFE e 96,2% para a CK. O método foi avaliado em 1344 imagens do conjunto de dados CK, em que são selecionados os 6 últimos quadros de cada sequência das 7 expressões faciais de 32 indivíduos. Para validação os conjuntos JAFFE e CK foram utilizados e separados em 50% para treinamento e 50% para testes. Este protocolo de validação se difere desta proposta e das demais da literatura e impede uma comparação mais justa, contudo, a abordagem deste trabalho conseguiu taxas de acerto superiores tanto para JAFFE, quanto para CK. Ainda em Wang et al. [16] uma face deve ser

representada com características de dois extratores de características, enquanto que neste trabalho são consideradas as características extraídas por WLD ou LBP.

Na abordagem de Zhang et al. [89] foi utilizado Gabor Filter para extração de características, PCA e LDA para redução de características e SVM para classificação. A validação do método foi realizada dividindo cada um dos conjuntos de dados JAFFE, CK e TFEID em duas partições, sendo uma parte composta por metade dos sujeitos e utilizada para o treinamento do classificador, e a segunda metade para validação, o processo é repetido 3 vezes. A proposta aqui apresentada conseguiu uma taxa de acerto de 98,1% e 99,6% para CK e TFEID respectivamente, enquanto em Zhang et al. [89] foi obtido 91% com1693 imagens da CK, e 95% para TFEID com 268 imagens. A diferença de desempenho é mais notável para JAFFE, uma vez que o método aqui desenvolvido alcançou 99,1% com 211 faces e em Zhang et al. [89] a taxa de acerto foi de 71% para 213 faces. Contudo deve-se considerar que a dificuldade em classificar uma expressão no protocolo de Zhang et al. [89] é maior, pois os sujeitos de treinamento e teste são diferentes. Ainda o trabalho de Zhang et al. [89] apresenta uma etapa de pré-processamento mais robusta que este estudo, em que as faces são rotacionadas e escaladas de maneira a manter os olhos alinhados horizontalmente à uma distância normalizada. Um modelo facial é aplicado para remover o fundo (Figura 6.14) e para melhorar o contraste da imagem é *aplicado Contrast Limited Adaptive Histogram Equalization* (CLAHE) que ameniza os efeitos de ruídos.

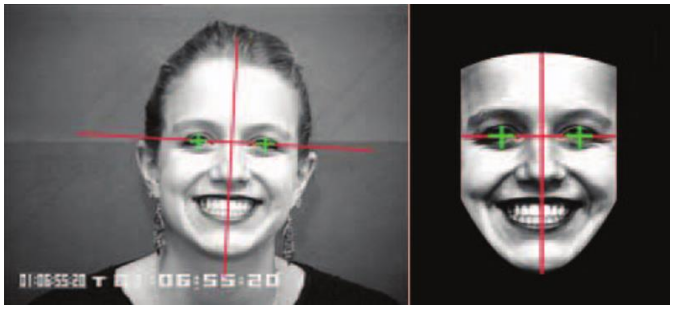


Figura 6.14: Pré-processamento de [89]. A imagem da direita representa o resultado gerado.

Por fim o trabalho de Zavaschi et al. [17] que possui maior taxa de acerto e utiliza pontos fiduciais. Deve-se considerar que nesta abordagem há uma restrição, os pontos faciais devem ser determinados manualmente o que descarta o erro gerado por um algoritmo e demanda trabalho humano para identificar e rotular estes pontos nas imagens. Zavaschi et al. [17] utilizou validação cruzada com 10 partições para verificar o desempenho da abordagem. A taxa de

acerto para JAFFE foi de 96,2%, enquanto que o método com redução em pares atingiu 99,1%, uma diferença de aproximadamente 3%. Para o conjunto CK os resultados são diferentes, enquanto o método apresentado conseguiu 98,1%, o trabalho de Zavaschi et al. [17] alcançou 99,4%. Essa comparação é menos coerente devido a diferença do número de imagens utilizadas em cada trabalho. Neste estudo foram utilizadas 414 imagens, sendo compostas pela última imagem de cada sequência e com rótulo fornecido pelo conjunto CK. Enquanto que em Zavaschi et al. [17] não está claro como foram obtidas as 1281 imagens utilizadas no experimento. No conjunto CK são fornecidos os rótulos para apenas 327 sequências, assim para utilizar mais exemplos o autor deve rotular sequências desconhecidas conforme sua opinião. Outra maneira é em vez de utilizar somente a última expressão de cada sequência, incluir por exemplo, as 6 últimas imagens da sequência semelhante ao trabalho de Wang et al. [16]. Desta forma as expressões dos sujeitos passam a ser frequentes tanto no subconjunto de treino quanto no teste, assim o protocolo de validação fica semelhante em testar o modelo com o conjunto de treinamento. O método proposto para reconhecer expressões faciais necessita apenas de um extrator de características independente do conjunto de dados, enquanto que em Zavashi et al. [17] pode ser verificado que para o conjunto JAFFE os melhores resultados são obtidos por meio de 5 extratores de características, sendo um LBP Uniforme e os demais Gabor Filters, e também 5 classificadores SVM. Para o conjunto CK o cenário é melhor, são utilizados um LBP Uniforme e um Gabor Filter para produzir as características que são classificadas por duas SVM's.

De modo a avaliar se as diferenças são significativas entre os trabalhos da literatura e o método proposto, foram realizadas análises estatísticas dos resultados para cada conjunto de dados. Na Tabela 6.13 são apresentados as médias e os desvios padrões para os resultados obtidos sobre o conjunto JAFFE. Os trabalhos de Zhang et al. [89] e Liu et al. [53] não foram incluídos na comparação por não fornecer a matriz de confusão. Com o teste de Shapiro-Wilk foi verificado que os dados da proposta não seguem uma distribuição normal, por isso foi utilizado o teste não paramétrico U de Mann-Whitney. Com relação ao trabalho de Zavaschi et al. [17] a proposta não apresentou diferenças significativas, no entanto, foi verificado que o desempenho do método proposto é significativamente superior do que Wang et al. [16].

Tabela 6.13: Desempenho por expressão para o conjunto JAFFE

Expressões	**Proposta**[4]	**Zavaschi et al** [17]	**Wang et al** [16]
Raiva	100,0	96,7	93,3
Medo	100,0	100,0	88,9
Alegria	100,0	90,3	100,0
Tristeza	96,8	93,5	99,3
Nojo	100,0	96,6	88,9
Neutro	100,0	100,0	93,3
Surpresa	97,0	100,0	100,0
Média	99,1	96,7	94,8
Desvio Padrão	1,5	3,7	5,0

Na Tabela 6.14 são apresentados os desempenhos de reconhecimento para cada tipo de expressão obtido no conjunto CK. Os trabalhos de Zhang et al. [89] e Zavaschi et al. [17] não apresentam os desempenho para cada expressão, por isso foram excluídos da comparação. O teste de Shapiro-Wilk demonstrou que os dados da proposta não pertencem a uma distribuição normal, portanto foi utilizado o teste não paramétrico U de Mann-Whitney para comparar os métodos. Segundo os testes realizados, os resultados obtidos com a proposta foram significativamente maior que Bashar et al. [15], mas não foi encontrada diferenças com o trabalho de Wang et al. [16].

[4] Foi considerado o conjunto de técnicas que produz maior taxa de acerto para o conjunto JAFFE, ou seja, extração de características com LBP, seleção de atributos em pares com KW e classificação com SVM.

Tabela 6.14: Desempenho por expressão para o conjunto CK

Expressões	**Proposta**[5]	**Wang et al.** [16]	**Bashar et al.** [15]
Raiva	95,5	95,5	94,5
Medo	100,0	92,7	95,5
Alegria	100,0	98,3	93,2
Tristeza	89,3	94,0	93,3
Nojo	100,0	96,9	93,2
Neutro	98,3	95,9	91,8
Surpresa	98,8	97,7	92,6
Média	*97,4*	*95,9*	*93,4*
Desvio Padrão	*3,9*	*2,0*	*1,2*

O trabalho de Zhang et al. [89] é o único trabalho da literatura avaliado com o conjunto TFEID e não apresenta os desempenhos para cada tipo expressão, devido a este fato, não foi possível realizar comparações com o respectivo conjunto de dados.

Com base nas comparações realizadas com outras propostas para reconhecer expressões faciais deve-se rejeitar a hipótese de que o "o *reconhecimento de expressões faciais através da classificação com estratégia Um-Contra-Um consegue obter desempenho superior em relação aos trabalhos da literatura"*, pois os testes estatísticos não indicaram melhoras significativas em relação a determinados trabalhos.

De modo a avaliar a eficiência do método desenvolvido em relação aos demais trabalho, uma comparação é apresentada na Tabela 6.15, em que o número de atributos processados por cada método é utilizado como métrica de eficiência. Esta métrica possui princípio semelhante ao trabalho de Last et al. [43], o qual também avalia o custo computacional através do número de atributos utilizado por cada classificador. Em Bashar et al. [15] são utilizados 21504 atributos com SVM Multiclasse Um-Contra-Todos, o que na prática o vetor de características é aplicado em 7 SVM Binárias, ou seja, ao final são processados $21504 \times 7 = 150528$ atributos, o mesmo acontece para Zavashi et al. [17] e Liu et al [53]. O fato só não se repete para o KNN, mas em contrapartida deve ser calculado a distância para cada exemplo de treinamento e conforme os resultados de Liu et al [53], a medida que a dimensão aumenta, as diferenças de

[5] Foi considerado o conjunto de técnicas que produz maior taxa de acerto para o conjunto CK, ou seja, extração de características com LBP, seleção de atributos em pares com IG e classificação com SVM.

tempo entre a SVM e o KNN passam a ser mais significativas, sendo a SVM mais rápida. Com base nos fatos apresentados a Figura 6.15 mostra o número total de atributos processados por cada abordagem. Apesar de serem necessário 21 classificadores, a redução de características em pares é capaz de selecionar um pequeno número de atributos para cada SVM em relação as demais abordagens, ainda como descrito anteriormente, é capaz obter um desempenho equivalente ou superior aos melhores resultados da literatura, portanto, a hipótese de que "*a redução de atributos e classificação Um-Contra-Um possui custo computacional maior do que os trabalhos da literatura*" deve ser rejeitada.

Tabela 6.15: Comparação de dimensionalidade com outras propostas

Trabalho	**Conjunto de dados**	**Dimensão (extrator)**	**Classificação**
Bashar et al. [15]	CK	21504 (MTP)	SVM Multiclasse Um-Contra-Todos
Zavaschi et al. [17]	JAFFE	2478 (LBP) + 160 (Gabor Scale) + 3x 100 (Gabor Orientation)	O vetor de atributos gerado por cada extrator de características é classificado por uma SVM Multiclasse Um-Contra-Um.
Zavaschi et al. [17]	CK	2478 (LBP) + 160 (Gabor Scale)	
Wang et al. [16]	JAFFE	1440 (HOG) + 2560 (WLD)	As características são classificadas com KNN (JAFFE)
Liu et al .[53]	JAFFE	6x 144 (WLD)	6x SVM Multiclasse Um-Contra-Um (JAFFE)
Zhang et al. [89]	JAFFE e CK	3x 6 (PCA+LDA)	3x SVM Multiclasse Um-Contra-Um (CK e CK)
Proposta	JAFFE	21x 48[6] (LBP)	21x SVM Binária
Proposta	CK	21x 108[7] (LBP)	21x SVM Binária

[6] Utilizou-se a dimensão média do número de atributos selecionados com a redução de dimensionalidade em pares com KW

[7] Utilizou-se a dimensão média do número de atributos selecionados com a redução de dimensionalidade em pares com IG

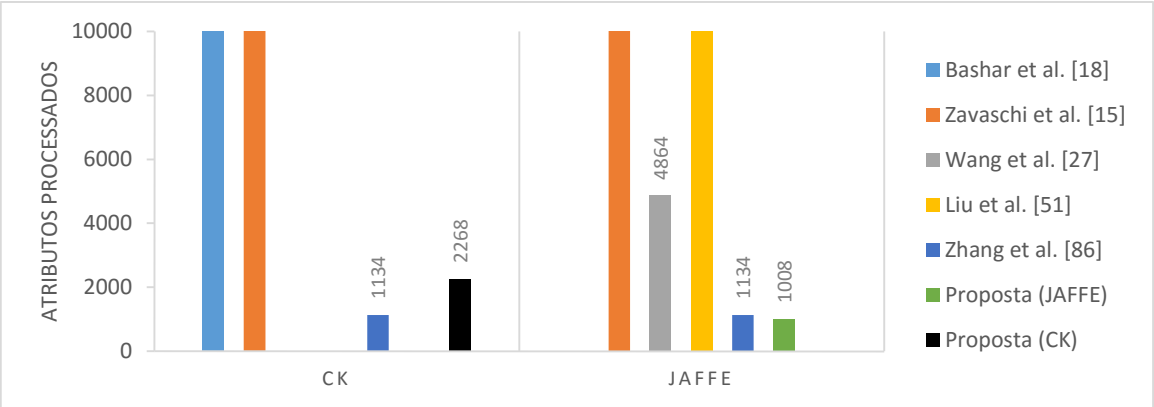


Figura 6.15: Número de atributos processados

Capítulo 7

Conclusão

Neste documento foi apresentado um método de reconhecimento automático de expressões faciais. Para isso foram utilizadas as etapas de detecção facial, extração de características, redução de dimensionalidade e classificação. Para detecção facial foi utilizado o algoritmo proposto por Viola-Jones, e a extração de características foi executada pelo LBP e WLD. Duas estratégias de redução de dimensionalidade foram seguidas, uma baseada em *filter* com CFS e outra composta por *filter* e *wrapper* com IG e KW. A classificação foi avaliada com SVM e KNN, e para validação foram utilizados os conjuntos de dados JAFFE, CK e TFEID.

Conforme apresentado na literatura a extração de características baseada em geometria ainda não é robusta para detectar os pontos fiduciais com precisão. As variações de luminosidade e baixas resoluções influenciam as técnicas para detecção dos pontos fiduciais, afetando diretamente no desempenho dos classificadores, por este motivo foi explorado extratores de características baseados em textura. Ainda alguns métodos baseados em características geométricas necessitam que os pontos fiduciais sejam indicados manualmente na etapa de treinamento, demandando esforço humano, e também as técnicas utilizadas para obter as características geométricas, como a AAM, possuem parâmetros iniciais difíceis de serem determinados.

Com o problema da alta dimensionalidade gerada pelos extratores de características foi apresentado um modelo para reduzir o número de atributos. O presente trabalho propõe em aplicar a seleção de atributos por pares de expressões e classificação Um-Contra-Um. Esta estratégia foi comparada com a seleção de atributos utilizando todas as classes, sem a redução de atributos e com outros trabalhos da literatura.

A partir de experimentos realizados com modelo de seleção de atributos avaliado foi possível identificar os seguintes fatos:

- Quanto a hipótese de que "*a seleção de atributos em pares de expressões faciais consegue selecionar atributos mais discriminantes levando a uma maior taxa de acerto do a abordagem considerando todas as classes*", foi possível identificar uma melhora nos resultados em relação ao seu uso com classificação Um-Contra-Um. Em média, a seleção de atributos em pares conseguiu reconhecer corretamente 96,94% das expressões faciais, a redução de características com todo o conjunto atingiu 89,38% e sem redução de dimensionalidade foi obtido 86,80%. Testes estatísticos demonstram que a diferença é significativa e que a redução de dimensionalidade em pares consegue maiores taxas de acertos;

- Testes estatísticos demonstraram que para o uso de classificação com estratégia Um-Contra-Um a redução de atributos em pares é capaz de fornecer uma dimensionalidade significativamente menor em cada classificador do que a redução com todas as expressões;

- As confusões geradas na classificação não estão relacionadas com a proximidades entre as expressões, mas sim com as técnicas utilizadas. As confusões do método proposto e outros trabalhos da literatura produzem resultados diferentes, no entanto com as confusões geradas pelo método desenvolvido verificou-se semelhanças entre as expressões com maior taxa de acerto e as mais difíceis de serem identificadas, mesmo com diferentes conjuntos de dados;

- Os resultados mostram que apesar da proposta utilizar um conjunto de 21 classificadores o método não tem custo computacional mais elevado. Uma comparação mostrou que o método proposto é capaz de processar menos atributos em cada classificador do que outros trabalhos da literatura, demonstrando que a seleção de atributos em pares pode ser uma eficiente maneira de obter características mais discriminantes com menor dimensionalidade. No trabalho desenvolvido, a maior taxa de acerto para JAFFE foi obtida processando 1008 atributos, enquanto que grande parte dos trabalhos da literatura podem processar mais de 4000 atributos. O resultado se repete para CK, em que nesta proposta são processados 2268 atributos e na literatura o número ultrapassa os 10000. Testes estatísticos também demonstraram que em

relação aos trabalhos da literatura, a seleção em pares e classificação Um-Contra-Um produz resultados equivalentes ou superiores.

- Apesar do método de seleção de atributos em pares conseguir atributos discriminantes com menor dimensionalidade, deve-se avaliar que o processo de seleção possui um alto custo computacional, uma vez que deve ser aplicada a cada par de expressão, e utilizando o modelo híbrido eleva-se ainda mais o tempo de processamento. No entanto este processo é executado somente uma única vez e posteriormente a classificação é realizada rapidamente. Também deve-se avaliar a possibilidade de *overftting*, já que na estratégia *wrapper*, os mesmos dados para seleção de atributos são os mesmos usados para a validação. Através do cruzamento do conjunto de dados é possível verificar se o impacto do *overffiting* é significativo, nesta estratégia os classificadores são treinados com um conjunto e a validação é realizada em outro. Para isso é necessário que os conjuntos de dados possuam características semelhantes, ao contrário da JAFFE e CK que apresentam características muito distintas. Outra alternativa é utilizar o modelo *filter* para seleção de atributos, que conforme testes estatísticos o modelo *filter* não é significativamente inferior ao modelo híbrido.

Uma condição que deve ser verificada no método proposto corresponde a quantidade de expressões faciais utilizadas. Quando consideradas 7 expressões faciais são necessários 21 pares, mas ao utilizar 10 expressões são verificados 45 pares e com 12 expressões o número de pares possíveis é de 66. Assim é verificado que o número de pares gerados aumenta exponencialmente a medida que mais expressões são avaliadas e em determinadas condições o custo computacional pode ser caro.

Tanto a proposta deste estudo, quanto em grande parte dos trabalhos da literatura, ao avaliar uma expressão desconhecida o método é obrigado a classificar como alguma das expressões conhecidas e levando a uma interpretação errônea. Uma estratégia que possibilite distinguir uma expressão desconhecida seria mais adequada em situações reais.

O método proposto por este estudo para o reconhecimento de expressões faciais tem conseguido se aproximar dos melhores resultados obtidos da literatura com menor dimensionalidade e consequentemente menor custo computacional, demonstrando que o trabalho desenvolvido é de grande importância para a comunidade científica.

7.1. Trabalhos Futuros

Em trabalhos futuros, a seleção de atributos em pares deve ser avaliada utilizando a estratégia Um-Contra-Todos. Considerando as 7 expressões faciais, seriam necessários apenas 7 subconjuntos ao invés dos 21. Essa estratégia poderia ser uma alternativa para conseguir reduzir ainda mais o tempo de treinamento e o custo computacional.

Futuramente a avaliação da redução em pares deve ser executada em conjuntos de dados maiores como o *Facial Expression and Emotion Database* (FEED) [96] e o *MMI Facial Expression Database* [97]. Um segundo protocolo experimental utilizando um conjunto formado por JAFFE, CK, TFEID, FEED e MMI poderia ser conduzido para avaliar o método em cenários reais, pois desta forma haveria um número maior de exemplos para compor um subconjunto para treinamento e outro para validação, além de ser possível avaliar a capacidade de generalização, tal como o *overfitting*.

Por fim, para reduzir o tempo de extração de características de uma face, seria importante avaliar os resultados extraindo as características de sub-regiões estratégicas da face, como a área dos olhos, testa, bochechas e boca. Isso evitaria ter que percorrer toda a imagem e eliminaria regiões irrelevantes como o nariz.

Referências

[1] A. Dhall, “Context Based Facial Expression Analysis in the Wild,” in *Proceedings of the 3rd ACM conference on International conference on multimedia retrieval - ICMR ’13*, 2013, pp. 636–641.

[2] M. Hussain, S. A. Khan, N. Ullah, N. Riaz, and M. Nazir, “Computationally Efficient Invariant Facial Expression Recognition,” *Res. J. Recent Sci.*, vol. 3, no. 2, pp. 61–68, 2014.

[3] S. Zhang, L. Li, and Z. Zhao, “Facial expression recognition based on Gabor wavelets and sparse representation,” *2012 IEEE 11th Int. Conf. Signal Process.*, vol. 2, pp. 816–819, Oct. 2012.

[4] Z. Zhang, M. Lyons, M. Schuster, and S. Akamatsu, “Comparison between geometry-based and Gabor-wavelets-based facial expression recognition using multi-layer perceptron,” in *Third IEEE International Conference on Automatic Face And Gesture Recognition*, 1998.

[5] A. Mehrabian, “Communication Without Words,” *Psychol. Today*, vol. 2, no. 4, pp. 53–56, 1968.

[6] S. V. P. G. do Rosário, “Reprodução de Informação Associada a Expressões Faciais por Via do seu Reconhecimento,” Universidade Técnica de Lisboa, 2008.

[7] S. B. Hamann and R. Adolphs, “Normal recognition of emotional similarity between facial expressions following bilateral amygdala damage.,” *Neuropsychologia*, vol. 37, no. 10, pp. 1135–41, Sep. 1999.

[8] R. Sprengelmeyer, A. W. Young, K. Mahn, U. Schroeder, D. Woitalla, T. Büttner, W. Kuhn, and H. Przuntek, “Facial expression recognition in people with medicated and unmedicated Parkinson’s disease,” *Neuropsychologia*, vol. 41, no. 8, pp. 1047–1057, Jan. 2003.

[9] H. D. Critchley, P. Rotshtein, Y. Nagai, J. O’Doherty, C. J. Mathias, and R. J. Dolan, “Activity in the human brain predicting differential heart rate responses to emotional facial expressions,” *Neuroimage*, vol. 24, no. 3, pp. 751–62, Feb. 2005.

[10] K. M. Corcoran, S. R. Woody, and D. F. Tolin, “Recognition of facial expressions in obsessive-compulsive disorder,” *J. Anxiety Disord.*, vol. 22, no. 1, pp. 56–66, Jan. 2008.

[11] J. E. Martinez, D. C. Grassi, and L. G. Marques, “Análise da aplicabilidade de três instrumentos de avaliação de dor em distintas unidades de atendimento: ambulatório, enfermaria e urgência,” *Rev. Bras. Reumatol.*, vol. 51, no. 4, pp. 304–308, 2011.

[12] N. Perveen, S. Gupta, and K. Verma, “Facial expression recognition using facial characteristic points and Gini index,” *2012 Students Conf. Eng. Syst.*, pp. 1–6, Mar. 2012.

[13] L. Zhang, D. Tjondronegoro, and V. Chandran, “Evaluation of texture and geometry for dimensional facial expression recognition,” *Proc. - 2011 Int. Conf. Digit. Image Comput. Tech. Appl. DICTA 2011*, pp. 620–626, 2011.

[14] R. Verma and M. Y. Dabbagh, “Fast Facial Expression Recognition Based On Local Binary Patterns,” in *2013 26th IEEE Canadian Conference Of Electrical And Computer Engineering (CCECE)*, 2013, pp. 1–4.

[15] F. Bashar, A. Khan, F. Ahmed, and M. H. Kabir, “Robust facial expression recognition based on median ternary pattern (MTP),” *2013 Int. Conf. Electr. Inf. Commun. Technol.*, pp. 1–5, Feb. 2013.

[16] X. Wang, C. Jin, W. Liu, M. Hu, L. Xu, and F. Ren, “Feature Fusion of HOG and WLD for Facial Expression Recognition,” *Syst. Integr. (SII), 2013 IEEE/SICE Int. Symp.*, pp. 227–232, 2013.

[17] T. H. H. Zavaschi, A. S. Britto, L. E. S. Oliveira, and A. L. Koerich, “Fusion of feature sets and classifiers for facial expression recognition,” *Expert Syst. Appl.*, vol. 40, no. 2, pp. 646–655, Feb. 2013.

[18] M. Kyperountas, A. Tefas, and I. Pitas, “Pairwise facial expression classification,” *2009 IEEE Int. Work. Multimed. Signal Process.*, pp. 1–4, Oct. 2009.

[19] Z. Zeng, M. Pantic, G. I. Roisman, and T. S. Huang, “A survey of affect recognition methods: Audio, visual, and spontaneous expressions,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 31, no. 1, pp. 39–58, 2009.

[20] I. Megvii, “Face++,” 2015. [Online]. Available: http://www.faceplusplus.com/. [Accessed: 15-Feb-2015].

[21] X. Feng, B. Lv, Z. Li, and J. Zhang, “A Novel Feature Extraction Method for Facial Expression Recognition,” *Proc. 9th Jt. Conf. Inf. Sci.*, pp. 32–35, 2006.

[22] J. Ou, X. Bai, Y. Pei, L. Ma, and W. Liu, “Automatic Facial Expression Recognition Using Gabor Filter and Expression Analysis,” *2010 Second Int. Conf. Comput. Model. Simul.*, pp. 215–218, Jan. 2010.

[23] K. T. Song and S.-C. Chien, “Facial expression recognition based on mixture of basic expressions and intensities,” *2012 IEEE Int. Conf. Syst. Man, Cybern.*, pp. 3123–3128, Oct. 2012.

[24] C. Shan, S. Gong, and P. W. Mcowan, “Recognizing Facial Expressions at Low Resolution,” in *IEEE Conference on Advanced Video and Signal Based Surveillance*, 2005, pp. 330–335.

[25] M. W. Huang, Z. W. Wang, and Z. L. Ying, “A novel method of facial expression recognition based on GPLVM Plus SVM,” *IEEE 10th Int. Conf. Signal Process. Proc.*, no. 4, pp. 916–919, Oct. 2010.

[26] T. Kanade, J. Cohn, and Y. Tian, “Comprehensive database for facial expression analysis,” *Automatic Face and Gesture Recognition, 2000. Proceedings. Fourth IEEE International Conference on*, Grenoble, pp. 484–491, 28-Mar-2000.

[27] M. J. Lyons, “Automatic classification of single facial images,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 21, no. 12, pp. 1357–1362, 1999.

[28] “The Taiwanese Facial Expression Image Database.” [Online]. Available: http://bml.ym.edu.tw/tfeid/. [Accessed: 01-Jan-2015].

[29] P. Ekman and W. V. Friesen, “Constants across cultures in the face and emotion,” *J. Pers. Soc. Psychol.*, vol. 17, no. 2, pp. 124–129, Feb. 1971.

[30] F. G. A. M. da Costa, “Reconhecimento de Expressões Faciais,” Universidade de Trás-Os-Montes e Alto Douro, 2010.

[31] S. C. Tai and K. C. Chung, “Automatic facial expression recognition system using Neural Networks,” *TENCON 2007 - 2007 IEEE Reg. 10 Conf.*, vol. 2, no. 1, pp. 113–118, 2007.

[32] C. J. C. Juanjuan, Z. Z. Z. Zheng, S. H. S. Han, and Z. G. Z. Gang, “Facial expression recognition based on PCA reconstruction,” *Comput. Sci. Educ. (ICCSE), 2010 5th Int. Conf.*, pp. 195–198, Aug. 2010.

[33] R. Qasim, M. M. Shirazi, N. Arshad, I. Qureshi, and S. Zaidi, “Comparison and improvement of PCA and LBP efficiency for face recognition,” *2013 3rd IEEE Int. Conf. Comput. Control Commun.*, pp. 1–6, Sep. 2013.

[34] S. M. Lajevardi and M. Lech, “Facial expression recognition from image sequences using optimized feature selection,” *2008 23rd Int. Conf. Image Vis. Comput. New Zeal.*, pp. 1–6, 2008.

[35] Y. T. Y. Tian, “Evaluation of Face Resolution for Expression Analysis,” *2004 Conf. Comput. Vis. Pattern Recognit. Work.*, pp. 0–6, 2004.

[36] C. Martin, U. Werner, and H.-M. Gross, “A real-time facial expression recognition system based on Active Appearance Models using gray images and edge images,” *2008 8th IEEE Int. Conf. Autom. Face Gesture Recognit.*, pp. 1–6, 2008.

[37] P. Viola and M. Jones, “Rapid Object Detection using a Boosted Cascade of Simple Features,” in *Conference on Computer Vision and Pattern Recognition*, 2001, vol. 1, pp. I–511 – I–518.

[38] M. C. Sobia, V. Brindha, and A. Abudhahir, “Facial expression recognition using PCA based interface for wheelchair,” *Electron. Commun. Syst. (ICECS), 2014 Int. Conf.*, pp. 1–6, 2014.

[39] H. Sadeghi, A.-A. Raie, and M.-R. Mohammadi, “Facial expression recognition using geometric normalization and appearance representation,” in *2013 8th Iranian Conference on Machine Vision and Image Processing (MVIP)*, 2013, pp. 159–163.

[40] H. Deng, L. Jin, L. Zhen, and J. Huang, “A new facial expression recognition method based on local gabor filter bank and pca plus lda,” *Int. J. Inf. Technol.*, vol. 11, no. 11, pp. 86–96, 2005.

[41] Y. Z. Y. Zilu and Z. G. Z. Guoyi, “Facial Expression Recognition Based on NMF and SVM,” *2009 Int. Forum Inf. Technol. Appl.*, vol. 3, pp. 612–615, May 2009.

[42] X. Tan and B. Triggs, “Enhanced local texture feature sets for face recognition under difficult lighting conditions,” *IEEE Trans. Image Process.*, vol. 19, no. 6, pp. 1635–1650, 2010.

[43] Y.-L. T. Y.-L. Tian, T. Kanada, and J. F. Cohn, “Recognizing upper face action units for facial expression analysis,” *Proc. IEEE Conf. Comput. Vis. Pattern Recognition. CVPR 2000 (Cat. No.PR00662)*, vol. 1, no. 2, pp. 1–19, 2000.

[44] Y. T. Lisa, B. Arun, H. Sharat, P. Andrew, and R. Bolle, “Real World Real-time Automatic Recognition of Facial Expressions,” *Most*, 2003.

[45] C. Shan, S. Gong, and P. W. McOwan, “Facial expression recognition based on Local Binary Patterns: A comprehensive study,” *Image Vis. Comput.*, vol. 27, no. 6, pp. 803–816, May 2009.

[46] Z. Huang and F. Ren, “Facial Expression Recognition basead on Active Appearance Model And Scale-Invariant Feature Transform,” pp. 94–99, 2013.

[47] C. V. Ramireddy and K. V. K. Kishore, “Facial expression classification using Kernel based PCA with fused DCT and GWT features,” *2013 IEEE Int. Conf. Comput. Intell. Comput. Res.*, pp. 1–6, Dec. 2013.

[48] B. K. Dehkordi and J. Haddadnia, “Facial expression recognition with optimum accuracy based on Gabor filters and geometric features,” *Signal Process. Syst. (ICSPS), 2010 2nd Int. Conf.*, vol. 1, pp. V1–731–V1–733, Jul. 2010.

[49] Y. S. Huang, S. H. Chuang, and F. H. Cheng, “An AdaBoost-based facial expression recognition method,” *Mach. Learn. Cybern. (ICMLC), 2011 Int. Conf.*, vol. 4, pp. 10–13, 2011.

[50] K. Cho, Y. Kim, and Y. Lee, “Real-time Expression recognition System using Active Appearance Model and EFM,” in *2006 International Conference on Computational Intelligence and Security*, 2006, pp. 747–750.

[51] H. C. Choi and S.-Y. Oh, “Facial Identity and Expression Recognition by using Active Appearance Model with Efficient Second Order Minimization and Neural Networks,” *2007 Int. Symp. Comput. Intell. Robot. Autom.*, pp. 131–136, Jun. 2007.

[52] I. Cohen, N. Sebe, A. Garg, L. S. Chen, and T. S. Huang, “Facial expression recognition from video sequences: Temporal and static modeling,” *Comput. Vis. Image Underst.*, vol. 91, no. 1–2, pp. 160–187, 2003.

[53] S. Liu, Y. Zhang, and K. Liu, “Facial expression recognition under partial occlusion based on Weber Local Descriptor histogram and decision fusion,” *Proc. 33rd Chinese Control Conf.*, pp. 4664–4668, Jul. 2014.

[54] T. Ojala, M. Pietikäinen, and T. Mäenpää, “Multiresolution gray-scale and rotation invariant texture classification with local binary patterns,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 7, pp. 971–987, 2002.

[55] J. Chen, S. Shan, C. He, G. Zhao, S. Member, X. Chen, and M. Pietika, “WLD : A Robust Local Image Descriptor,” vol. 32, no. 9, pp. 1705–1720, 2010.

[56] J. Chen, S. Shan, C. He, G. Zhao, M. Pietikäinen, S. Member, X. Chen, and W. Gao, “WLD : A Robust Local Image Descriptor,” *Pattern Anal. Mach. Intell. IEEE Trans.*, vol. 32, no. 9, pp. 1705–1720, 2009.

[57] M. Abdulrahman, T. R. Gwadabe, F. J. Abdu, and A. Eleyan, “Gabor wavelet transform based facial expression recognition using PCA and LBP,” in *2014 22nd Signal Processing and Communications Applications Conference (SIU)*, 2014, no. Siu, pp. 2265–2268.

[58] C. Lin, F. Peng, B. Wang, W. Sun, and X. Kong, “Research on PCA and KPCA Self-Fusion Based MSTAR SAR Automatic Target Recognition Algorithm,” *J. Eletronic Sci. Technol.*, vol. 10, no. 4, pp. 352–357, 2012.

[59] X. Chen, J. Yang, and Z. Jin, “An Improved Linear Discriminant Analysis with L1-Norm for Robust Feature Extraction,” *2014 22nd Int. Conf. Pattern Recognit.*, pp. 1585–1590, 2014.

[60] Y. Cheon and D. Kim, “A Natural Facial Expression Recognition Using Differential-AAM and k-NNS,” *2008 Tenth IEEE Int. Symp. Multimed.*, pp. 220–227, Dec. 2008.

[61] M. Huang, Z. Wang, and Z. Ying, “Facial expression recognition using Stochastic Neighbor Embedding and SVMs,” *Proc. 2011 Int. Conf. Syst. Sci. Eng.*, no. June, pp. 671–674, 2011.

[62] H. Mliki, N. Fourati, S. Smaoui, and M. Hammami, “Automatic Facial Expression Recognition System,” *Computer Systems and Applications (AICCSA), 2013 ACS International Conference on*, Ifrane, pp. 1–4, 27-May-2013.

[63] D. S. Chen and Z. K. Liu, “Generalized Haar-Like Features for Fast Face Detection,” in *Machine Learning and Cybernetics, 2007 International Conference on*, 2007, no. August, pp. 19–22.

[64] R. Lienhart and J. Maydt, “An extended set of Haar-like features for rapid object detection,” *Proceedings. Int. Conf. Image Process.*, vol. 1, pp. I–900–I–903, 2002.

[65] T. Ojala, M. Pietikäinen, and D. Harwood, “A comparative study of texture measures with classification based on featured distributions,” *Pattern Recognit.*, vol. 29, no. l, pp. 51–59, 1996.

[66] S. L. Happy, A. George, and A. Routray, “A real time facial expression classification system using Local Binary Patterns,” *4th Int. Conf. Intell. Hum. Comput. Interact.*, pp. 1–5, 2012.

[67] G. Cheng, Y. Fang, Y. Tan, W. Dai, and Q. Cai, “A local difference coding algorithm for face recognition,” in *Proceedings - 4th International Congress on Image and Signal Processing, CISP 2011*, 2011, vol. 2, pp. 828–832.

[68] D. D. Souza and R. V Yampolskiy, “Natural vs Artificial Face Classification using Uniform Local Directional Patterns and Wavelet Uniform Local Directional Patterns,” *Comput. Vis. Pattern Recognit. Work. (CVPRW), 2014 IEEE Conf.*, pp. 27 – 33, 2014.

[69] a. Jain and D. Zongker, “Feature selection: evaluation, application, and small sample performance,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 19, no. 2, pp. 153–158, 1997.

[70] M. S. Sainin and R. Alfred, “A genetic based wrapper feature selection approach using Nearest Neighbour Distance Matrix,” *Conf. Data Min. Optim.*, no. June, pp. 237–242, 2011.

[71] H.-H. Hsu, C.-W. Hsieh, and M.-D. Lu, “A Hybrid Feature Selection Mechanism,” *2008 Eighth Int. Conf. Intell. Syst. Des. Appl.*, vol. 2, 2008.

[72] C. N. Hsu, H. J. Huang, and D. Schuschel, “The ANNIGMA-wrapper approach to fast feature selection for neural nets,” in *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, 2002, vol. 32, no. 2, pp. 207–212.

[73] E. Cantú-Paz, S. Newsam, and C. Kamath, “Feature selection in scientific applications,” *Proc. tenth ACM SIGKDD Int. Conf. Knowl. Discov. data Min.*, pp. 788–793, 2004.

[74] S. Dinakaran and P. R. J. Thangaiah, “Comparative Analysis of Filter-Wrapper Approach for Random Forest Performance on Multivariate Data,” *2014 Int. Conf. Intell. Comput. Appl.*, pp. 174–178, 2014.

[75] G. Wang, F. H. Lochovsky, and Q. Yang, “Feature selection with conditional mutual information maximin in text categorization,” *Proc. Thirteen. ACM Conf. Inf. Knowl. Manag. - CIKM ’04*, pp. 342–349, 2004.

[76] L. Patil and M. Atique, “A novel feature selection based on information gain using WordNet,” in *Science and Information Conference*, 2013, pp. 625–629.

[77] L. Wei and W. Xiao, “Improved Method of Feature Selection Based on Information Gain,” *Eng. Technol. (S-CET), 2012 Spring Congr.*, pp. 1–4, 2012.

[78] G. Muhammad, M. Hussain, F. Alenezy, A. M. Mirza, G. Bebis, and H. Aboalsamh, “Race Recognition Using Local Descriptors,” *Acoust. Speech Signal Process. (ICASSP), 2012 IEEE Int. Conf.*, pp. 1525–1528, 2012.

[79] Q. Zhu, L. Lin, M. L. Shyu, and S. C. Chen, “Feature Selection Using Correlation and Reliability Based Scoring Metric for Video Semantic Detection,” *2010 IEEE Fourth Int. Conf. Semant. Comput.*, pp. 462–469, Sep. 2010.

[80] M. A. Hall, “Correlation-based Feature Selection for Discrete and Numeric Class Machine Learning,” Hamilton, 2000.

[81] M. A. Hall, “Correlation-based Feature Selection for Machine Learning,” The University of Waikato, 1999.

[82] V. N. Vapnik, *Statistical Learning Theory*. New York: Wiley-Interscience, 1998.

[83] E. Fix and J. L. J. Hodges, *Discriminatory Analysis. Nonparametric Discrimination: Consistency Properties*, 3rd ed. International Statistical Institute (ISI), 1989.

[84] K. Facelli, A. C. Lorena, J. Gama, and A. C. P. L. F. Carvalho, *Inteligência Artificial: Uma Abordagem de Aprendizagem de Máquina*, 1st ed. Rio de Janeiro: LTC - Livros Técnicos e Científicos Editora LTA., 2011.

[85] A. Singh, A. Yadav, and A. Rana, “K-means with Three different Distance Metrics,” *Int. J. Comput. Appl.*, vol. 67, no. 10, pp. 13–17, 2013.

[86] D. Michie, D. J. Spiegelhalter, and C. C. Taylor, *Machine Learning , Neural and Statistical Classification*, 1st ed. New York, NY, USA: ACM New York, 1994.

[87] C. J. C. Burges, “A Tutorial on Support Vector Machines for Pattern Recognition,” *Data Min. Knowl. Discov.*, vol. 2, pp. 121–167, 1998.

[88] C. J. C. Junli and J. L. J. Licheng, “Classification mechanism of support vector machines,” *WCC 2000 - ICSP 2000. 2000 5th Int. Conf. Signal Process. Proceedings. 16th World Comput. Congr. 2000*, vol. 3, pp. 0–3, 2000.

[89] Z. Zhang, C. Fang, and X. Ding, “Facial expression analysis across databases,” *2011 Int. Conf. Multimed. Technol.*, pp. 317–320, Jul. 2011.

[90] X. Li, Q. Ruan, G. An, and Y. Jin, “Automatic 3D facial expression recognition based on polytypic Local Binary Pattern,” in *2014 12th International Conference on Signal Processing (ICSP)*, 2014, pp. 1030–1035.

[91] M. K. Chmarra, A. a Á. Cabrera, T. Van Beek, V. D’Amelio, M. S. Erden, and T. Thomiyama, “Revisiting the divide and conquer strategy to deal with complexity in product design,” *2008 IEEE/ASME Int. Conf. Mechatronics Embed. Syst. Appl. MESA 2008*, pp. 393–398, 2008.

[92] C. S. Dhir, N. Iqbal, and S. Lee, “Efficient feature selection based on information gain criterion for face recognition,” in *Information Acquisition, 2007. ICIA ’07. International Conference on*, 2007, pp. 523–527.

[93] P. Lucey, J. F. Cohn, T. Kanade, J. Saragih, Z. Ambadar, and I. Matthews, “The extended Cohn-Kanade dataset (CK+): A complete dataset for action unit and emotion-specified expression,” *2010 IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. - Work. CVPRW 2010*, no. July, pp. 94–101, 2010.

[94] H. Yan, M. H. Ang, and A. N. Poo, “Cross-dataset facial expression recognition,” *2011 IEEE Int. Conf. Robot. Autom.*, pp. 5985–5990, 2011.

[95] H. Ghaderi and P. Kabiri, “Fourier transform and correlation-based feature selection for fault detection of automobile engines,” *AISP 2012 - 16th CSI Int. Symp. Artif. Intell. Signal Process.*, no. Aisp, pp. 514–519, 2012.

[96] F. Wallhoff, “Facial Expressions and Emotion Database,” *Technische Universität München*. [Online]. Available: http://www.mmk.ei.tum.de/~waf/fgnet/feedtum.html. [Accessed: 01-Jan-2006].

[97] M. Pantic, M. Valstar, R. Rademaker, and L. Maat, “Web-based database for facial expression analysis,” *Multimed. Expo, 2005. ICME 2005. IEEE Int. Conf.*, pp. 1–5, 2005.