

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
INSTITUTO DE INFORMÁTICA
PROGRAMA DE PÓS-GRADUAÇÃO EM COMPUTAÇÃO

JOHN SOLDERA

**Reconhecimento Facial com Projeções
Ortogonais Preservadoras de Localidade
Customizadas para Maximizar Margens
Suaves**

Tese apresentada como requisito parcial
para a obtenção do grau de
Doutor em Ciência da Computação

Prof. Dr. Jacob Scharcanski
Orientador

Porto Alegre, novembro de 2015

CIP – CATALOGAÇÃO NA PUBLICAÇÃO

John Soldera,

Reconhecimento Facial com Projeções Ortogonais Preservadoras de Localidade Customizadas para Maximizar Margens Suaves /

John Soldera. – Porto Alegre: PPGC da UFRGS, 2015.

67 f.: il.

Tese (doutorado) – Universidade Federal do Rio Grande do Sul. Programa de Pós-Graduação em Computação, Porto Alegre, BR–RS, 2015. Orientador: Jacob Scharcanski.

1. Reconhecimento Facial. 2. Biometria. 3. Máquinas de Vetores de Suporte. 4. Métodos de Subespaço. 5. Reconhecimento de Padrões. I. Scharcanski, Jacob. II. Título.

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL

Reitor: Prof. Carlos Alexandre Netto

Pró-Reitor de Coordenação Acadêmica: Prof. Rui Vicente Oppermann

Pró-Reitora de Pós-Graduação: Prof^ª. Vladimir Pinheiro do Nascimento

Diretor do Instituto de Informática: Prof. Luís da Cunha Lamb

Coordenador do PPGC: Prof. Luigi Carro

Bibliotecária-chefe do Instituto de Informática: Beatriz Regina Bastos Haro

Face Recognition Using Customized Orthogonal Locality Preserving Projections with Soft Margin Maximization

ABSTRACT

Nowadays, face recognition by automatic techniques still is a challenging task since face images may be affected by changes in the scene, such as in the illumination, head pose or face expression. Also, face feature representation often requires several dimensions, which poses additional challenges for face recognition. In this thesis is proposed a novel face recognition method with the objective of to be robust to many issues which can affect the face features in practice and it is based on projections of high dimensional face image representations into lower dimensionality and highly discriminative spaces. This is achieved by a modified Orthogonal Locality Preserving Projections (OLPP) method that uses a supervised alternative locality definition scheme designed to preserve the face class (individuals) structure in the obtained lower dimensionality face feature space unlike the typical OLPP method which preserves the face data structure. Besides, a new kernel equation is proposed to calculate affinities among face samples, presenting better class structure preservation when compared to the heat kernel used by the typical OLPP method. The proposed method can work with sparse and dense face image representations (i.e. it can use sub-sets or all face image pixels), and a sparse and a dense feature extraction methods are proposed, which preserve the color information during the feature extraction process from the facial images improving on the typical OLPP method which uses grayscale low-resolution face images. New test face images are classified in the obtained lower dimensionality feature space using a trained soft margins Support Vector Machine (SVM), so it performs better than the nearest neighbor rule used in the typical OLPP method. A set of experiments was designed to evaluate the proposed method under various conditions found in practice (such as changes in head pose, face expression, illumination, and in the presence of occlusion artifacts). The experimental results were obtained using five challenging public face databases (namely, PUT, FEI, FERET, Yale and ORL). These experiments confirm that the proposed feature extraction method integrated to the proposed transformation to a discriminative lower dimensionality space using the alternative classification scheme with SVM and soft margins obtains higher recognition rates than the OLPP method itself and methods representative of the state-of-the-art even when are used color (RGB) face images in high resolution (PUT, FEI and FERET face databases) as well as grayscale face images in low resolution (Yale and ORL face databases).

Keywords: Face Recognition, Biometrics, Support Vector Machines, Sub-space Methods, Pattern Recognition.

RESUMO

Atualmente, o reconhecimento facial por técnicas automáticas é ainda uma tarefa desafiadora uma vez que as imagens faciais podem ser afetadas por mudanças na cena, tais como na iluminação, na pose da cabeça, ou na expressão facial. Além disso, a representação de faces por feições faciais geralmente requer diversas dimensões, o que impõe desafios adicionais ao reconhecimento facial. Nessa tese, é proposto um novo método de reconhecimento facial com o objetivo de ser robusto a muitos dos fatores que podem afetar as feições faciais na prática e se baseia em determinar transformações do espaço original de feições faciais de alta dimensionalidade para um espaço de baixa dimensionalidade que apresenta maior discriminação das classes de dados faciais (indivíduos). Isso é realizado através da aplicação de um método *Projeções Ortogonais Preservadoras de Localidade* (*Orthogonal Locality Preserving Projections* - OLPP) modificado, que usa esquemas de definição de localidade supervisionados que têm o objetivo de preservar a estrutura das classes de dados faciais no espaço resultante de baixa dimensionalidade, diferentemente do método OLPP típico que preserva a estrutura dos dados faciais. Dessa forma, as classes se tornam mais compactas, preservando a métrica de classificação. O método proposto pode trabalhar tanto com representações densas como esparsas de imagens faciais (ou seja, ele pode usar subconjuntos ou todos os *pixels* das imagens faciais), sendo proposto nessa tese um método de extração de feições faciais esparsas e um método de extração de feições faciais densas que preservam a informação de cor das imagens faciais apresentando melhora em relação ao método OLPP típico que usa imagens em escalas de cinza em baixa resolução. Novas imagens faciais de teste são classificadas no espaço de baixa dimensionalidade obtido usando *Máquinas de Vetores de Suporte* (*Support Vector Machines* - SVM) treinadas com margens suaves, apresentando maior eficiência do que a regra do vizinho mais próximo usada no método OLPP típico. Um conjunto de experimentos foi projetado para avaliar o método proposto sob várias condições encontradas na prática (como mudanças na pose, expressão facial, iluminação e a presença de artefatos que causam oclusão facial). Os resultados experimentais foram obtidos usando cinco bases de imagens faciais públicas (a PUT, a FEI, a FERET, a Yale e a ORL). Esses experimentos confirmam que os esquemas propostos de extração de feições faciais integrados à transformação proposta para um espaço discriminativo de baixa dimensionalidade empregando o esquema alternativo de classificação usando SVM com margens suaves obtêm maiores taxas de reconhecimento do que o próprio método OLPP e métodos representativos do estado da arte mesmo quando são usadas imagens coloridas em alta resolução (das bases de imagens faciais PUT, FEI e FERET) como imagens faciais em escalas de cinza em baixa resolução (das bases Yale e ORL).

Palavras-chave: Reconhecimento Facial, Biometria, Máquinas de Vetores de Suporte, Métodos de Subespaço, Reconhecimento de Padrões.

LISTA DE ABREVIATURAS E SIGLAS

AMI	Informação Mútua (<i>Mutual Information</i>)
ASM	Modelos de Forma Ativos (<i>Active Shape Models</i>)
LDA	Análise Discriminante Linear (<i>Linear Discriminant Analysis</i>)
LPP	Projeções Preservadoras de Localidade (<i>Locality Preserving Projections</i>)
MvDA	Análise Discriminante Multi-Visões (<i>Multi-view Discriminant Analysis</i>)
NN	Vizinho Mais Próximo (<i>Nearest Neighbor</i>)
OLPP	Projeções Ortogonais Preservadoras de Localidade (<i>Orthogonal Locality Preserving Projections</i>)
OLPP1	Definição Clássica de Localidade no OLPP
OLPP2	Definição Proposta de Localidade no OLPP
OLPP3	Definição Proposta de Localidade no OLPP
OLPP4	Definição Proposta de Localidade no OLPP
ONPP	Projeções Ortogonais Preservadoras de Vizinhança (<i>Orthogonal Neighborhood Preserving Projections</i>)
PCA	Análise das Componentes Principais (<i>Principal Component Analysis</i>)
PDM	Modelo de Distribuição de Pontos (<i>Point Distribution Model</i>)
PIN	Número de Identificação Pessoal (<i>Personal Identification Number</i>)
RVM	Máquinas de Vetores Relevantes (<i>Relevance Vector Machines</i>)
SRKDA	Análise Discriminante com Regressão Espectral com Kernels (<i>Spectral Regression Kernel Discriminant Analysis</i>)
SVD	Decomposição em Valores Singulares (<i>Singular Value Decomposition</i>)
SVM	Máquinas de Vetores de Suporte (<i>Support Vector Machines</i>)

LISTA DE SÍMBOLOS

h	Índice de uma Imagem Facial no Conjunto de Imagens de Treinamento.
s	Índice de uma Imagem Facial no Conjunto de Imagens de Treinamento.
x_h	h -Ésima Imagem Facial no Conjunto de Imagens de Treinamento.
x_s	s -Ésima Imagem Facial no Conjunto de Imagens de Treinamento.
x_m	Média das Imagens Faciais no Conjunto de Imagens de Treinamento.
n	Quantidade de Imagens Faciais.
X	Matriz com todas as Imagens Faciais de Treinamento.
e	Número de Autovetores no Método PCA (e nas suas Aplicações).
S	Matriz de Afinidades (Similaridades).
S_{hs}	Afinidade entre as Imagens Faciais x_h e x_s .
ϵ	Raio Máximo de Definição de Vizinhaça.
t	Largura de Banda dos Dados Faciais.
k	Quantidade de Vizinhos Usada na Definição de Vizinhaça.
T	Operador Matriz Transposta.
L	Matriz Laplaciana do Grafo de Vizinhaça.
D	Matriz Diagonal com o Grau dos Vértices.
W_{LPP}	Matriz de Transformação do método LPP.
p	Índice de um Autovetor em uma Base de Autovetores.
\vec{g}_p	p -Ésimo Autovetor do Espaço de Baixa Dimensionalidade.
d	Dimensionalidade Escolhida para o Espaço de Baixa Dimensionalidade.
\vec{g}_d	d -Ésimo Autovetor do Espaço de Baixa Dimensionalidade.
x	Imagem Facial no Espaço Original de Feições.
y	Imagem Facial x Projetada em um Espaço de Baixa Dimensionalidade.
W	Matriz de Transformação para o Espaço de Baixa Dimensionalidade.
W_{PCA}	Matriz de Transformação do Método PCA.
\vec{v}_e	O e -Ésimo Autovetor.

$G^{(p-1)}$	Matriz Parcial de Autovetores Contendo Até o $(p - 1)$ -Ésimo Autovetor no Método OLPP.
$H^{(p-1)}$	Matriz Intermediária no Cálculo do p -Ésimo Autovetor no Método OLPP.
$M^{(p)}$	Matriz para Cálculo do p -Ésimo Autovetor no Método OLPP.
W_{OLPP}	Matriz de Transformação do Método OLPP.
\mathcal{E}	Erro de Reconstrução de Imagens Faciais.
P	Matriz de Pesos de Reconstrução de Coordenadas de Imagens Faciais.
p_{hs}	Peso que x_h Tem na Reconstrução de x_s .
\mathcal{Y}	Matriz com o Erro de Reconstrução de cada Imagem Facial y_h .
I	Matriz Identidade.
z	Direção que Maximiza a Separabilidade das Classes no Método LDA.
S_B	Matriz de Dispersão entre Classes (<i>Between-Class Scatter Matrix</i>).
S_W	Matriz de Dispersão Intra-Classes (<i>Within-Class Scatter Matrix</i>).
A	Quantidade de Classes de Dados Faciais Disponíveis no Conjunto de Dados de Treinamento.
B	Quantidade de Imagens Faciais por Classe de Dados Faciais.
a	a -Ésima Classe de Dados Faciais Presente na Base de Treinamento.
b	b -Ésima Imagem Facial Pertencente a uma Determinada Classe de Dados Faciais.
μ_a	Média das Imagens Faciais da Classe de Dados Faciais a .
μ_t	Média das Imagens Faciais de todas as Classes de Dados Faciais.
x_{ab}	b -Ésima Imagem Facial Pertencente à a -Ésima Classe de Dados Faciais.
l	l -Ésima Dimensão Obtida no Método LDA.
\vec{z}_l	l -Ésimo Autovetor da Projeção Discriminativa Obtida no Método LDA.
λ_l	l -Ésima Constante no Cálculo de Autovetores no Método LDA.
κ	Função de Kernel.
K	Matriz de Kernel no Método SRKDA.
K_{hs}	Valor Obtido com a Aplicação da Função de Kernel κ às Imagens Faciais x_h e x_s .
σ	Desvio Padrão.
α_k	Constante do Kernel Sigmoid.
R	Matriz de Localidade Definida no Método SRKDA.
$Tr(\dots)$	Traço de uma Matriz.
W_{MvDA}	Matriz de Transformação do método MvDA.
D_{MvDA}	Matriz de Dispersão entre Classes do método MvDA.

S_{MvDA}	Matriz de Dispersão Intra-Classes do método MvDA.
$AMI(...)$	Cálculo da Informação Mútua que Compara Dois <i>Clusters</i> de Dados.
o	o -Ésimo Canal de Cor.
Z	Quantidade de Canais de Cores Disponíveis em uma Imagem Facial.
c	c -Ésimo <i>Landmark</i> de um PDM (ou Topologia de <i>Landmarks</i>).
C	Quantidade de <i>Landmarks</i> presentes em um PDM (ou Topologia de <i>Landmarks</i>).
$\vec{\mu}_{F_{o,c}}$	Vetor com as Médias de Intensidade de Cor do o -Ésimo Canal de Cor Apresentadas na Vizinhança do c -Ésimo Ponto no PDM Considerando todas as Imagens Faciais de Treinamento.
a_c	Peso do c -Ésimo <i>Landmark</i> de um PDM.
μ_c	Média na Vizinhança do <i>Landmark</i> c em todas as Imagens Faciais de Treinamento.
σ_c	Desvio Padrão na Vizinhança do <i>Landmark</i> c em todas as Imagens Faciais de Treinamento.
Q	Subconjunto de Q <i>Landmarks</i> de um Total de P <i>Landmarks</i> .
P	Número de <i>Landmarks</i> Totais de uma Topologia.
E	Quantidade de Topologias Possíveis Usando Q Combinações de P <i>Landmarks</i> Totais.
V	Tamanho das Janelas de <i>Pixels</i> Extraídas nas Vizinhanças dos <i>Landmarks</i> .
$ $	Operador de Concatenação para Formar um Vetor.
f	f -Ésimo Canal de Cor de uma Imagem Facial (sendo R, G ou B).
x_{ab}^f	Feições Extraídas do f -Ésimo Canal de Cor da b -Ésima Imagem Facial da a -Ésima Classe de Dados Faciais.
x_{abc}^f	Feições Extraídas do f -Ésimo Canal de Cor do c -Ésimo <i>Landmark</i> da b -Ésima Imagem Facial da a -Ésima Classe de Dados Faciais.
M	Comprimento de uma Imagem Facial.
N	Largura de uma Imagem Facial.
$S_{ab}^{a'b'}$	Afinidade entre a Imagem Facial x_{ab} e a Imagem Facial $x_{a'b'}$.
U_{SVD}	Matriz com Autovetores do SVD.
S_{SVD}	Matriz com Autovalores do SVD.
V_{SVD}	Matriz com Autovetores do SVD.
r	Porcentagem de Preservação de Autovalores no Método PCA.
λ	Constante Usada no Cálculo de Autovetores.
\vec{v}	Vetor em um Problema de Autovetores.
d_{max}	Número Máximo de Dimensões após a Aplicação de PCA.

β	Índice de Autovalor no método PCA usando SVD.
x_{uv}	Imagem Facial de Teste de Índice v com Classe de Dados Faciais u Desconhecida (no Espaço Original de Feições).
y_{uv}	Imagem Facial de Teste de Índice v com Classe de Dados Faciais u Desconhecida (no Espaço Discriminativo de Baixa Dimensionalidade).
m	Tamanho da Margem de Separação entre Duas Classes no Método SVM.
\vec{w}	Vetor Normal ao Hiperplano de Separação no Método SVM.
q	Deslocamento do Hiperplano de Separação em Relação à Origem no Método SVM.
$sign(\dots)$	Função Sinal (Retorna 1 ou -1).
x_{ij}	j -Ésima Imagem Facial Pertencente a i -Ésima Classe de Dados Faciais no Espaço Original de Feições.
y_{ij}	j -Ésima Imagem Facial Pertencente a i -Ésima Classe de Dados Faciais no Espaço Discriminativo de Baixa Dimensionalidade.
z_i	Rótulo de uma Imagem Facial de Classe i no método SVM.
$F(\dots)$	Função de Classificação Aplicada a Imagens Faciais no Método SVM.
C_{SVM}	Constante de Ajuste de Precisão da Margem no Método SVM com Margens Suaves.
ξ_{ij}	Variável de Folga Usada no Problema de Otimização do Método SVM com Margens Suaves.
α	Constante de Ajuste da Importância nas Medições das Posições de <i>Landmarks</i> em Relação a Erros de Posicionamento no Método ASM Eficiente.

LISTA DE FIGURAS

Figura 2.1:	Extração Típica de Feições Densas de Imagens Faciais.	21
Figura 2.2:	Ilustração representando uma projeção bidimensional do espaço determinado pelas <i>Laplacianfaces</i> na base de imagens faciais Radboud com eixos (dimensões) denotados por z_1 e z_2	24
Figura 2.3:	Taxonomia de métodos em reconhecimento facial.	30
Figura 3.1:	Esquema com os fluxos de treinamento e teste do método proposto.	32
Figura 3.2:	Topologia de <i>landmarks</i> adotada nesse trabalho exemplificada em duas bases de imagens.	34
Figura 3.3:	Ilustração de um problema típico de SVM com duas classes.	40
Figura 4.1:	Poses disponíveis na base de imagens faciais PUT usadas nos experimentos.	45
Figura 4.2:	Poses disponíveis na base de imagens faciais FEI usadas nos experimentos.	45
Figura 4.3:	Poses disponíveis na base de imagens faciais FERET usadas nos experimentos.	46
Figura 4.4:	Poses disponíveis na base de imagens faciais Yale usadas nos experimentos.	46
Figura 4.5:	Poses disponíveis na base de imagens faciais ORL usadas nos experimentos.	47
Figura 4.6:	Tempos de processamento médios para treinamento e teste no método proposto (OLPP4 e SVM) em segundos.	57

LISTA DE TABELAS

Tabela 4.1:	Comparação de métodos aplicados a reconhecimento facial (identificação facial) usando representações esparsas na base PUT.	48
Tabela 4.2:	Comparação de métodos aplicados a reconhecimento facial (identificação facial) usando representações esparsas na base FEI.	48
Tabela 4.3:	Comparação de métodos aplicados a reconhecimento facial (identificação facial) usando representações esparsas na base FERET.	49
Tabela 4.4:	Comparação de métodos aplicados a reconhecimento facial (identificação facial) usando representações densas na base PUT.	50
Tabela 4.5:	Comparação de métodos aplicados a reconhecimento facial (identificação facial) usando representações densas na base FEI.	51
Tabela 4.6:	Comparação de métodos aplicados a reconhecimento facial (identificação facial) usando representações densas na base FERET.	51
Tabela 4.7:	Comparação de métodos aplicados à identificação facial usando a base de imagens faciais Yale na abordagem clássica densa com imagens em escala de cinza em baixa resolução.	52
Tabela 4.8:	Comparação de métodos aplicados à identificação facial usando a base de imagens faciais ORL na abordagem clássica densa com imagens em escala de cinza em baixa resolução.	53
Tabela 4.9:	Tempos de treinamento por imagem do método proposto usando os esquemas OLPP1, OLPP2, OLPP3 e OLPP4 com SVM (em segundos). "RGB" indica experimentos com imagens coloridas e "Cinza" indica experimentos com imagens em tons de cinza.	58
Tabela 4.10:	Tempos de teste por imagem do método proposto usando os esquemas OLPP1, OLPP2, OLPP3 e OLPP4 com SVM (em segundos). "RGB" indica experimentos com imagens coloridas e "Cinza" indica experimentos com imagens em tons de cinza.	58
Tabela 4.11:	Tempos de treinamento por imagem do método proposto usando os esquemas OLPP1, OLPP2, OLPP3 e OLPP4 com o classificador do vizinho mais próximo (em segundos). "RGB" indica experimentos com imagens coloridas e "Cinza" indica experimentos com imagens em tons de cinza.	58
Tabela 4.12:	Tempos de teste por imagem do método proposto usando os esquemas OLPP1, OLPP2, OLPP3 e OLPP4 com o classificador do vizinho mais próximo (em segundos). "RGB" indica experimentos com imagens coloridas e "Cinza" indica experimentos com imagens em tons de cinza.	59

SUMÁRIO

1	INTRODUÇÃO	14
2	RECONHECIMENTO FACIAL: CONCEITOS E ESTADO DA ARTE	19
2.1	Extração de Feições Faciais e Reconhecimento Facial	20
2.2	Métodos Baseados em Feições Densas	22
2.3	Métodos Baseados em Feições Esparsas	28
3	MÉTODO DE RECONHECIMENTO FACIAL PROPOSTO	31
3.1	Método Proposto	31
3.2	Extração de Feições Faciais Esparsas Coloridas	32
3.3	Extração de Feições Faciais Densas Coloridas	35
3.4	Definição de Vizinhança dos Dados Faciais	35
3.5	Redução de Dimensionalidade	37
3.6	Reconhecimento Facial	39
4	RESULTADOS EXPERIMENTAIS	42
4.1	Materiais e Métodos Usados nos Experimentos	44
4.2	Experimentos com Feições Faciais Esparsas	47
4.3	Experimentos com Feições Faciais Densas	49
4.4	Experimentos com Feições Faciais Densas em Imagens de Baixa Resolução	51
4.5	Discussão sobre os Resultados Experimentais	53
5	CONCLUSÕES E TRABALHOS FUTUROS	60
	REFERÊNCIAS	63

1 INTRODUÇÃO

Com a necessidade cada vez maior de segurança em sistemas e instalações, com a crescente evolução tecnológica de computadores e dispositivos de captura de informação biométrica disponíveis no mercado e com a pesquisa e evolução de novas técnicas para representar e reconhecer dados biométricos de alta dimensionalidade, aplicações que usam biometria estão se tornando cada vez mais robustas e confiáveis. Além disso, a autenticação de usuários baseada em biometria apresenta vantagens em relação a outros métodos atuais de validação de acesso a instalações e sistemas de informação, tais como o uso de senhas, PINs e cartões de acesso, pois elimina problemas como perda ou empréstimo de cartões de acesso, clonagem de PINs ou danos causados aos mesmos, o que poderia causar falhas como autenticar um usuário não autorizado ou ainda impossibilitar a autenticação do usuário verdadeiro.

Conseqüentemente, a tecnologia biométrica está atraindo o interesse de pesquisadores e engenheiros (SCHARCANSKI; PROENÇA; DU, 2014) (VERDOOLAEGE et al., 2014), particularmente em modalidades biométricas como reconhecimento de íris (RAHULKAR; HOLAMBE, 2012) (CHOU et al., 2010), leitura de impressões digitais (LEE et al., 2008) (ZHANG et al., 2011), reconhecimento facial (CAI et al., 2006) (BEHAINE; SCHARCANSKI, 2012) (MIKA et al., 1999) (BETTA et al., 2013), reconhecimento de gestos (DARDAS; GEORGANAS, 2011) (REN et al., 2013), entre outros. Além disso, as técnicas empregadas nessa área também são eficientes em aplicações semelhantes como estimativa de idade (GUO et al., 2009) (WU; TURAGA; CHELLAPPA, 2012), identificação de etnia (GUO; WANG, 2012) (LU et al., 2012), entre outros.

O reconhecimento facial é uma das modalidades de biometria menos invasivas disponíveis na atualidade quando comparado a outras modalidades de reconhecimento biométrico como íris, retina, mãos, digitais, voz e outros, pois tais modalidades biométricas exigem ações voluntárias dos usuários, sendo no reconhecimento facial necessário apenas que o usuário olhe diretamente para a câmera por um instante. Por outro lado, o reconhecimento facial é uma das modalidades biométricas mais desafiadoras, pois tem um grau maior de falsos positivos e falsos negativos devido a variações comuns nas imagens faciais, como mudanças na iluminação, na pose da cabeça, expressão facial, idade, ocorrência de oclusão, entre outros.

Na autenticação por biometria são utilizadas amostras coletadas anteriormente para servirem no processo de reconhecimento do usuário. Assim, na modalidade de biometria por reconhecimento facial, geralmente são usadas fotos coletadas anteriormente para servirem como conhecimento prévio no processo de reconhecimento de uma nova imagem facial do mesmo usuário, que podem sofrer de variações comuns nas feições faciais, dificultando o reconhecimento facial. Considera-se o reconhecimento facial menos eficiente que outras modalidades biométricas como reconhecimento de íris e de impressões

digitais, entretanto, nesse trabalho, pretende-se oferecer melhor tratamento aos problemas inerentes às feições faciais e elevar a robustez e confiabilidade do reconhecimento facial como aplicação biométrica.

Existem diversos métodos de reconhecimento facial atualmente disponíveis na literatura. Idealmente, um método de reconhecimento facial eficiente deve ser apto a processar diferentes tipos de imagens faciais, tanto em alta resolução, como em baixa resolução, e tanto em cores como em escalas de cinza. Além disso, um método de reconhecimento facial deve ser apto a processar adequadamente dados de alta dimensionalidade, que são as imagens faciais, sendo que as mesmas são intrinsecamente ruidosas e podem conter artefatos e/ou dados atípicos (por exemplo, mudanças na pose da cabeça, expressão facial, ocorrência de oclusão, entre outros). Infelizmente, muitos dos métodos propostos na literatura não têm um bom desempenho em tais condições de reconhecimento de imagens faciais, levando a imprecisões ou resultados errôneos. Nesse trabalho, propõe-se desenvolver um método de reconhecimento facial (identificação facial) que consiga processar adequadamente dados faciais de alta dimensionalidade, projetando-os para espaços discriminativos de baixa dimensionalidade que preservam as classes de dados faciais, enquanto é adotado um esquema alternativo de classificação de imagens faciais que seja robusto a muitos dos problemas inerentes às feições faciais que podem ocorrer na prática, ajudando a tratar melhor feições faciais ruidosas e dados atípicos.

Uma abordagem bastante comum usada para reduzir a dimensionalidade da representação de faces por feições faciais é a determinação de uma projeção de dados para um espaço de baixa dimensionalidade que compacte dados faciais enquanto ao mesmo tempo são preservadas características discriminativas dos dados faciais originais. Através de um processo de treinamento em que são usadas amostras faciais coletadas anteriormente, os métodos da abordagem baseada na aparência modelam a variabilidade dos dados faciais para criar uma projeção para um subespaço com dimensões reduzidas, onde é feita a classificação de novas imagens faciais (TURK; PENTLAND, 1991) (HE et al., 2005) (CAI et al., 2006) (BELHUMEUR; HESPANHA; KRIEGMAN, 1997).

O método *Eigenfaces* (TURK; PENTLAND, 1991) usa *Análise das Componentes Principais* (*Principal Component Analysis* - PCA) para mapear imagens faciais de baixa resolução em escala de cinza para um espaço de baixa dimensionalidade enquanto se propõe a preservar a maior parte da variabilidade dos dados faciais. Entretanto, o método *Eigenfaces* não garante a preservação da estrutura das classes de dados faciais (indivíduos) no espaço de baixa dimensionalidade (HE et al., 2005) (CAI et al., 2006). Alternativamente, o método *Laplacianfaces* (HE et al., 2005) foi proposto com o intuito de preservar a estrutura local dos dados faciais no espaço de baixa dimensionalidade, causando um aumento da discriminação das classes de dados faciais (indivíduos). Entretanto, a projeção obtida pelo método *Laplacianfaces* não é ortogonal, distorcendo a métrica de classificação. Como solução, método *Projeções Ortogonais Preservadoras de Localidade* (*Orthogonal Locality Preserving Projections* - OLPP) (CAI et al., 2006) modifica o método *Laplacianfaces* de forma que a matriz de projeção final seja ortogonal.

Os métodos mencionados acima garantem a preservação da estrutura local dos dados faciais no espaço de baixa dimensionalidade, entretanto não é garantida a separação das classes de dados faciais. Como em muitas aplicações práticas de reconhecimento facial, para cada classe de dados faciais são designadas imagens faciais de treinamento, a informação de classe dos dados faciais de treinamento já é conhecida, sendo possível determinar uma projeção para um espaço de baixa dimensionalidade com o objetivo de preservar a estrutura das classes de dados faciais no espaço de baixa dimensionalidade.

Alguns métodos procuram preservar a estrutura das classes de dados faciais no espaço de baixa dimensionalidade, como o método *Análise Discriminante Linear* (*Linear Discriminant Analysis* - LDA) (BELHUMEUR; HESPANHA; KRIEGMAN, 1997), onde os vetores de projeção *Fisherfaces* são geralmente obtidos maximizando a covariância entre as classes enquanto simultaneamente é minimizada a covariância interna das classes. Outro método baseado em LDA aplicado ao reconhecimento facial é o *Análise Discriminante Multi-Visões* (*Multi-view Discriminant Analysis* - MvDA) (KAN et al., 2012) que cria projeções discriminativas dos dados faciais sob diferentes perspectivas e as combina em uma única projeção linear final.

Os métodos apresentados anteriormente geralmente usam imagens em escalas de cinza que são reduzidas de tamanho para diminuir a complexidade computacional. Nessa abordagem, vetores de feições de alta dimensionalidade são criados concatenando todos os *pixels* das imagens faciais de forma que uma imagem facial seja um ponto em um espaço de alta dimensionalidade. Nesse processo de redução dos dados, detalhes sutis das feições faciais (imagens faciais) podem ser perdidos, impactando negativamente na discriminação das classes de dados faciais (indivíduos) no espaço de baixa dimensionalidade obtido pela projeção linear determinada. Além disso, geralmente esses métodos utilizam a regra do vizinho mais próximo (COVER; HART, 1967) para realizar a classificação de novas imagens faciais, porém a regra do vizinho mais próximo não é robusta a dados atípicos (por exemplo, quando ocorre mudança na iluminação ou na expressão facial) (DEVROYE; GYÖRFI; LUGOSI, 1996).

Em uma abordagem diferente, são determinados espaços discriminativos de baixa dimensionalidade usando mapeamentos não-lineares das imagens faciais, aplicando outros classificadores. Por exemplo, Pham e Venkatesh (PHAM; VENKATESH, 2011) propõem mapear imagens faciais para um *manifold de Grassmann* (ABSIL; MAHONY; SEPULCHRE, 2004), e realizar a classificação de novas imagens faciais usando *Máquinas de Vetores de Suporte* (*Support Vector Machines* - SVM). No método de Jia e Martinez (JIA; MARTINEZ, 2009) imagens faciais usando máscaras de oclusão são reconhecidas por SVM. Entretanto, tais métodos apresentam grande complexidade computacional, dificultando a sua aplicação prática, necessitando que as imagens faciais sejam reduzidas de tamanho.

Os métodos descritos acima processam imagens inteiras para criar representações vetoriais das mesmas, ou seja, todos os *pixels* das imagens faciais são concatenados, formando uma representação densa das imagens faciais, reduzindo o tamanho das imagens a fim de diminuir a complexidade computacional. Entretanto, em outra abordagem baseada em feições locais, feições são extraídas de localizações específicas nas imagens faciais através da identificação de um conjunto discreto de localizações (*landmarks*) demarcadas nas imagens faciais.

Os *landmarks* podem ser manualmente marcados ou automaticamente detectados usando vários métodos (EFRATY et al., 2011) (ZHAO et al., 2011) (RAPP et al., 2011) baseados em *Modelos de Forma Ativos* (*Active Shape Models* - ASM) (COOTES et al., 1995). Através dos *landmarks*, são extraídos conjuntos discretos de feições das imagens faciais. Alguns métodos que coletam informação esparsa das imagens faciais são o método *Modelo de Formas Ativas com Feições Ótimas Invariantes* (*Active shape models with invariant optimal features*) (SUKNO et al., 2007) e o método ASM Eficiente (BEHAINE; SCHARCANSKI, 2012) que designa diferentes pesos aos *landmarks*, estimando a capacidade discriminativa dos mesmos. O método *Estimação de Idade e Verificação Facial sob Envelhecimento Usando Landmarks* (*Age estimation and face verification across aging*

using landmarks) (WU; TURAGA; CHELLAPPA, 2012) extrai as posições dos *landmarks* como feições e as mapeia para um *manifold de Grassmann* (ABSIL; MAHONY; SEPULCHRE, 2004) e realiza a classificação facial no espaço tangente do *manifold* usando SVM e *Máquinas de Vetores Relevantes (Relevant Vector Machines - RVM)* (TIPPING, 2001). Tais métodos são muito sensíveis à oclusão e erros de posicionamento de *landmarks* e podem ser adversamente afetados por artefatos, como mudanças na pose da cabeça ou iluminação.

Nas décadas passadas, os dispositivos de captura de imagens faciais que eram utilizados em aplicações de reconhecimento facial obtinham imagens de baixa qualidade em escalas de cinza e em baixa resolução. Porém com o contínuo aperfeiçoamento dos dispositivos oferecidos no mercado como câmeras e processadores, a qualidade das imagens faciais obtidas aumentou consideravelmente fazendo que muitos dos métodos de reconhecimento facial disponíveis na literatura possam ter se tornados obsoletos, pois não possuíam informação relevante das imagens faciais que agora passaram a ser coloridas e em alta resolução. Um método de reconhecimento facial robusto deve ser capaz de processar os mais variados tipos de imagens faciais, desde imagens em escala de cinza em baixa resolução até imagens coloridas em alta resolução.

Nessa tese, é proposto um novo método de reconhecimento facial com o objetivo de ser robusto a muitos dos problemas inerentes às feições faciais que ocorrem em aplicações práticas, como mudanças na iluminação, na pose da cabeça ou na expressão facial. O método proposto realiza diversas melhorias em relação ao método OLPP típico (CAI et al., 2006). Primeiramente, é proposto um novo método de extração de feições faciais esparsas coloridas de imagens faciais, que processa imagens coloridas em alta resolução usando topologias de *landmarks* a fim de demarcar posições importantes nas imagens faciais que favorecem a discriminação de classes de dados faciais (indivíduos). A informação biométrica é coletada na vizinhança dos *landmarks*, descartando o restante dos *pixels* das imagens faciais, que contém informação redundante e ambígua. Como resultado desse novo esquema de extração de feições, são obtidos vetores de feições faciais de alta dimensionalidade assim como no método OLPP típico que converte imagens em escalas de cinza em baixa resolução para a representação vetorial.

Além disso, um novo método de extração de feições faciais densas coloridas é proposto e tem o objetivo de converter imagens coloridas para a representação vetorial concatenando todos os *pixels* das imagens faciais, diferentemente do método OLPP típico (CAI et al., 2006) que converte imagens em escalas de cinza em baixa resolução para a representação vetorial. Dessa forma, os vetores de feições obtidos no método proposto retém a informação de cor das feições faciais, assim possibilitando uma maior discriminação de classes de dados faciais (indivíduos) do que simples feições em escalas de cinza (CAI et al., 2006), pois a informação de cor é relevante no reconhecimento facial (BEHAINE; SCHARCANSKI, 2012) (SOLDERA; BEHAINE; SCHARCANSKI, 2015) (KOKIOPOULOU; SAAD, 2009).

Porém as feições faciais extraídas, tanto densas como esparsas, apresentam alta dimensionalidade e não apresentam uma boa discriminação de classes de dados faciais. A fim de tornar as feições faciais mais compactas e discriminativas, muitos métodos determinam uma projeção para um espaço de baixa dimensionalidade que preserva informação discriminativa das classes de dados faciais (TURK; PENTLAND, 1991) (HE et al., 2005) (CAI et al., 2006) (BELHUMEUR; HESPANHA; KRIEGMAN, 1997). Como preservar as classes de dados faciais no espaço de baixa dimensionalidade é melhor que preservar a estrutura dos dados faciais (SOLDERA; BEHAINE; SCHARCANSKI, 2015) (KOKI-

OPOULOU; SAAD, 2009), é proposta usar uma abordagem de reconhecimento facial supervisionada nessa tese usando um método *Projeções Ortogonais Preservadoras de Localidade* (*Locality Preserving Projections* - OLPP) modificado que usa esquemas propostos de definição de localidade supervisionados, preservando a estrutura das classes de dados faciais no espaço de baixa dimensionalidade, diferentemente do método OLPP típico que não é supervisionado, pois preserva apenas a estrutura dos dados faciais no espaço de baixa dimensionalidade. Dessa forma, as classes se tornam mais compactas, preservando melhor a métrica de classificação.

O método OLPP típico (CAI et al., 2006) e muitos outros métodos (TURK; PEN-TLAND, 1991) (HE et al., 2005) (CAI et al., 2006) (BELHUMEUR; HESPANHA; KRIGMAN, 1997) usam o classificador do vizinho mais próximo no espaço de baixa dimensionalidade para realizar a classificação de novas imagens faciais, porém esse classificador tipicamente usado na abordagem baseada na aparência pode apresentar falhas quando aplicado a dados de alta dimensionalidade ruidosos com ocorrência de dados atípicos (COVER; HART, 1967) (DEVROYE; GYÖRFI; LUGOSI, 1996). Portanto, é proposto usar um método de classificação alternativo aplicando SVM com margens suaves de forma a traçar as fronteiras entre classes de dados faciais distintas, permitindo ao método proposto lidar melhor com dados ruidosos e dados atípicos, apresentando maiores taxas de reconhecimento que o próprio método OLPP e métodos representativos do estado da arte.

Muitos sistemas de autenticação de usuários usando alguma modalidade de biometria incluem uma verificação de prova de vida, para que, além da autenticação biométrica, seja verificado se a fonte da informação biométrica é realmente uma pessoa, pois existe a possibilidade da fonte de informação ser uma foto ou um dispositivo que visa a simular as características biométricas do indivíduo. Entretanto, nessa tese, o foco é estudar técnicas que possam realizar reconhecimento facial de forma automática, ficando prova de vida a cargo de trabalhos futuros.

Outro objetivo do método proposto é apresentar um bom desempenho computacional tanto no treinamento como teste de novas imagens faciais, com o objetivo de ser aplicado em situações práticas de reconhecimento facial, necessitando apenas de uma câmera fotográfica e um computador para operar, podendo processar tanto imagens de baixa resolução em escalas de cinza, como imagens coloridas de alta resolução.

Essa tese é estruturada como segue. O Capítulo 2 apresenta conceitos e o estado da arte na área de reconhecimento facial. O Capítulo 3 apresenta o método proposto, onde a Seção 3.2 introduz o método proposto de extração de feições faciais coloridas esparsas baseado em topologias de *landmarks*, a Seção 3.3 apresenta o método proposto de extração de feições faciais coloridas densas, a Seção 3.4 apresenta os novos esquemas de definição de localidade propostos incluindo o kernel proposto para calcular afinidades entre imagens faciais, a Seção 3.5 apresenta a determinação do espaço discriminativo de baixa dimensionalidade onde são projetadas as imagens faciais e a Seção 3.6 apresenta a abordagem de classificação que usa SVM com margens suaves a fim de tratar melhor imagens ruidosas e dados atípicos. No Capítulo 4, o método proposto e diversos métodos representativos do estado da arte são comparados incluindo uma discussão sobre os resultados experimentais e as conclusões são apresentadas no Capítulo 5.

2 RECONHECIMENTO FACIAL: CONCEITOS E ESTADO DA ARTE

O reconhecimento facial é uma das abordagens de reconhecimento biométrico mais utilizadas na atualidade e pode ser dividido em duas categorias, onde a primeira é a verificação facial, onde é verificado se uma nova imagem facial de teste pertence a um determinado usuário (indivíduo), sendo um problema de correspondência de um-para-um. A segunda categoria é a identificação facial, onde dada uma nova imagem facial de teste, verifica-se a qual indivíduo essa imagem facial pertence dado um conjunto de imagens faciais de treinamento pertencentes a diversos indivíduos, sendo um problema de correspondência de um-para-muitos (JAFRI; ARABNIA, 2009).

Um problema inerente ao reconhecimento facial em ambas as abordagens é a elevada dimensionalidade da representação de faces por feições faciais (pois cada *pixel* em uma imagem facial pode ser compreendido como sendo uma dimensão no espaço de feições original). Além disso, o processamento com dados de alta dimensionalidade pode implicar em ineficiência e resultados imprecisos devido não somente à grande necessidade de memória e processamento à computação de tais dados, mas também à necessidade de um grande número de amostras (imagens faciais) necessárias para tornar dados multidimensionais suficientemente amostrados, situação que pode levar à chamada *maldição da dimensionalidade* (VALIANT, 1984).

Um método de reconhecimento facial eficiente deve ser apto a processar adequadamente dados de alta dimensionalidade e deve suportar o crescimento do número de usuários a serem reconhecidos (indivíduos), bem como suportar a adição de novas imagens faciais. Felizmente existem diversas abordagens propostas na literatura para realizar a redução de dimensionalidade de dados de alta dimensionalidade de forma a preservar características discriminativas dos dados faciais, criando um espaço de baixa dimensionalidade, onde novas imagens faciais podem ser mapeadas e classificadas. Entretanto nem toda a transformação que compacta os dados é discriminativa.

Outros fatores que podem influenciar a eficiência do reconhecimento facial são a pose da cabeça, expressão facial, variações de iluminação do ambiente, variação de idade, oclusão, entre outros (NASTAR; MITSCHKE, 1998). Tais variações faciais afetam a qualidade das feições obtidas na extração de feições das imagens faciais, levando a vetores de feições ruidosos, com grande incidência de dados atípicos. É muito importante que os métodos de reconhecimento facial sejam tolerantes a tais variações nas imagens faciais, tanto no treinamento como no reconhecimento facial (SHAOGANG GONG; PSARROU, 2000).

2.1 Extração de Feições Faciais e Reconhecimento Facial

Os métodos de reconhecimento facial disponíveis na literatura podem ser divididos em duas categorias principais (JAFRI; ARABNIA, 2009); onde a primeira corresponde aos métodos que usam feições densas, também chamados de métodos baseados na aparência (*appearance-based methods*) (TURK; PENTLAND, 1991) (HE et al., 2005) (CAI et al., 2006), também denominados métodos holísticos (JAFRI; ARABNIA, 2009), onde são processados todos os *pixels* das imagens faciais, levando a uma representação densa das feições faciais que é usada no aprendizado de forma a representar a variabilidade da aparência facial a partir de um conjunto de imagens faciais de treinamento; e os métodos que usam feições esparsas, também chamados de métodos baseados em feições locais (*local-feature-based methods*), onde são avaliadas posições específicas nas imagens faciais (*landmarks*), de forma que são extraídos subconjuntos dos *pixels* das imagens faciais, posições de *landmarks*, proporções biométricas, entre outros, levando a uma representação esparsa das feições faciais.

De forma geral, um método de reconhecimento facial apresenta três estágios distintos durante seu processamento: extração de feições faciais, redução de dimensionalidade e reconhecimento facial. A extração de feições é o primeiro passo de qualquer método de reconhecimento facial e geralmente nesse estágio são processadas imagens faciais para gerar vetores de feições faciais ou, diferentemente, localizações importantes nas imagens faciais (*landmarks*) podem ser marcadas manualmente ou automaticamente para que sejam extraídas informações esparsas das imagens faciais, como cores nas vizinhanças dos *landmarks*, proporções geométricas, posições, etc.

Em muitos dos métodos presentes na literatura, são extraídos vetores de feições faciais a partir de imagens faciais em escalas de cinza (TURK; PENTLAND, 1991) (HE et al., 2005) (CAI et al., 2006) de forma que, ao final do processo de extração de feições faciais, cada imagem facial é convertida para uma forma vetorial através da concatenação de todos os *pixels* presentes nas imagens faciais. Considerando imagens faciais com tamanho $M \times N$ *pixels*, os métodos da abordagem baseada na aparência (TURK; PENTLAND, 1991) (HE et al., 2005) (CAI et al., 2006), que comumente usam imagens em escalas de cinza, obtêm vetores de feições com dimensionalidade MN , de forma que cada *pixel* da imagem original representa uma dimensão no vetor de feições resultante, como exemplificado na Figura 2.1.

Na Figura 2.1, é apresentado um exemplo de extração típica de feições de imagens faciais conforme a abordagem baseada na aparência (TURK; PENTLAND, 1991) (HE et al., 2005) (CAI et al., 2006), onde à esquerda é apresentada uma imagem facial de entrada na forma de uma matriz de *pixels* e à direita é apresentado o vetor de feições resultante extraído da imagem facial. Os quadrados coloridos representam os *pixels* que são movidos da imagem facial na forma de matriz para a imagem facial na forma vetorial na ordem indicada na figura. Consequentemente, o vetor de feições obtido concatena todos os *pixels* da imagem facial original e pode ser considerado como um ponto em um espaço MN -dimensional, que é frequentemente um espaço de alta dimensionalidade.

A representação de imagens faciais através de vetores de feições faciais tende a apresentar alta dimensionalidade, o que reduz a eficiência das feições faciais no reconhecimento facial, portanto, a fim de aumentar a qualidade das feições faciais e ao mesmo tempo compactá-las, é realizada a redução de dimensionalidade da representação de faces por feições faciais. Como as feições originais das imagens faciais não apresentam boa separabilidade de classes de dados faciais, é importante realizar a redução de dimensionalidade da representação de faces por feições faciais de forma a preservar características

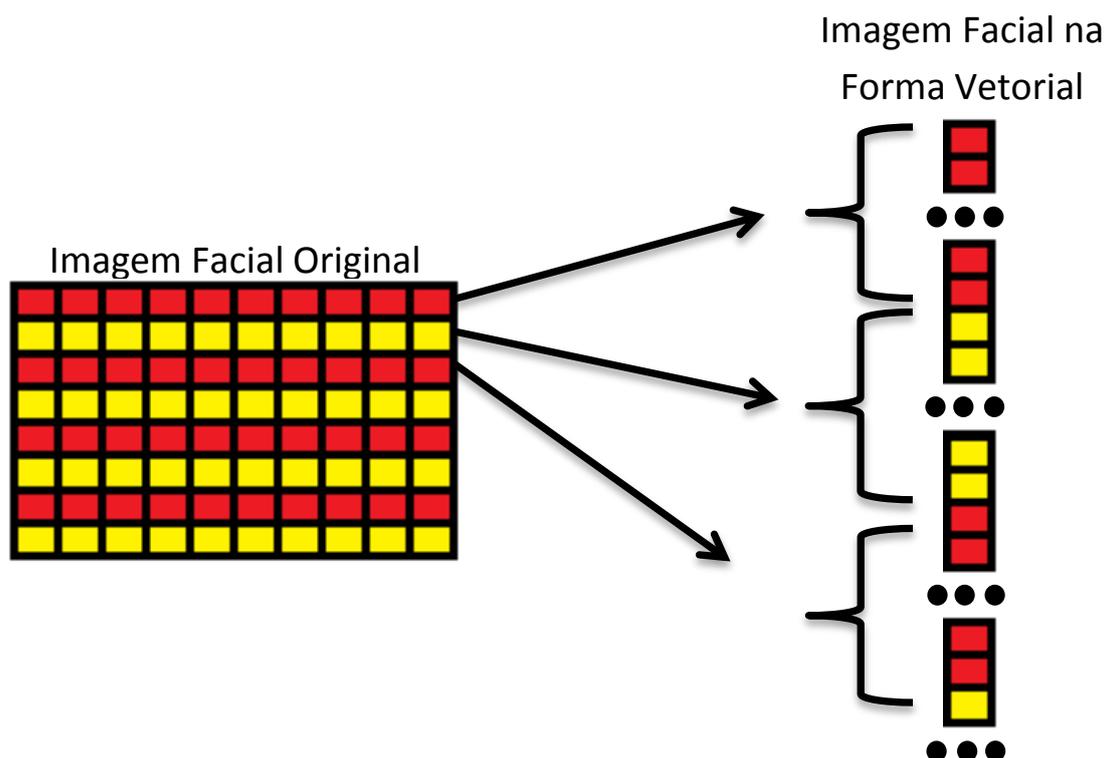


Figura 2.1: Extração Típica de Feições Densas de Imagens Faciais.

discriminativas dos dados faciais no espaço de baixa dimensionalidade (BELHUMEUR; HESPANHA; KRIEGMAN, 1997) (TURK; PENTLAND, 1991) (HE et al., 2005) (CAI et al., 2006).

Uma vez que as imagens faciais tenham sido projetadas para o espaço de baixa dimensionalidade, o reconhecimento (classificação) de novas imagens faciais ocorre nesse subespaço. A regra do vizinho mais próximo (COVER; HART, 1967) é um dos classificadores mais aplicados na classificação de novas imagens faciais (TURK; PENTLAND, 1991) (HE et al., 2005) (CAI et al., 2006), além de outros classificadores, como *Máquinas de Vetores de Suporte* (*Support Vector Machines - SVM*) (VAPNIK, 1998).

Alguns dos principais métodos da abordagem baseada na aparência são os métodos *Eigenfaces* (TURK; PENTLAND, 1991), *Laplacianfaces* (HE et al., 2005), *Projeções Ortogonais Preservadoras de Localidade* (*Orthogonal Locality Preserving Projections - OLPP*), *Análise Discriminante Linear* (*Linear Discriminant Analysis - LDA*) (BELHUMEUR; HESPANHA; KRIEGMAN, 1997), *Análise Discriminante Multi-Visões* (*Multi-view Discriminant Analysis - MvDA*) (KAN et al., 2012), entre outros.

Enquanto os métodos da abordagem baseada na aparência extraem vetores de feições de imagens faciais inteiras, como exemplificado na Figura 2.1, os métodos da abordagem baseada na aparência extraem posições de *landmarks*, proporções biométricas e também vetores de feições, concatenando os *pixels* na vizinhança dos *landmarks*, como no método ASM Eficiente (BEHAINE; SCHARCANSKI, 2012), e outros (SUKNO et al., 2007) (WU; TURAGA; CHELLAPPA, 2012).

2.2 Métodos Baseados em Feições Densas

Um dos primeiros métodos a estudar a variabilidade das imagens faciais é o trabalho de Sirovich e Kirby (SIROVICH; KIRBY, 1987), onde é introduzida a base teórica para a representação de faces por feições faciais de baixa dimensionalidade utilizada no popular método *Eigenfaces* (TURK; PENTLAND, 1991), que visa a realizar a classificação de imagens faciais nesse espaço de baixa dimensionalidade que preserva a maior parte da variabilidade das imagens faciais. Como normalmente ocorre na abordagem baseada na aparência, imagens faciais de treinamento são usadas para modelar a variabilidade das imagens faciais.

Como proposto inicialmente no método de Sirovich e Kirby (SIROVICH; KIRBY, 1987), imagens faciais são convertidas para escalas de cinza e reduzidas a um tamanho fixo¹, implicando em uma alta dimensionalidade da representação de faces por feições faciais. Entretanto, verifica-se que apenas um pequeno número de dimensões seria necessário para realizar o reconhecimento facial, portanto uma transformação que mapeia imagens faciais do espaço de alta dimensionalidade para um espaço de baixa dimensionalidade é proposta no trabalho de Sirovich e Kirby (SIROVICH; KIRBY, 1987) de forma a preservar a maior parte da variabilidade dos dados faciais.

Considerando cada imagem facial (independente de sua classe de dados faciais) como x_h e a face média (média de todas as imagens faciais de treinamento) como $x_m = \sum_{h=1}^n x_h/n$, com n imagens faciais, a matriz $X = [x_1 - x_m, x_2 - x_m, \dots, x_n - x_m]$ contém todas as imagens de treinamento (subtraídas da média). A decomposição espectral de X pelo método de Karhunen-Loeve (GONZALEZ; WOODS, 2001) permite determinar as principais componentes da distribuição das imagens faciais.

A fim de compactar dados preservando a maior parte da informação facial, são selecionados os e maiores autovalores (e seus autovetores correspondentes), criando a base de um espaço de baixa dimensionalidade, onde cada imagem facial é representada como sendo uma combinação linear dos autovetores obtidos, chamados de *Eigenpictures*, que são as "imagens padrão" para representação facial. A representação final de baixa dimensionalidade é dada pelos pesos que multiplicam cada autovetor.

A representação facial na forma de vetores de pesos pode ser muito útil no reconhecimento facial, como proposto no método *Eigenfaces* (TURK; PENTLAND, 1991), onde é proposto um *framework* que projeta representações de faces por feições faciais de alta dimensionalidade para um *espaço facial* de baixa dimensionalidade caracterizado pelas *Eigenfaces*, que são o equivalente às *Eigenpictures*. A classificação de uma nova imagem facial ocorre determinando a imagem facial do conjunto de dados de treinamento mais próxima no espaço de baixa dimensionalidade. Por outro lado, se a imagem facial de teste estiver distante de qualquer outra imagem facial de treinamento, acima de um limiar, ela é considerada como desconhecida (ou não-face).

A transformação de qualquer imagem facial para o espaço determinado pelas *Eigenfaces* ocorre pela projeção da imagem facial para o subespaço de baixa dimensionalidade, preservando grande parte da variação dos dados originais, no sentido de reconstrução dos dados. Entretanto, o método *Eigenfaces* nem sempre apresenta uma boa separabilidade entre diferentes classes de dados faciais (indivíduos) no espaço de baixa dimensionalidade.

¹No método de Sirovich e Kirby (SIROVICH; KIRBY, 1987), todas as imagens faciais são reduzidas ao tamanho fixo de 128×128 *pixels*, levando a uma representação facial original com $128^2 = 16384$ dimensões.

A fim de garantir a preservação de características discriminativas dos dados faciais no espaço de baixa dimensionalidade, foi proposto o método *Laplacianfaces* (HE et al., 2005) que usa *Projeções Preservadoras de Localidade* (*Locality Preserving Projections - LPP*) para determinar a transformação para o espaço de baixa dimensionalidade de forma a preservar a estrutura local dos dados faciais.

Como primeiro passo do método LPP, é determinado um grafo de localidade, que é dado pela matriz de similaridade S , que indica a vizinhança de cada imagem facial presente no treinamento. A similaridade S_{hs} entre as imagens faciais x_h e x_s é dada da seguinte forma:

$$S_{hs} = \begin{cases} e^{-\|x_h - x_s\|^2/t}, & \|x_h - x_s\|^2 < \varepsilon; \\ 0, & \text{caso contrário;} \end{cases}, \quad (2.1)$$

onde t é a largura de banda dos dados faciais (pode ser estimada (PARK; MARRON, 1990) ou determinada experimentalmente) e ε define o raio da localidade, de forma que x_h e x_s são vizinhos se eles se encontram a uma distância um do outro menor que ε . Adicionalmente, a matriz S pode ser definida de outra forma (HE et al., 2005), onde a condição de distância é substituída por x_h estar entre os k vizinhos mais próximos de x_s ou x_s estar entre os k vizinhos mais próximos de x_h .

$$S_{hs} = \begin{cases} e^{-\|x_h - x_s\|^2/t}, & \text{se } x_h \text{ está entre os } k \text{ vizinhos de } x_s; \\ 0, & \text{caso contrário;} \end{cases} \quad (2.2)$$

Como consequência, o cálculo da dissimilaridade adiciona uma pesada penalidade quando duas imagens faciais não estão próximas no espaço de alta dimensionalidade, fazendo com que ambas fiquem isoladas no espaço de baixa dimensionalidade. No próximo passo, é necessário determinar a base de vetores do subespaço reduzido que preserva a localidade dos dados indicada pela matriz de similaridades S , que representa o grafo de localidades. Portanto, o seguinte problema de autovalores é solucionado:

$$XLX^T \vec{v} = \lambda XDX^T \vec{v} \quad (2.3)$$

Onde $X = [x_1, x_2, \dots, x_n]$ é a matriz com todas as imagens de treinamento, D é a matriz de valência (diagonal) com $D_{hh} = \sum_s S_{hs}$ e $L = D - S$ é a matriz Laplaciana do grafo, conforme a teoria espectral dos grafos (CHUNG, 1997). A matriz D oferece uma medida natural no espaço dos dados faciais, portanto quanto maior é D_{hh} , mais importante é x_h . Como resultado da solução da Equação 2.3, a matriz de transformação $W_{LPP} = (\vec{g}_1, \vec{g}_2, \dots, \vec{g}_d)$ é obtida selecionando os d menores autovalores (e seus autovetores g_p correspondentes, com $1 \leq p \leq d$), de forma que d representa a dimensionalidade escolhida para o espaço de baixa dimensionalidade. Para projetar uma nova imagem facial x no espaço de baixa dimensionalidade, é aplicada a seguinte transformação:

$$x \rightarrow y = W^T x, \quad (2.4)$$

onde,

$$W = W_{PCA} W_{LPP}, \quad (2.5)$$

e,

$$W_{LPP} = (\vec{g}_1, \vec{g}_2, \dots, \vec{g}_d), \quad (2.6)$$

onde y é uma representação com d dimensões da imagem facial x , e W é a matriz de transformação final. PCA é usada para reduzir a influência de ruído e para compactar

dados projetando as imagens faciais x_h para um subespaço onde quase toda a variabilidade dos dados é preservada mantendo as principais componentes com maior variância, que originam a matriz $W_{PCA} = (\vec{v}_1, \vec{v}_2, \dots, \vec{v}_e)$. A Figura 2.2 ilustra um caso de projeção bidimensional do subespaço das *Laplacianfaces*, onde diferentes expressões faciais e poses (imagens faciais) pertencentes à base de imagens faciais Radboud (LANGNER et al., 2010) são mapeadas para esse subespaço. Pode-se observar que existe um suave caminho de transição entre expressões faciais e poses similares. Dessa forma, imagens faciais similares são mapeadas para localizações vizinhas nesse subespaço.

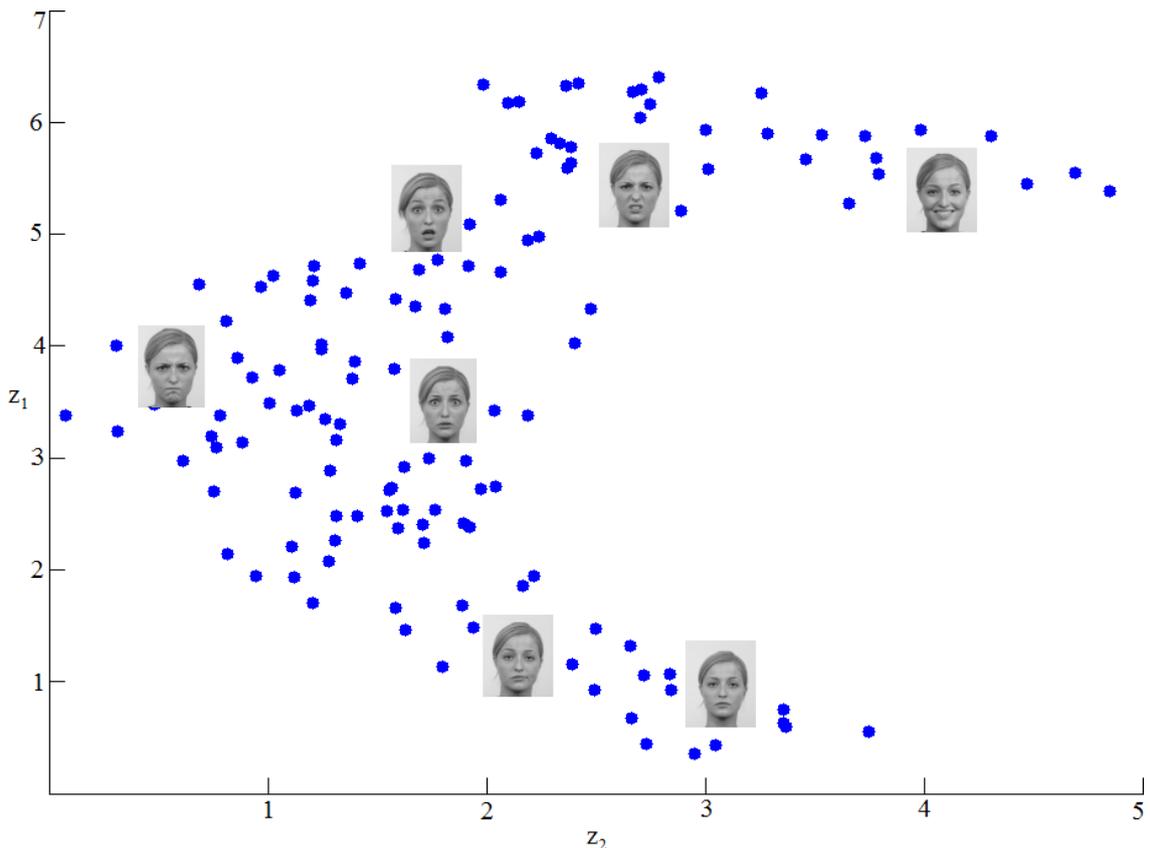


Figura 2.2: Ilustração representando uma projeção bidimensional do espaço determinado pelas *Laplacianfaces* na base de imagens faciais Radboud com eixos (dimensões) denotados por z_1 e z_2 .

O algoritmo LPP pode ser compreendido como sendo uma forma de obter uma aproximação linear ótima do operador Laplacian-Beltrami no *manifold* das faces, onde LPP preserva a estrutura local dos dados faciais (HE et al., 2005). Entretanto LPP não é ortogonal e como consequência, a métrica do subespaço das *Laplacianfaces* pode conter distorção na forma das classes de dados faciais, potencialmente prejudicando o reconhecimento facial.

O método *Projeções Ortogonais Preservadoras de Localidade (Orthogonal Locality Preserving Projections - OLPP)* (CAI et al., 2006) foi proposto para melhorar a preservação da estrutura local dos dados faciais e, conseqüentemente, incrementar a discriminação das classes de dados faciais no espaço de baixa dimensionalidade. Em outras palavras, o método OLPP adiciona a propriedade de ortogonalidade à matriz de transformação obtida pelo método LPP causando uma melhor preservação da métrica de classificação facial no espaço de baixa dimensionalidade.

Como OLPP é uma extensão do método LPP, o seu primeiro passo é aplicar a projeção PCA W_{PCA} , descartando todos os autovalores de baixa magnitude com o objetivo de preservar a maior parcela dos dados faciais no sentido de reconstrução dos dados, assim, compactando as feições faciais originais enquanto características discriminativas das mesmas são mantidas. Dessa forma, os dados faciais se tornam descorrelacionados e o número de dimensões se torna igual ao posto (*rank*) da matriz de dados. Da mesma forma que no método LPP, é computada a matriz com os pesos que medem as similaridades entre as imagens faciais de treinamento, usando a Equação 2.1 ou Equação 2.2, de forma que cada imagem facial x_h tenha apenas k vizinhos (com similaridade não-nula).

Da mesma forma que no LPP, é definida a matriz diagonal D e a matriz Laplaciana L . No método OLPP, são determinados os autovetores $\{\vec{g}_1, \vec{g}_2, \dots, \vec{g}_d\}$ que formam a base ortogonal da transformação para o subespaço de baixa dimensionalidade, que são calculados como segue: o vetor \vec{g}_1 é o autovetor associado com o menor autovalor de $(XDX^T)^{-1}X LX^T$. Para determinar os outros autovetores \vec{g}_p , com $2 \leq p \leq d$, de forma que a projeção final seja ortogonal, através de desenvolvimento algébrico (CAI et al., 2006), são definidos $G^{(p-1)}$ e $H^{(p-1)}$:

$$G^{(p-1)} = [\vec{g}_1, \dots, \vec{g}_{p-1}], \quad (2.7)$$

$$H^{(p-1)} = [G^{(p-1)}]^T (XDX^T)^{-1} G^{(p-1)}, \quad (2.8)$$

onde cada vetor \vec{g}_p da base é o autovetor de:

$$M^{(p)} = \{I - (XDX^T)^{-1} G^{(p-1)} [H^{(p-1)}]^{-1} [G^{(p-1)}]^T\} (XDX^T)^{-1} X LX^T, \quad (2.9)$$

associado com o menor autovalor de $M^{(p)}$, onde d é a dimensionalidade escolhida para o espaço de baixa dimensionalidade. Uma vez que os vetores da base ortogonal $\{\vec{g}_1, \vec{g}_2, \dots, \vec{g}_d\}$ forem obtidos, a transformação OLPP é determinada. Considerando $W_{OLPP} = [\vec{g}_1, \dots, \vec{g}_d]$, a transformação para o espaço de baixa dimensionalidade ocorre da seguinte forma:

$$x \rightarrow y = W^T x, \quad (2.10)$$

onde, $W = W_{PCA} W_{OLPP}$; y é uma representação de dimensionalidade d da imagem facial x e W é a matriz de transformação final que mapeia vetores de imagens faciais do espaço de alta dimensionalidade original para o espaço de baixa dimensionalidade, preservando a estrutura da métrica de classificação facial. Por fim, a classificação facial ocorre pelo classificador do vizinho mais próximo da mesma forma que no método LPP.

Outro método que segue a linha do OLPP é o método *Projeções Ortogonais Preservadoras de Vizinhança (Orthogonal Neighborhood Preserving Projections - ONPP)* (KOKIOPOULOU; SAAD, 2007), que tem o objetivo de criar uma projeção para um espaço de baixa dimensionalidade de forma que ocorra a preservação da geometria local intrínseca dos dados faciais e também a preservação da geometria global dos dados faciais no espaço de baixa dimensionalidade.

Primeiramente, no método ONPP, é criada uma matriz de afinidades, onde para cada imagem facial, são determinados os seus k vizinhos mais próximos de forma a definir a localidade dos dados considerando que cada imagem facial e seus k vizinhos formam uma estrutura localmente linear. Com base nisso, o método ONPP reconstrói cada imagem facial através de uma combinação linear de seus k vizinhos mais próximos, onde o erro de reconstrução \mathcal{E} de todas as imagens faciais é dado por:

$$\mathcal{E}(P) = \sum_h \left\| x_h - \sum_s p_{hs} x_s \right\|_2^2, \quad (2.11)$$

sendo que o objetivo é determinar a matriz de pesos P que minimiza o erro de reconstrução \mathcal{E} , pois cada peso p_{hs} indica a contribuição que x_s (s -ésima imagem facial de treinamento) tem na reconstrução de x_h (h -ésima imagem facial de treinamento), e se x_s não for um dos k vizinhos mais próximos de x_h , seu peso é definido como zero. Por fim, a matriz P é determinada satisfazendo a restrição: $\sum_s p_{hs} = 1$. Considerando que cada imagem facial original x_h é mapeada para y_h no espaço de baixa dimensionalidade, considera-se que os pesos p_{hs} também reconstróem y_h a partir dos mesmos k vizinhos no espaço de baixa dimensionalidade. Portanto, um novo problema de otimização é proposto, onde o objetivo é minimizar o erro de reconstrução \mathcal{Y} das imagens faciais no espaço de baixa dimensionalidade:

$$\mathcal{Y} = \sum_h \left\| y_h - \sum_s p_{hs} y_s \right\|_2^2, \quad (2.12)$$

porém, agora os pesos p_{hs} são fixos e o objetivo se torna determinar a matriz \mathcal{Y} que contém as coordenadas y_h das imagens faciais no espaço de baixa dimensionalidade que minimiza o erro de reconstrução. Para se obter \mathcal{Y} , devem-se seguir duas restrições, $\sum_h y_h = 0$ (as coordenadas dos vetores das imagens faciais devem estar centradas na origem) e $\frac{1}{n} \sum_h y_h y_h^T = I$ (as coordenadas dos vetores faciais de cada imagem facial devem ter covariância unitária), onde I indica a matriz identidade. Tal problema de otimização pode ser solucionado eficientemente utilizando-se do cálculo de autovetores.

De forma geral, os métodos apresentados anteriormente modelam a localidade das imagens faciais na forma de um grafo, onde se usam pesos maiores para os vizinhos mais similares (ou que tem maior contribuição), e pesos menores (ou nulos) para as imagens faciais que são dissimilares, assim criando uma aproximação de baixa dimensionalidade para o *manifold* das faces com o objetivo de preservar no espaço de baixa dimensionalidade a estrutura dos dados faciais (abordagem não supervisionada). Porém, uma vez que são dadas as imagens faciais de treinamento e seus rótulos (indicando a quais classes de dados faciais as imagens faciais de treinamento pertencem), essa informação pode ser aproveitada no processo de determinação do espaço de baixa dimensionalidade, ajudando a criar uma projeção que tem o objetivo de preservar a estrutura das classes de dados faciais no espaço de baixa dimensionalidade (abordagem supervisionada).

Com o objetivo de preservar melhor a estrutura das classes de dados faciais (indivíduos), foi proposto o método *Análise Discriminante Linear* (*Linear Discriminant Analysis* - LDA), que é um método popular de redução de dimensionalidade e classificação (BELHUMEUR; HESPANHA; KRIEGMAN, 1997), onde os vetores de projeção (*Fisherfaces*) são obtidos maximizando a covariância entre as classes e ao mesmo tempo minimizando a covariância interna das classes no espaço de baixa dimensionalidade, determinando a direção z que maximiza o seguinte problema de otimização:

$$\max_z \frac{z^T S_B z}{z^T S_W z}, \quad (2.13)$$

onde,

$$S_B = \sum_{a=1}^A (\mu_a - \mu_t)(\mu_a - \mu_t)^T, \quad (2.14)$$

$$S_W = \sum_{a=1}^A \left(\sum_{b=1}^B (x_{ab} - \mu_a)(x_{ab} - \mu_a)^T \right) \quad (2.15)$$

onde S_B é a matriz de dispersão entre classes (*between-class*) e S_W é a matriz de dispersão intra-classes (*within-class*); μ_t é a média global das imagens faciais; μ_a é a média das imagens faciais da classe de dados faciais a ; x_{ab} representa a b -ésima imagem facial na a -ésima classe de dados faciais; e A e B são a quantidade de classes de dados faciais (indivíduos) e a quantidade de imagens faciais por classe de dados faciais, respectivamente. Através de desenvolvimento algébrico (BELHUMEUR; HESPANHA; KRIEGMAN, 1997), o problema de otimização na Equação 2.13 pode ser resolvido na forma de um problema de autovalores, como segue:

$$S_B \vec{z}_l = \lambda_l S_W \vec{z}_l, \quad \text{para } l = 1, 2, 3, \dots, d, \quad (2.16)$$

onde d é a dimensionalidade do espaço reduzido, criando a base de autovetores $[\vec{z}_1, \vec{z}_2, \dots, \vec{z}_d]$ para o espaço de baixa dimensionalidade, onde a classificação ocorre pela regra do vizinho mais próximo.

O método LDA típico é linear e pode apresentar imprecisões para problemas não separáveis linearmente. Para lidar melhor com essa limitação, extensões de LDA que empregam funções de *kernel* têm sido propostas na literatura (MIKA et al., 1999) (BAUDAT; ANOUAR, 2000) (CAI; HE; HAN, 2011). A ideia principal de métodos baseados em *kernels* é mapear as imagens faciais de entrada para um espaço de maior dimensionalidade através de um mapeamento não-linear, onde o produto interno nesse espaço de feições pode ser computado usando uma função de *kernel* sem que seja necessário determinar o mapeamento não-linear explicitamente (SCHOLKOPF; SMOLA, 2001). O método *Análise Discriminante com Regressão Espectral com Kernels* (*Efficient Kernel Discriminant Analysis via Spectral Regression - SRKDA*) (CAI; HE; HAN, 2011) foi proposto como uma extensão do método LDA para o uso de *kernels* e generaliza outros métodos baseados em LDA (MIKA et al., 1999) (BAUDAT; ANOUAR, 2000). SRKDA (CAI; HE; HAN, 2011) usa decomposição espectral de grafos e técnicas de regularização para melhorar a eficiência computacional, sendo que o problema de otimização apresentado na Equação 2.13 é definido de uma forma alternativa:

$$\max_z \frac{z^T K R K z}{z^T K K z}, \quad (2.17)$$

onde K é a matriz com a computação da função *kernel*, onde $K_{hs} = \kappa(x_h, x_s)$, sendo $\kappa(\cdot, \cdot)$ uma função de *kernel*. Algumas das funções de *kernel* mais comuns são o *kernel* Gaussiano $\kappa(x_h, x_s) = \exp(-\|x_h - x_s\|^2 / 2\sigma^2)$ (onde σ define a largura de banda dos dados) e o *kernel* de Sigmoid $\kappa(x_h, x_s) = x_h^T x_s + \alpha_k$ (onde α_k é o parâmetro de ajuste). Por fim, R é a matriz que define a localidade dos dados, dada na seguinte forma:

$$R_{hs} = \begin{cases} 1/B & \text{se } x_h \text{ e } x_s \text{ são da mesma classe} \\ 0 & \text{caso contrário.} \end{cases} \quad (2.18)$$

O método *Análise Discriminante Multi-Visões* (*Multi-view Discriminant Analysis - MvDA*) (KAN et al., 2012) procura por uma projeção para um espaço comum combinando múltiplas visões dos dados para a realização de um reconhecimento facial robusto. Mais especificamente, MvDA simultaneamente resolve múltiplas projeções lineares otimizando um coeficiente de Rayleigh, com o objetivo de maximizar a variabilidade entre

classes e minimizar a variabilidade interna das classes no espaço comum de baixa dimensionalidade, que é obtido pelo seguinte problema de otimização:

$$\arg \min_{g_1, g_2, \dots, g_p} \text{Tr} \left(\frac{W_{MvDA}^T D_{MvDA} W_{MvDA}}{W_{MvDA}^T S_{MvDA} W_{MvDA}} \right), \quad (2.19)$$

onde $W_{MvDA} = [\vec{g}_1, \vec{g}_2, \dots, \vec{g}_p]$ representa a transformação para o espaço de baixa dimensionalidade, D_{MvDA} representa a matriz de dispersão entre classes combinando as múltiplas visões dos dados e S_{MvDA} representa a matriz de dispersão intra-classes combinando as múltiplas visões dos dados.

Todos esses métodos processam imagens faciais inteiras em baixa resolução e em escalas de cinza devido principalmente à necessidade de reduzir a complexidade computacional das operações matriciais nos mesmos métodos. Consequentemente, o reconhecimento de imagens faciais de alta resolução implica na redução do tamanho das imagens faciais. Nesse processo de redução dos dados, alguns detalhes nas imagens faciais podem ser perdidos. Além disso, nesses métodos, a classificação facial geralmente é baseada na regra do vizinho mais próximo (COVER; HART, 1967), que não é robusta a ruído e dados atípicos (por exemplo, quando ocorre mudança na iluminação ou expressão facial) (DEVROYE; GYÖRFI; LUGOSI, 1996).

Alguns pesquisadores propuseram-se a identificar espaços de baixa dimensionalidade que discriminem classes de dados faciais usando mapeamentos não-lineares, aplicando outros classificadores de imagens faciais. Por exemplo, Pham e Venkatesh (PHAM; VENKATESH, 2011) propõem mapear imagens faciais para um *manifold de Grassmann* (ABSIL; MAHONY; SEPULCHRE, 2004), e então determinar a projeção discriminativa de classes de dados faciais para um espaço de baixa dimensionalidade que maximiza o tamanho das margens das classes de dados faciais na abordagem de classificação por *Máquinas de Vetores de Suporte* (*Support Vector Machines* - SVM). Entretanto, o método proposto por Pham e Venkatesh (PHAM; VENKATESH, 2011) é caro computacionalmente e requer pequenos conjuntos de dados (ou seja, imagens em baixa resolução).

Existem outros métodos propostos na literatura que também usam SVM e tentam superar limitações no reconhecimento facial como a ocorrência de oclusões. Jia e Martinez (JIA; MARTINEZ, 2009) propõem um método para reconhecimento facial sob oclusão usando máscaras de oclusão que são incorporadas ao processo de aprendizagem de um classificador robusto. Entretanto esse método trabalha apenas nas feições faciais no nível das imagens faciais, não apresentando pós-processamento dos dados faciais, como em (BELHUMEUR; HESPANHA; KRIEGMAN, 1997) ou (CAI et al., 2006), e pode ser adversamente afetado por artefatos comuns na cena.

2.3 Métodos Baseados em Feições Esparsas

Os métodos da abordagem baseada na aparência (*appearance-based methods*) descritos anteriormente podem usar imagens faciais inteiras, obtendo uma representação facial densa. Entretanto, outros métodos trabalham com uma abordagem diferente e tentam obter representações faciais esparsas identificando um conjunto discreto de marcações (*landmarks*) nas imagens faciais. Esses métodos são comumente chamados de métodos baseados em feições locais (*local-feature-based methods*), onde os *landmarks* representam posições importantes nas imagens faciais, permitindo extrair informação de cor, posições de *landmarks* ou proporções biométricas (JAFRI; ARABNIA, 2009).

Os *landmarks* podem ser detectados usando *Modelos de Forma Ativos* (*Active Shape*

Models - ASM) (COOTES et al., 1995) de forma que um conjunto discreto de feições faciais possa ser extraído das imagens faciais. A maioria dos métodos esparsos usa ASM treinados e consideram todos os *landmarks* igualmente importantes no reconhecimento facial, como no método *Modelo de Formas Ativas com Feições Ótimas Invariantes* (*Active shape models with invariant optimal features*) (SUKNO et al., 2007). Entretanto, na prática, alguns *landmarks* podem oferecer uma capacidade de discriminação das classes de dados faciais maiores que outros *landmarks*, e alguns métodos designam mais peso para *landmarks* que apresentem melhor discriminação das classes de dados faciais. Em particular, o método ASM Eficiente (BEHAINE; SCHARCANSKI, 2012) designa diferentes pesos aos *landmarks*, estimando a capacidade de discriminação das classes de dados faciais a_c de cada *landmark* pelo uso de Informação Mútua (AMI):

$$a_c = \frac{1}{Z} \sum_{o=1}^Z \frac{1}{\sum_{c_1=1}^C AMI(\vec{\mu}_{F_{o,c_1}}, \vec{\mu}_{F_{o,c}})}, \quad (2.20)$$

onde c é o índice de cada *landmark* pertencente a um *Modelo de Distribuição de Pontos* (*Point Distribution Model* - PDM) com C *landmarks*, onde Z indica a quantidade de canais de cor disponíveis nas imagens faciais e $\vec{\mu}_{F_{o,c}}$ representa um vetor com as médias de intensidade de cor do o -ésimo canal de cor apresentadas na vizinhança do c -ésimo ponto no PDM considerando todas as imagens de treinamento. O vetor $\vec{\mu}_{F_{o,c}}$ contém um elemento para cada classe de dados faciais presente no conjunto de dados de treinamento. Por fim, AMI representa o cálculo da informação mútua entre dois conjuntos (VINH; EPPS; BAILEY, 2010). Para representar feições, cada indivíduo presente no treinamento é representado por uma mistura de Gaussianas, onde cada mistura é dada por:

$$\frac{1}{\sum_{c_1=1}^C a_{c_1}} \sum_{c=1}^C \frac{a_c}{\sqrt{(2\pi\sigma_c^2)}} e^{-\frac{(x-\mu_c)^2}{2\sigma_c^2}}, \quad (2.21)$$

onde σ_c e μ_c representam a média e o desvio padrão na vizinhança de cada *landmark* c compreendendo todas as imagens de treinamento de um determinado indivíduo. É usada uma abordagem probabilística para realizar a classificação facial. O método ASM Eficiente (BEHAINE; SCHARCANSKI, 2012) tenta tornar ASMs mais confiáveis, mas requer um posicionamento mais sofisticado de *landmarks* e pode sofrer de algumas deficiências (especialmente quando os dados faciais não apresentam distribuição modal).

Mapeamentos não-lineares de representações faciais esparsas também têm sido propostas para aumentar a discriminação das classes de dados faciais. Por exemplo, o método *Estimação de Idade e Verificação Facial sob Envelhecimento Usando Landmarks* (*Age estimation and face verification across aging using landmarks*) (WU; TURAGA; CHELLAPPA, 2012) representa imagens faciais através de um conjunto de posições de *landmarks*, obtendo um mapeamento não-linear de cada imagem facial para um ponto em um *manifold de Grassmann*, sendo que a fase de reconhecimento ocorre no espaço tangente do *manifold* usando *Máquinas de Vetores de Suporte* (*Support Vector Machines* - SVM) (VAPNIK, 1998) e *Máquinas de Vetores Relevantes* (*Relevant Vector Machines* - RVM) (TIPPING, 2001). Desafortunadamente, esse método usa as posições dos *landmarks* como feições e pode ser adversamente afetado por artefatos, como mudanças na pose da cabeça ou iluminação.

Foram apresentados diversos métodos capazes de realizar reconhecimento facial, porém a maioria deles pode sofrer penalidades na extração de feições, especialmente no caso da presença de dados ruidosos e dados atípicos. Além disso, é necessário mapear feições

faciais de alta dimensionalidade para um espaço discriminativo de baixa dimensionalidade, pois o espaço original de feições (imagens) geralmente não apresenta boa discriminação das classes de dados faciais, entretanto muitos métodos trabalham no espaço de feições originais de alta dimensionalidade. E por fim, muitos dos métodos apresentados utilizam a regra do vizinho mais próximo para realizar a classificação que não é robusta a dados atípicos e ruído.

A seguir, na Figura 2.3, é apresentada uma taxonomia com os principais métodos de reconhecimento facial da abordagem de reconhecimento facial baseada na aparência e da abordagem baseada em feições locais, apresentados nesse trabalho.

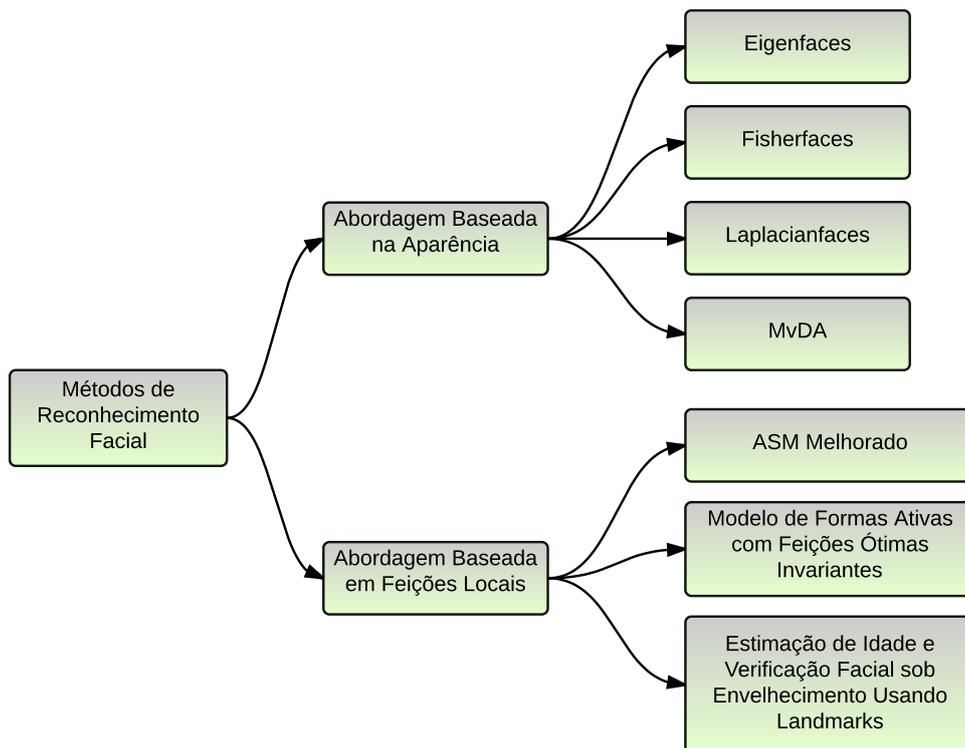


Figura 2.3: Taxonomia de métodos em reconhecimento facial.

3 MÉTODO DE RECONHECIMENTO FACIAL PROPOSTO

Conforme estado da arte, os métodos de reconhecimento facial disponíveis na literatura podem ser divididos em duas categorias principais: densos ou esparsos (JAFRI; ARABNIA, 2009), correspondendo a abordagens distintas de reconhecimento facial, onde são extraídas feições de todos os *pixels* das imagens faciais (TURK; PENTLAND, 1991) (HE et al., 2005) (CAI et al., 2006), ou extraídas feições de subconjuntos dos *pixels* das imagens faciais, respectivamente (SUKNO et al., 2007) (BEHAINE; SCHARCANSKI, 2012).

Os métodos que utilizam feições com representação densa (como o método OLPP típico (CAI et al., 2006)) avaliam a aparência das imagens faciais, sendo que a mesma é usada no aprendizado para representar a variabilidade da aparência facial a partir de um conjunto de imagens faciais de treinamento, que são geralmente imagens em baixa resolução em escalas de cinza. No método OLPP típico, as feições faciais são obtidas na forma de vetores de feições através da concatenação de todos os *pixels* das imagens faciais. Entretanto, tais vetores de feições faciais têm uma tendência a não apresentar boa separabilidade das classes de dados faciais (indivíduos), portanto, no método OLPP típico, é determinada uma projeção para um subespaço de baixa dimensionalidade que preserva a estrutura dos dados faciais. Além disso, novas imagens faciais de teste são classificadas nesse subespaço usando o classificador do vizinho mais próximo (COVER; HART, 1967) tipicamente usado na abordagem baseada na aparência (TURK; PENTLAND, 1991) (HE et al., 2005).

3.1 Método Proposto

Nesse trabalho, são propostas diversas melhorias em relação ao método OLPP típico (CAI et al., 2006) que processa imagens faciais em escalas de cinza, sendo que é proposto processar tanto representações coloridas esparsas de imagens faciais (Seção 3.2) como representações coloridas densas (Seção 3.3), ao contrário da abordagem baseada na aparência que processa imagens em escalas de cinza. Além disso, é proposta uma modificação do método OLPP de forma a definir novos esquemas de definição de localidade supervisionados com o objetivo de preservar a estrutura das classes de dados faciais no espaço de baixa dimensionalidade (Seções 3.4 e 3.5) ao contrário do método OLPP típico que não é supervisionado, pois preserva a estrutura dos dados faciais. Por fim, é utilizado um esquema de classificação diferente que emprega *Máquinas de Vetores de Suporte* (*Support Vector Machines* - SVM) com margens suaves (Seção 3.6) que é robusto a dados ruidosos e dados atípicos, de forma a superar o classificador do vizinho mais próximo usado no método OLPP típico, que é suscetível a dados ruidosos e atípicos.

Uma descrição esquemática do processo de treinamento e teste do método proposto é

ilustrada na Figura 3.1, onde (a) indica o fluxo de treinamento e (b) indica o fluxo de teste. Em (a), inicialmente, feições são extraídas de todas as imagens de treinamento, obtendo-se uma representação esparsa colorida (conforme proposto na Seção 3.2) ou densa (em escalas de cinza (CAI et al., 2006) ou colorida como proposto na Seção 3.3). Como primeiro passo da determinação da projeção para o espaço de baixa dimensionalidade, é definido um grafo de localidade (Seção 3.4) aplicando o esquema de localidade escolhido (OLPP1, OLPP2, OLPP3 ou OLPP4, definidos na Seção 3.4, correspondendo a formas diferentes de se definir localidades), o parâmetro de largura de banda t e o número de vizinhos k (segundo o método OLPP típico). A transformação PCA é determinada pelo parâmetro de reconstrução r , e a transformação final W para o subespaço d -dimensional é obtida (Equação 3.1 ou Equação 3.2) definindo $d \subset [1, d_{max}]$ (Seção 3.5). No último passo, o classificador SVM com margens suaves é treinado usando o parâmetro C_{SVM} (Seção 3.6).

Na Figura 3.1, é apresentado também o fluxo de teste, onde em (b) é realizada a classificação de novas imagens faciais de teste, iniciando com a extração das feições das imagens faciais, gerando vetores de feições que são projetados para o espaço discriminativo de baixa dimensionalidade, e por fim, o classificador SVM com margens suaves é aplicado para realizar o reconhecimento facial de novas imagens faciais nesse espaço.

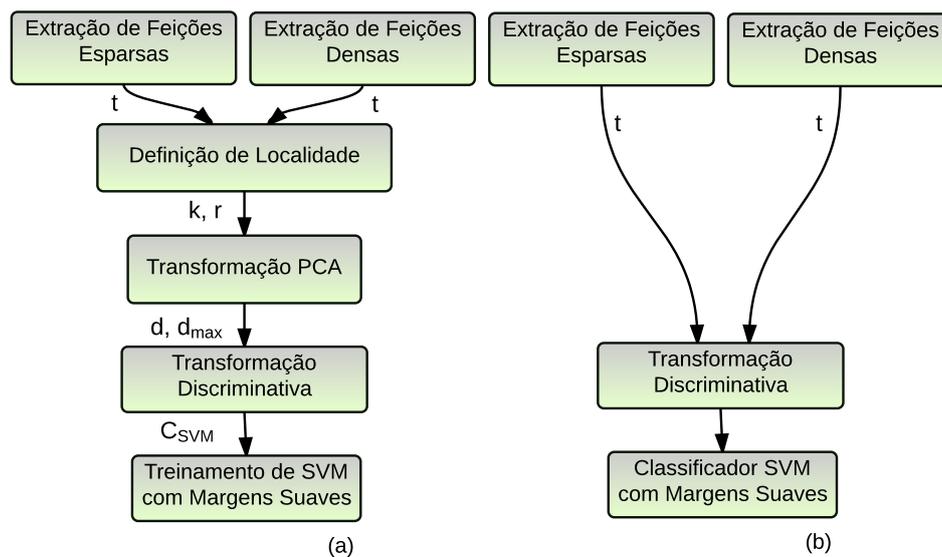


Figura 3.1: Esquema com os fluxos de treinamento e teste do método proposto.

3.2 Extração de Feições Faciais Esparsas Coloridas

Em muitos métodos da abordagem baseada na aparência, imagens faciais em escalas de cinza em baixa resolução são usadas durante o treinamento para determinar uma transformação para um espaço de baixa dimensionalidade que preserve a estrutura dos dados faciais. Porém, no processo de classificação de uma nova imagem facial, podem ocorrer dados atípicos se a aparência do indivíduo não for bem representada no conjunto de dados de treinamento (por exemplo, se o indivíduo aparece com uma diferente expressão facial ou pose da cabeça). A fim de superar problemas comuns das feições faciais, é proposto nesse trabalho um método de extração de feições faciais esparsas coloridas de imagens faciais coloridas em alta resolução que tem foco em extrair feições de posições

importantes nas imagens faciais (*landmarks*) causando um aumento da discriminação das classes de dados faciais (indivíduos) no espaço de baixa dimensionalidade em comparação com o método OLPP típico (CAI et al., 2006). Desta forma, *pixels* redundantes das imagens faciais são descartados, enquanto são preservados detalhes faciais relevantes tais como olhos e sobrancelhas, facilitando a representação de poses e expressões faciais diversas (SOLDERA; BEHAINE; SCHARCANSKI, 2015), além de preservar a cor na extração de feições faciais, pois a cor é relevante no reconhecimento facial (BEHAINE; SCHARCANSKI, 2012) (SOLDERA; BEHAINE; SCHARCANSKI, 2015) (KARIMI; DEVROYE, 2007).

No método proposto de extração de feições faciais esparsas, os vetores de feições faciais são obtidos concatenando informação de cor na vizinhança de cada posição biométrica de imagens faciais coloridas em alta resolução. Portanto, essa representação vetorial não precisa de um novo treinamento para se adaptar à maioria das poses e expressões faciais uma vez que essa amostragem seletiva é tolerante a imprecisões nas posições dos *landmarks*.

Considerando P *landmarks* possíveis em imagens faciais, geralmente não é necessário utilizar todos esses *landmarks* uma vez que eles frequentemente retêm informação redundante. Portanto, é usado um subconjunto dos *landmarks* que levam a uma melhor discriminação de classes de dados faciais (indivíduos), ou seja, é escolhida a melhor topologia de *landmarks*. Encontrar a melhor topologia de *landmarks* é computacionalmente caro, uma vez que o número de combinações de Q *landmarks* de um conjunto de P *landmarks* leva a um total de $E = (P!/(P - Q)!)$ possíveis topologias. Mesmo se for computada a capacidade de discriminação de classes de dados faciais de cada possível topologia de *landmarks* usando dados de treinamento, adicionar novas imagens faciais pode requerer recomputar a topologia escolhida, o que não é prático.

Uma solução viável é encontrar uma topologia de *landmarks* que seja adequada (que preserve informação biométrica importante e discriminativa das classes de dados faciais) através da seleção de *landmarks* em imagens faciais em posições relevantes, tais como boca, olhos, nariz, sobrancelhas, entre outras posições faciais. A topologia apresentada na Figura 3.2 foi obtida através de experimentos nos quais foram avaliadas várias topologias que marcam informação discriminativa nas imagens faciais, e essa topologia apresentou melhores resultados experimentalmente, sendo utilizada para extração de feições esparsas nessa tese. Além disso, é proposto nesse trabalho introduzir *landmarks* adicionais, que são igualmente espaçados e interpolados entre cada par de *landmarks* adjacentes e relevantes como ilustrado na Figura 3.2, onde exemplos de imagens faciais da base de imagens faciais FEI são apresentados em (a) e (b) e exemplos de imagens faciais da base de imagens faciais PUT são apresentados em (c) e (d), com os *landmarks* manualmente marcados. As bases de imagens faciais usadas nesse trabalho são apresentadas na Seção 4.1. Os *landmarks* originais (marcados) são exibidos em vermelho e os interpolados são exibidos em azul. A topologia apresentada na Figura 3.2 é também utilizada no Capítulo 4 para coletar importante informação biométrica nos experimentos com imagens faciais.

Através dos *landmarks* interpolados, a representação facial se torna mais robusta a oclusões, mudanças na pose da cabeça e nas condições de iluminação, melhorando a qualidade das feições extraídas, levando a uma topologia de *landmarks* final com $C > Q$ *landmarks*, onde $C - Q$ é o número de *landmarks* interpolados entre *landmarks* relevantes adjacentes. Foi determinado através de experimentos que a interpolação de *landmarks* melhora a discriminação das classes de dados faciais (indivíduos), melhorando a quali-

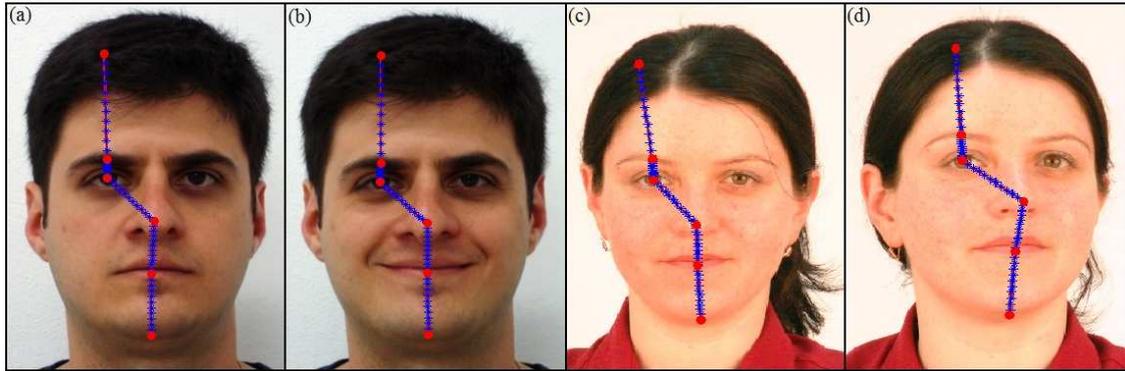


Figura 3.2: Topologia de *landmarks* adotada nesse trabalho exemplificada em duas bases de imagens.

dade das feições faciais obtidas (SOLDERA; BEHAINE; SCHARCANSKI, 2015).

De forma geral, imagens faciais coloridas em alta resolução contêm muitas posições (*landmarks*) em que potencialmente se pode extrair informação biométrica. Por outro lado, usar uma topologia de *landmarks* possibilita que seja usado apenas um subconjunto dessa vasta informação facial, implicando em uma maior eficiência na representação de feições faciais e na classificação facial, pois apenas posições importantes e discriminativas são selecionadas, como olhos, nariz e sobrancelhas, descartando informação facial imprecisa e ambígua que existe nas imagens faciais.

Os métodos baseados na aparência geralmente reduzem imagens faciais em escalas de cinza para reduzir a complexidade computacional das operações matriciais gerando *pixels* redundantes ou perdendo informação relevante na extração de feições. Por outro lado, o método proposto seleciona posições faciais relevantes (ou seja, *landmarks* relevantes selecionados e suas vizinhanças) para extrair vetores de feições, e como a topologia de *landmarks* proposta inclui *landmarks* interpolados entre cada par de *landmarks* adjacentes relevantes, detalhes discriminativos das classes de dados faciais são melhores preservados. Os *landmarks* podem ser posicionados manualmente, ou podem ser automaticamente detectados utilizando várias abordagens (EFRATY et al., 2011) (ZHAO et al., 2011) (RAPP et al., 2011) e (BURGOS-ARTIZZU; PERONA; DOLLAR, 2013).

A detecção automática de *landmarks* adiciona erros às posições obtidas dos *landmarks*, geralmente no intervalo de 1% a 9% de acordo com o método de detecção de *landmarks* e com a base de imagens faciais usados (BURGOS-ARTIZZU; PERONA; DOLLAR, 2013). Atualmente, há diversos métodos automáticos de detecção de *landmarks* na literatura, constituindo uma área muito ativa de pesquisa (EFRATY et al., 2011) (ZHAO et al., 2011) (RAPP et al., 2011) (BURGOS-ARTIZZU; PERONA; DOLLAR, 2013), entretanto o foco do trabalho proposto é realizar reconhecimento facial usando seleção de feições seguida de aplicação de redução de dimensionalidade e classificação facial, aplicado a representações de imagens densas ou esparsas, portanto, nesse trabalho, é considerado que as feições faciais (*landmarks*) já são fornecidas para representações faciais esparsas, ou seja, considera-se que as marcações já são dadas juntamente com as imagens faciais, sendo manualmente marcadas.

No método proposto de extração de feições faciais esparsas, cada imagem facial é representada por um vetor de tamanho fixo obtido pela concatenação das feições capturadas nos *pixels* internos às janelas de tamanho $V \times V$ centradas em cada *landmark* relevante ou interpolado na imagem facial, considerando que uma topologia de *landmarks* contém

C landmarks. Geralmente imagens em escalas de cinza (luminância) são usadas na abordagem de reconhecimento facial baseada na aparência, entretanto, no método proposto imagens coloridas são usadas para melhorar a discriminação das classes de dados faciais (BEHAINE; SCHARCANSKI, 2012) (SOLDERA; BEHAINE; SCHARCANSKI, 2015) (KOKIOPOULOU; SAAD, 2009).

O vetor extraído da b -ésima imagem facial pertencente à a -ésima classe de dados faciais é denotado por $x_{ab} = [x_{ab}^R | x_{ab}^G | x_{ab}^B]$, onde R, G, B são os canais de cor, '|' é usado para indicar concatenação, os vetores $x_{ab}^R = [x_{ab1}^R | x_{ab2}^R | \dots | x_{abC}^R]$, $x_{ab}^G = [x_{ab1}^G | x_{ab2}^G | \dots | x_{abC}^G]$ e $x_{ab}^B = [x_{ab1}^B | x_{ab2}^B | \dots | x_{abC}^B]$ representam os canais de cor, e x_{abc}^f , $f \in \{R, G, B\}$, é uma janela vetorizada (em cores) com tamanho V^2 centrada em cada *landmark* c . Consequentemente, x_{ab} é um vetor com $3CV^2$ dimensões.

3.3 Extração de Feições Faciais Densas Coloridas

Como comumente realizado nos métodos da abordagem baseada na aparência (TURK; PENTLAND, 1991) (HE et al., 2005) (CAI et al., 2006), as imagens faciais são representadas de forma densa por vetores de tamanho fixo obtidos pela concatenação de todos os *pixels* das imagens faciais. Geralmente imagens em escalas de cinza (luminância) em baixa resolução são usadas na abordagem baseada na aparência (TURK; PENTLAND, 1991) (HE et al., 2005) (CAI et al., 2006). Por outro lado, como a informação de cor é relevante no reconhecimento facial (BEHAINE; SCHARCANSKI, 2012) (SOLDERA; BEHAINE; SCHARCANSKI, 2015) (KARIMI; DEVROYE, 2007), é proposto nessa tese um método de extração de feições faciais densas coloridas que preserva a informação de cor na extração de feições faciais densas para melhorar a discriminação das classes de dados faciais (indivíduos).

No método proposto de extração de feições faciais densas coloridas, o vetor de feições extraído da b -ésima imagem facial pertencente à a -ésima classe de dados faciais (indivíduo) é denotado por $x_{ab} = [x_{ab}^R | x_{ab}^G | x_{ab}^B]$, onde R, G, B são os canais de cor e '|' indica concatenação. Cada vetor x_{ab}^R , x_{ab}^G e x_{ab}^B indica um canal de cor vetorizado da b -ésima imagem facial pertencente à a -ésima classe de dados faciais, que apresenta tamanho MN , onde M e N são o comprimento e largura das imagens faciais, respectivamente. Consequentemente, x_{ab} é um vetor com $3MN$ dimensões.

3.4 Definição de Vizinhança dos Dados Faciais

Tanto o esquema proposto de extração de feições faciais esparsas coloridas apresentado na Seção 3.2, como o esquema proposto de extração de feições faciais densas coloridas apresentado na Seção 3.3 levam à obtenção de vetores de feições faciais de alta dimensionalidade. Além disso, vetores de feições de alta dimensionalidade são também esperados no caso da utilização de feições típicas densas em escalas de cinza, como na abordagem de reconhecimento facial baseada na aparência, que converte imagens faciais em escalas de cinza para a forma vetorial (TURK; PENTLAND, 1991) (HE et al., 2005) (CAI et al., 2006).

Independentemente do método de extração de feições faciais empregado, a representação de feições faciais geralmente apresenta alta dimensionalidade, e por isso as feições faciais têm uma tendência a sofrerem de algumas imprecisões, como a dificuldade de se amostrar suficientemente dados faciais de alta dimensionalidade, pois amostras faciais (imagens faciais) são geralmente escassas e ao mesmo tempo apresentam alta dimen-

onalidade. Essa condição comumente chamada de *Maldição da Dimensionalidade* (VALIANT, 1984) pode levar a resultados imprecisos ou incorretos quando não é tratada devidamente.

Uma técnica amplamente utilizada em reconhecimento facial para aumentar a qualidade das feições faciais extraídas das imagens faciais e ao mesmo tempo torná-las mais compactas é determinar um subespaço de baixa dimensionalidade que preserve características discriminativas das feições faciais originais de alta dimensionalidade. Portanto as imagens faciais são mapeadas para esse subespaço de baixa dimensionalidade onde é realizada a classificação facial de novas imagens faciais. Um método que realiza a redução de dimensionalidade dos dados faciais é o OLPP típico (CAI et al., 2006), que cria um grafo de localidade que tem a função de descrever a geometria local dos dados faciais que é preservada no subespaço de baixa dimensionalidade obtido através de uma transformação linear.

Considerando A classes de dados faciais (indivíduos) disponíveis nos dados de treinamento e B imagens faciais de treinamento por classe de dados faciais, e considerando x_{ab} como a b -ésima imagem facial pertencente à a -ésima classe de dados faciais, para definir a localidade dos dados, o método OLPP típico (CAI et al., 2006) constrói um grafo de localidade que é descrito por uma matriz de afinidades (similaridades) S de forma que duas imagens faciais x_{ab} e $x_{a'b'}$ são vizinhas se $x_{a'b'}$ está entre as k imagens faciais mais próximas de x_{ab} (e vice-versa). Esses k vizinhos são as imagens faciais $x_{a'b'}$ com maior afinidade à x_{ab} .

No método OLPP típico (CAI et al., 2006), a afinidade $S_{ab}^{a'b'}$ entre imagens faciais x_{ab} e $x_{a'b'}$ é obtida usando o kernel do calor:

$$S_{ab}^{a'b'} = \begin{cases} e^{-\|x_{ab}-x_{a'b'}\|^2/t}, & \text{se } x_{ab} \text{ está entre os } k \text{ vizinhos de } x_{a'b'}; \\ 0, & \text{caso contrário;} \end{cases} \quad (3.1)$$

onde t é a largura de banda dos dados faciais (pode ser estimada (PARK; MARRON, 1990) ou determinada experimentalmente). A exponencial na Equação 3.1 afasta no espaço de baixa dimensionalidade imagens faciais que não estão próximas no espaço de feições de entrada enquanto aproxima imagens faciais similares. Nesse trabalho, não são consideradas situações em que as imagens faciais foram obtidas com falhas severas, como por exemplo, uma imagem toda preta, o que poderia levar a uma falha de processamento do método proposto, ficando essa verificação a cargo de trabalhos futuros.

O esquema de definição de localidade apresentado acima usado pelo método OLPP típico não é supervisionado, ou seja, os k vizinhos mais próximos $x_{a'b'}$ de cada imagem facial x_{ab} são escolhidos entre quaisquer classes de dados faciais a' pertencentes ao conjunto de dados de treinamento e a afinidade $S_{ab}^{a'b'}$ de x_{ab} com cada imagem facial $x_{a'b'}$ é calculada usando a Equação 3.1 (CAI et al., 2006), o que leva a uma preservação da estrutura local dos dados faciais. Por conveniência, denomina-se esse esquema de localidade não supervisionado como OLPP1 (na criação do grafo S , não é considerada as classes que as imagens faciais pertencem).

Entretanto, como normalmente ocorre em aplicações de reconhecimento facial, para cada classe de dados faciais conhecida (indivíduo), um conjunto de imagens faciais de treinamento são fornecidas. Portanto, a informação de classe disponível nos dados de treinamento pode ser usada em uma abordagem de reconhecimento facial supervisionada com o objetivo de preservar a estrutura das classes no espaço de baixa dimensionalidade (BELHUMEUR; HESPANHA; KRIEGMAN, 1997) (KOKIOPOULOU; SAAD, 2009). Como a abordagem supervisionada tende a separar melhor as classes faciais (BE-

LHUMEUR; HESPANHA; KRIEGMAN, 1997) (KOKIOPOULOU; SAAD, 2009), é proposto um novo esquema de definição de localidade, denominado OLPP2, que é quando se restringem os vizinhos $x_{a'b'}$ de x_{ab} a aqueles que pertencem à mesma classe facial que x_{ab} , ou seja, $a = a'$ (SOLDERA; BEHAINE; SCHARCANSKI, 2015).

Entretanto, o kernel do calor na Equação 3.1 usado para calcular afinidades nos esquemas de definição de localidade OLPP1 e OLPP2 pode não levar a uma boa preservação da estrutura das classes no espaço de baixa dimensionalidade, pois imagens faciais que pertencem à mesma classe de dados faciais podem ser mapeadas distantes no espaço de baixa dimensionalidade quando há baixa afinidade $S_{ab}^{a'b'}$ entre as mesmas, o que corresponde ao caso de imagens faciais de mesma classe que estejam em diferentes poses da cabeça ou expressões faciais.

Por outro lado, compactar as classes faciais no espaço de baixa dimensionalidade tende a melhor preservar a métrica de classificação uma vez que as fronteiras das classes são melhor definidas. Portanto, é proposto definir afinidades entre imagens faciais alternativamente através da definição de afinidades iguais para imagens faciais consideradas vizinhas. Portanto a matriz S é obtida usando o seguinte cálculo de afinidade:

$$S_{ab}^{a'b'} = \begin{cases} 1, & \text{se } x_{ab} \text{ está entre os } k \text{ vizinhos de } x_{a'b'}; \\ 0, & \text{caso contrário;} \end{cases} \quad (3.2)$$

ainda considerando os k vizinhos de acordo com o critério da distância Euclidiana. Além disso, são propostos dois novos esquemas de definição de localidade, OLPP3 (não supervisionado) e OLPP4 (supervisionado). Em ambos os esquemas, as afinidades $S_{ab}^{a'b'}$ entre pares de imagens faciais x_{ab} e $x_{a'b'}$ são obtidas usando Eq. 3.2, porém OLPP4 considera a informação de classe das imagens faciais de treinamento, enquanto OLPP3 não utiliza essa informação.

3.5 Redução de Dimensionalidade

É determinada nesse trabalho uma projeção para um espaço discriminativo de baixa dimensionalidade através da aplicação da modificação proposta do método *Projeções Ortogonais Preservadoras de Localidade (Orthogonal Locality Preserving Projections - OLPP)* (CAI et al., 2006)) de forma que quatro esquemas distintos de definição de localidade podem ser usados, OLPP1 (conforme o método OLPP (CAI et al., 2006)) e os esquemas propostos OLPP2 (SOLDERA; BEHAINE; SCHARCANSKI, 2015), OLPP3 e OLPP4, apresentados na Seção 3.4. Como resultado da aplicação de qualquer um dos esquemas de definição de localidade OLPP1, OLPP2, OLPP3 ou OLPP4, a matriz de afinidades S obtida descreve a geometria de dados a ser preservada no espaço de baixa dimensionalidade (SOLDERA; BEHAINE; SCHARCANSKI, 2015).

Considerando a matriz de dados $X = [x_{11}, \dots, x_{1B}; \dots; x_{A1}, \dots, x_{AB}]$ obtida com todas as imagens faciais de treinamento, com tamanho $3CV^2 \times AB$ (quando usadas feições esparsas coloridas conforme Seção 3.2), $3MN \times AB$ (quando usadas feições densas coloridas conforme Seção 3.3) ou $MN \times AB$ (quando usadas feições densas em escalas de cinza (CAI et al., 2006)), o método OLPP convencional (CAI et al., 2006) obtém a projeção que mapeia as imagens faciais de todas as classes de dados faciais para um subespaço de baixa dimensionalidade. Até o restante dessa Seção, será apresentado um método de projeção de subespaço similar ao OLPP (CAI et al., 2006), entretanto é importante deixar claro que essa discussão é válida para OLPP1, OLPP2, OLPP3 e OLPP4, que são os esquemas de localidade apresentados nesse trabalho.

Para que sejam obtidas as bases ortogonais da transformação para o subespaço de baixa dimensionalidade, a matriz diagonal D (com elementos $D_{ab}^{ab} = \sum_{a',b'} S_{ab}^{a'b'}$) e a matriz Laplaciana $L = D - S$ são computadas (CHUNG, 1997). A transformação para o subespaço com a dimensionalidade escolhida d é obtida através do mapeamento $x_{ab} \rightarrow y_{ab}$, que é representado por:

$$x_{ab} \rightarrow y_{ab} = W^T x_{ab}, \quad (3.3)$$

$$W = W_{PCA} W_{AOLPP}, \quad (3.4)$$

onde $W_{PCA} = [\vec{v}_1, \dots, \vec{v}_e]$ é obtido projetando cada representação de imagem facial x_{ab} no subespaço PCA e descartando as componentes principais com menores autovalores de forma a manter somente uma porcentagem r das componentes originais nos dados projetados y_{ab} (ou seja, no subespaço W_{PCA}). Os autovetores $\vec{v}_1, \vec{v}_2, \dots, \vec{v}_e$ são aqueles com os maiores autovalores, e são obtidos resolvendo o problema de autovalores $XX^T \vec{v} = \lambda \vec{v}$ usando *Decomposição em Valores Singulares (Singular Value Decomposition - SVD)*, decompondo a matriz de dados X em $[U_{SVD}, S_{SVD}, V_{SVD}]$. Considera-se d_{max} como o número máximo de autovetores de forma que $\sum_{\beta=1}^{d_{max}} S_{SVD}(\beta, \beta) \leq r \sum_{\beta=1}^{\min(AB, \#features)} S_{SVD}(\beta, \beta)$, com $d_{max} \subset [1, \min(AB, \#features)]$, onde $\#features$ representa o número de feições independente da representação de feições aplicada.

O parâmetro r é arbitrariamente escolhido no intervalo $[0, 1]$ e o erro de reconstrução obtido de X com um conjunto reduzido de d_{max} autovetores (componentes principais) é minimizado de acordo com os mínimos quadrados (TURK; PENTLAND, 1991). A dimensionalidade desejada para o subespaço de baixa dimensionalidade é escolhida como $d \subset [1, d_{max}]$. Diferentes valores para r e d levam a diferentes taxas de reconhecimento faciais. Isso ocorre por que para cada valor de r e d , projeções de subespaço sensivelmente diferentes são determinadas. Deve ser observado que as acurácias mais altas não são necessariamente obtidas pelo uso dos valores de parâmetros mais altos uma vez que a taxa de reconhecimento depende da separabilidade de classes de dados faciais no subespaço obtido. Portanto, diferentes intervalos de valores de r e d são avaliados repetidamente, baseados nos dados disponíveis, e os valores de r e o d que levam às maiores taxas de reconhecimento são determinados com base nesses experimentos que são gerenciados manualmente, o que é uma tarefa custosa em termos de tempo. Mais detalhes da escolha desses parâmetros serão discutidos no Capítulo 4.

Definido um valor para d , a dimensionalidade do subespaço onde as imagens faciais projetadas y_{ab} residem, é necessário determinar a base ortogonal que define esse subespaço. Considerando $\{\vec{g}_1, \vec{g}_2, \dots, \vec{g}_d\}$ como sendo os vetores da base ortogonal, é definido:

$$G^{(p-1)} = [\vec{g}_1, \dots, \vec{g}_{p-1}], \quad (3.5)$$

$$H^{(p-1)} = [G^{(p-1)}]^T (XDX^T)^{-1} G^{(p-1)}. \quad (3.6)$$

Os vetores da base ortogonal $\{\vec{g}_1, \vec{g}_2, \dots, \vec{g}_d\}$ são computados como segue: o vetor da base \vec{g}_1 é o autovetor associado com o menor autovalor de $(XDX^T)^{-1}XDX^T$, e \vec{g}_p , com $2 \leq p \leq d$, é o autovetor de:

$$M^{(p)} = \left\{ I - (XDX^T)^{-1} G^{(p-1)} [H^{(p-1)}]^{-1} [G^{(p-1)}]^T \right\} \cdot (XDX^T)^{-1} XDX^T, \quad (3.7)$$

associado com o menor autovalor de $M^{(p)}$. O espaço de baixa dimensionalidade pretendido é definido por $W_{AOLPP} = [\vec{g}_1, \dots, \vec{g}_d]$, onde d é o número escolhido de dimensões, com $2 \leq p \leq d$. Como resultado da projeção descrita na Equação 3.3, y_{ab} é uma representação d -dimensional da imagem facial original x_{ab} .

De acordo com o método OLPP convencional (CAI et al., 2006), uma nova imagem facial x_{uv} com pose da cabeça v e classe de dados faciais desconhecida u é classificada projetando x_{uv} no espaço de baixa dimensionalidade usando a Equação 3.3 para obter y_{uv} , e então a regra de classificação pelo vizinho mais próximo é aplicada para encontrar a imagem facial mais próxima a y_{uv} nesse espaço de baixa dimensionalidade. Considerando que a imagem facial mais próxima é y_{ab} , então y_{uv} pode ser designada à classe de dados faciais a . Entretanto, uma abordagem de classificação diferente é proposta nesse trabalho para lidar melhor com ruído e dados atípicos, como discutido na Seção 3.6.

3.6 Reconhecimento Facial

Em reconhecimento de padrões, é possível medir a probabilidade de erro de um classificador, e se determinou que a probabilidade de erro da regra do vizinho mais próximo é delimitada como sendo acima do dobro da probabilidade de erro de Bayes (COVER; HART, 1967) (DEVROYE; GYÖRFI; LUGOSI, 1996). Portanto, o classificador do vizinho mais próximo comumente usado na abordagem de reconhecimento facial baseada na aparência pode ser insuficiente, especialmente por que esse classificador frequentemente falha com dados atípicos e amostras de dados de alta dimensionalidade (COVER; HART, 1967) (DEVROYE; GYÖRFI; LUGOSI, 1996).

Por outro lado, a abordagem de classificação por *Máquinas de Vetores de Suporte* (*Support Vector Machines* - SVM), onde não é necessário assumir nenhuma hipótese sobre a distribuição dos dados (BOSER; GUYON; VAPNIK, 1992)(CORTES; VAPNIK, 1995), tem se provado eficiente em muitas aplicações de classificação, como reconhecimento facial (PHAM; VENKATESH, 2011) (JIA; MARTINEZ, 2009), estimação de idade (GUO et al., 2009) (WU; TURAGA; CHELLAPPA, 2012), classificação de imagens hiperespectrais (HE et al., 2014), e outras aplicações envolvendo dados de alta dimensionalidade. Portanto, a fim de lidar melhor com problemas de classificação em reconhecimento facial, é proposto realizar a classificação de novas imagens faciais usando SVM com margens suaves.

Inicialmente, considera-se o caso típico de classificação com duas classes sem sobreposição. De acordo com a formulação típica do SVM, duas classes distintas podem ser separadas determinando o hiperplano de separação que apresenta a maior margem entre ambas, como exemplificado na figura 3.3, que representa uma ilustração de um problema típico de SVM com duas classes. Nessa figura, os eixos são empíricos, sendo o hiperplano de separação posicionado no meio da margem de separação, que é máxima. Círculos azuis representam imagens faciais de uma classe e os quadrados vermelhos representam imagens faciais da outra classe.

O objetivo do SVM típico é maximizar a margem $m = \frac{2}{\|\vec{w}\|}$, separando as imagens faciais de duas classes a e a' . Considerando \vec{w} como o vetor normal ao hiperplano de separação, e q como o deslocamento em relação à origem do hiperplano de separação ao longo da direção de \vec{w} , o hiperplano de separação $\vec{w}^T y_{ij} + q = 0$ será posicionado no meio dos hiperplanos de apoio $\vec{w}^T y_{ij} + q = 1$ e $\vec{w}^T y_{ij} + q = -1$, e pode ser determinado por otimização, como segue (BOSER; GUYON; VAPNIK, 1992)(CORTES; VAPNIK, 1995):

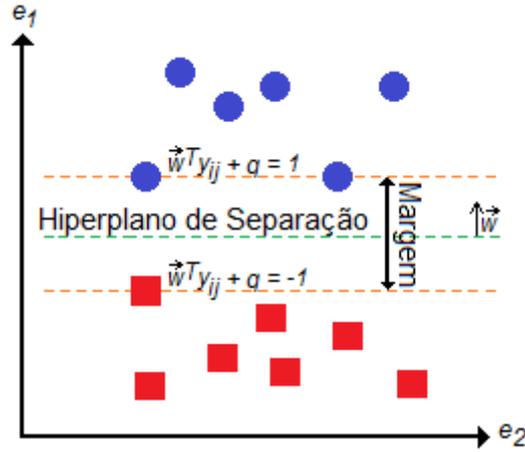


Figura 3.3: Ilustração de um problema típico de SVM com duas classes.

$$\min_{\vec{w}, q} \frac{1}{2} \|\vec{w}\|^2, \quad (3.8)$$

$$\text{sujeito à } z_i(\vec{w}^T y_{ij} + q) \leq 1, \quad \forall j \text{ na classe } i \in \{a, a'\}, \quad (3.9)$$

de forma que x_{ij} é um vetor de feições faciais de alta dimensionalidade representando a imagem facial de classe i e pose da cabeça j projetado em y_{ij} em um espaço de baixa dimensionalidade usando a Equação 3.3. Além disso, z_i é uma variável que define os rótulos das classes, sendo $z_i = -1$ quando $i = a$, e $z_i = 1$ quando $i = a'$. Como resultado desse problema de otimização, q e \vec{w} são obtidos, definindo o hiperplano de separação que maximiza o tamanho da margem definida por $2/\|\vec{w}\|$. Finalmente, a função de classificação tem a seguinte forma:

$$F(y_{ij}) = \text{sign}(\vec{w}^T y_{ij} + q), \quad (3.10)$$

onde o sinal de $F(\cdot)$ indica a atribuição da classe. Se $F(y_{ij})$ é positivo, a imagem facial y_{ij} é atribuída à classe a' , e se é negativa, a imagem facial y_{ij} é atribuída à classe a .

Relaxando esse problema de otimização, o caso de classificação com duas classes com sobreposição pode ser tratado como um problema de margem suave. Nesse caso, o hiperplano de separação é posicionado com certa quantidade de folga em relação às imagens faciais ao longo das fronteiras das duas classes. Esta abordagem permite alguns dados atípicos serem classificados erroneamente, mas tende a aumentar a acurácia geral uma vez que as fronteiras das classes são definidas otimamente. Portanto, um novo problema de otimização de duas classes é definido com variáveis de folga ξ_{ij} usando a seguinte função de custo e restrições (FAN; CHEN; LIN, 2005)(CHANG; LIN, 2011):

$$\min_{w, q, \xi} \frac{1}{2} \|\vec{w}\|^2 + C_{SVM} \sum_{i=\{a, a'\}, j} \xi_{ij}, \quad (3.11)$$

sujeito à:

$$z_i(\vec{w}^T y_{ij} + q) \leq 1 - \xi_{ij}, \quad \xi_{ij} > 0, \quad \forall j \text{ na classe } i, \quad (3.12)$$

onde C_{SVM} é um parâmetro que define o equilíbrio entre erro de classificação e precisão de posicionamento da margem. Considerando um problema de classificação SVM multi-classes, a classificação de uma nova imagem facial pode ser realizada eficientemente

usando a abordagem de um-contra-um (*one-against-one*) (CHANG; LIN, 2011)(KRESSEL, 1999). Nesse caso, $A(A - 1)/2$ classificadores são computados e cada classificador é treinado para discriminar somente entre duas classes distintas. Consequentemente, o seguinte problema de otimização é resolvido para cada possível par de classes a e a' , obtendo o vetor $\vec{w}^{aa'}$ (FAN; CHEN; LIN, 2005)(CHANG; LIN, 2011):

$$\min_{\vec{w}^{aa'}, q^{aa'}, \xi^{aa'}} \frac{1}{2} \|\vec{w}^{aa'}\|^2 + C_{SVM} \sum_{i \in \{a, a'\}, j} \xi_{ij}^{aa'}, \quad (3.13)$$

sujeito à $\xi_{ij}^{aa'} \geq 0$, e:

$$\begin{aligned} (\vec{w}^{aa'})^T y_{ij} + q^{aa'} &\geq 1 - \xi_{ij}^{aa'}, \text{ se } y_{ij} \text{ está na classe } a, \\ (\vec{w}^{aa'})^T y_{ij} + q^{aa'} &\leq -1 + \xi_{ij}^{aa'}, \text{ se } y_{ij} \text{ está na classe } a'. \end{aligned}$$

Similar a Equação 3.10, a função de classificação corresponde à distância normal (com sinal) da imagem facial y_{ij} ao hiperplano separando as classes a e a' , e é dada por:

$$F(y_{ij}) = \text{sign}((\vec{w}^{aa'})^T y_{ij} + q^{aa'}). \quad (3.14)$$

Quando o hiperplano determinado pelos vetores $\vec{w}^{aa'}$ é encontrado para cada par de classes a e a' , $A(A - 1)/2$ testes de classificação binária são executados com a função de classificação na Equação 3.14 para classificar uma nova imagem facial y_{uv} . Em cada teste, a imagem facial y_{uv} é repetidamente designada a uma classe de duas, lançando um voto para a classe designada. Ao fim de $A(A - 1)/2$ testes de classificação, a imagem facial y_{uv} é atribuída à classe com o maior número de votos. No caso de um empate, uma classe é escolhida aleatoriamente entre as classes empatadas (CHANG; LIN, 2011)(KRESSEL, 1999).

A abordagem de classificação com SVM usando margens suaves pode lidar com dados ruidosos e dados atípicos uma vez que ela minimiza o erro de classificação no caso de duas classes com sobreposição, superando as limitações do classificador do vizinho mais próximo. Além disso, a abordagem com SVM usando margens suaves também pode lidar melhor com dados de alta dimensionalidade e problemas não separáveis linearmente (FAN; CHEN; LIN, 2005) (KRESSEL, 1999) (comuns em dados faciais (HE et al., 2005)).

4 RESULTADOS EXPERIMENTAIS

Foram realizados testes com o objetivo de comparar a eficiência do método de reconhecimento facial proposto em relação ao método OLPP típico (CAI et al., 2006) e métodos representativos do estado da arte, usando tanto imagens em alta resolução como em baixa resolução, e tanto imagens coloridas (RGB) como imagens em escalas de cinza. Para isso foram utilizadas 5 bases de imagens tipicamente usadas em reconhecimento facial que apresentam diversos fatores que influenciam o reconhecimento facial na prática, como variações na pose da cabeça, expressão facial, iluminação, presença artefatos de oclusão, etnia, gênero, entre outros. As bases de imagens faciais escolhidas são a PUT (ANDRZEJ Kasiński ANDRZEJ FLOREK, 2008), a FEI (THOMAZ; GIRALDI, 2010), a FERET (PHILLIPS et al., 1998), a Yale (BELHUMEUR; HESPANHA; KRIEGMAN, 1997) e a ORL (SAMARIA; HARTER, 1994), sendo os resultados apresentados nas Seções 4.2, 4.3 e 4.4. A discussão sobre os resultados está na Seção 4.5.

A seguir, são apresentados resultados obtidos através de um conjunto de experimentos designados para avaliar e comparar o método proposto com outros métodos representativos do estado da arte sob várias condições desafiadoras encontradas na prática (tais como variações de pose da cabeça, expressão facial e iluminação). Os experimentos de reconhecimento facial (identificação facial) realizados são divididos em três conjuntos, no primeiro (Seção 4.2), o método proposto de extração de feições faciais esparsas coloridas é aplicado usando três bases de imagens faciais diferentes coloridas em alta resolução, a PUT (ANDRZEJ Kasiński ANDRZEJ FLOREK, 2008), a FEI (THOMAZ; GIRALDI, 2010) e a FERET (PHILLIPS et al., 1998) (descritas na Seção 4.1). Essas populares bases de imagens faciais apresentam um alto número de indivíduos em diferentes poses, com os detalhes da face (por exemplo, os olhos) claramente visíveis em imagens coloridas (RGB) de alta resolução. Os experimentos mencionados acima são apresentados nas Tabelas 4.1, 4.2 e 4.3, incluindo também resultados para as mesmas imagens convertidas para escalas de cinza.

No segundo conjunto de experimentos (Seção 4.3), o método proposto para redução de dimensionalidade e classificação de imagens faciais é avaliado usando representações densas de feições faciais (como na abordagem baseada na aparência). Nesses experimentos, foram usadas as mesmas imagens faciais coloridas em alta resolução das bases PUT (ANDRZEJ Kasiński ANDRZEJ FLOREK, 2008), FEI (THOMAZ; GIRALDI, 2010) e FERET (PHILLIPS et al., 1998) (descritas na Seção 4.1) usando o método proposto de extração de feições faciais densas coloridas proposto na Seção 3.3. Além disso, as mesmas imagens faciais foram convertidas para escalas de cinza a fim de comparar o método proposto com o método típico de extração de feições faciais densas da abordagem baseada na aparência (TURK; PENTLAND, 1991) (HE et al., 2005) (CAI et al., 2006), que converte imagens faciais em escalas de cinza para forma vetorial. Os experimentos mencionados

acima são apresentados nas Tabelas 4.4, 4.5 e 4.6.

Por último, a generalidade do método proposto de reconhecimento facial é avaliada através de outro conjunto de testes que usa bases de imagens faciais em baixa resolução e em escalas de cinza tipicamente usadas em reconhecimento facial, (por exemplo, no método OLPP (CAI et al., 2006)). As bases de imagens faciais Yale (BELHUMEUR; HESPANHA; KRIEGMAN, 1997) e ORL (SAMARIA; HARTER, 1994) forneceram as imagens em baixa resolução em escalas de cinza usadas nesses experimentos (descritas na Seção 4.1). Feições densas em escalas de cinza foram obtidas nesses experimentos segundo a abordagem baseada na aparência (TURK; PENTLAND, 1991) (HE et al., 2005) (CAI et al., 2006) e usadas nos experimentos com o método proposto e métodos representativos do estado da arte. Os experimentos mencionados acima são apresentados nas Tabelas 4.7 e 4.8.

Os métodos indicados nas Tabelas 4.1, 4.2, 4.3, 4.4, 4.5, 4.6, 4.7 e 4.8 são: *Projeções Ortogonais Preservadoras de Localidade (Orthogonal Locality Preserving Projections - OLPP)* (CAI et al., 2006), *ASM Eficiente* (BEHAINE; SCHARCANSKI, 2012), *Máquinas de Vetores de Suporte (Support Vector Machines - SVM)* (FAN; CHEN; LIN, 2005)(CHANG; LIN, 2011), *Análise Discriminante com Regressão Espectral com Kernels (Efficient Kernel Discriminant Analysis via Spectral Regression - SRKDA)* (SCHOLKOPF; SMOLA, 2001), *Análise Discriminante Multi-Visões (Multi-view Discriminant Analysis - MvDA)* (KAN et al., 2012), *Eigenfaces* (TURK; PENTLAND, 1991), *Fisherfaces* (BELHUMEUR; HESPANHA; KRIEGMAN, 1997) e *Laplacianfaces* (HE et al., 2005). O método OLPP é apresentado usando quatro esquemas distintos de definição de localidade, onde OLPP1 se refere ao esquema de definição de localidade usado pelo método OLPP típico (CAI et al., 2006) e OLPP2, OLPP3 e OLPP4 são os esquemas propostos de definição de localidade, levando à melhora da separabilidade de classes de dados faciais, ou seja, indivíduos (Seção 3.4).

Para cada um dos métodos mencionados nas Tabelas 4.1, 4.2, 4.3, 4.4, 4.5, 4.6, 4.7 e 4.8, os parâmetros de cada método que levam aos melhores resultados experimentais foram obtidos através da aplicação de repetidos testes em intervalos de parâmetros, possibilitando que sejam criados refinamentos sucessivos dos parâmetros, até que valores de parâmetros que levam aos melhores resultados em cada método foram determinados, sendo o processo de determinação dos parâmetros realizado manualmente, o que é uma tarefa custosa em termos de tempo. O parâmetro d no método *Eigenfaces*, *Laplacianfaces* e OLPP representa a dimensionalidade do espaço reduzido, t representa a largura de banda da função de *kernel* usada para definir localidade baseada em k vizinhos, e r representa a proporção usada no PCA para compactar dados, que também é usada por *Fisherfaces* e MvDA para reduzir a dimensionalidade dos dados.

Em todos os experimentos onde a classificação ocorre por SVM com margens suaves, o parâmetro C_{SVM} foi definido como 1, e a implementação adotada de SVM com margens suaves foi a LIBSVM (CHANG; LIN, 2011) (que resolve o problema de otimização na Equação 3.13 usando um *kernel* linear). No método ASM Eficiente (BEHAINE; SCHARCANSKI, 2012), o parâmetro α foi definido como 1 refletindo mais importância para medidas locais nas vizinhanças dos *landmarks*. No método SRKDA, o *kernel* Gaussiano com desvio padrão $\sigma = 20000$ foi usado, e os demais parâmetros de cada método estão definidos nas tabelas 4.1, 4.2, 4.3, 4.4, 4.5, 4.6, 4.7 e 4.8, onde "NN" denota o classificador do vizinho mais próximo, "OLPP1 + NN" corresponde ao algoritmo OLPP típico (CAI et al., 2006) e "SVM" representa o classificador SVM com margens suaves. As taxas médias de reconhecimento facial para cada método apresentadas nas mesmas

tabelas foram obtidas através da divisão do número de acertos pelo número total de experimentos.

4.1 Materiais e Métodos Usados nos Experimentos

Os métodos utilizados nos experimentos apresentados no Capítulo 4 são os métodos *Projeções Ortogonais Preservadoras de Localidade* (*Orthogonal Locality Preserving Projections* - OLPP) (CAI et al., 2006), ASM Eficiente (BEHAINE; SCHARCANSKI, 2012), *Máquinas de Vetores de Suporte* (*Support Vector Machines* - SVM) (FAN; CHEN; LIN, 2005)(CHANG; LIN, 2011), *Análise Discriminante com Regressão Espectral com Kernels* (*Efficient Kernel Discriminant Analysis via Spectral Regression* - SRKDA) (SCHOLKOPF; SMOLA, 2001), *Análise Discriminante Multi-Visões* (*Multi-view Discriminant Analysis* - MvDA) (KAN et al., 2012), *Eigenfaces* (TURK; PENTLAND, 1991), *Fisherfaces* (BELHUMEUR; HESPANHA; KRIEGMAN, 1997) e *Laplacianfaces* (HE et al., 2005). O método OLPP é apresentado usando quatro esquemas distintos de definição de localidade, onde OLPP1 se refere ao esquema de definição de localidade usado pelo método OLPP típico (CAI et al., 2006) e OLPP2, OLPP3 e OLPP4 são os esquemas propostos de definição de localidade, levando à melhora da separabilidade de classes de dados faciais, ou seja, indivíduos (Seção 3.4).

A base de imagens faciais PUT (ANDRZEJ Kasiński ANDRZEJ FLOREK, 2008) apresenta imagens faciais coloridas (RGB) em alta resolução (2048×1536 pixels), e foi criada para disponibilizar dados confiáveis para algoritmos de visão computacional para o objetivo de localização de faces, extração de feições e reconhecimento facial. A base apresenta imagens faciais com variação de gênero tiradas contra um plano de fundo uniforme sob condições de iluminação parcialmente controladas apresentando pequenas variações de iluminação. A principal fonte de variação na aparência facial são mudanças na pose (com a cabeça levantando ou abaixando e virando da esquerda para a direita, e vice-versa).

Experimentos serão realizados (Capítulo 4) com todas as 100 classes de dados faciais (indivíduos) disponíveis na base de imagens faciais PUT usando as poses pertencentes ao subconjunto de aprendizado da PUT (*PUT Learning Set*), que contém 2200 imagens faciais. Nesse subconjunto, existem 22 poses (imagens faciais) disponíveis para cada classe de dados faciais como exemplificado na Figura 4.1, porém não é uma boa estratégia usar todas as imagens para treinamento, pois em aplicações práticas, poucas imagens de treinamento são disponíveis por indivíduo, sendo melhor usar menos imagens por teste. Cada método citado anteriormente foi executado 10 vezes usando o subconjunto de teste da base de imagens faciais PUT (*PUT Learning Set*). Em cada execução, 9 poses (imagens faciais) em cada classe de dados faciais foram aleatoriamente selecionadas para treinamento e uma pose (imagem facial) em cada classe de dados faciais foi aleatoriamente selecionada para teste (sem contar as imagens escolhidas para treinamento). A seleção aleatória aproxima melhor situações reais, uma vez que em aplicações práticas pode-se ter imagens de treinamento ou teste com poses e expressões faciais diversas.

A base de imagens faciais FEI (THOMAZ; GIRALDI, 2010) apresenta imagens faciais coloridas (RGB) em alta resolução (640×480 pixels), e é usada em diversos algoritmos de visão computacional, principalmente em reconhecimento facial. Essa base de imagens faciais apresenta imagens faciais com variação de gênero tiradas contra um plano de fundo branco uniforme com o corpo dos indivíduos na posição frontal, com rotação da cabeça, de forma que a cabeça vire da esquerda para a direita e vice-versa, incluindo variações na



Figura 4.1: Poses disponíveis na base de imagens faciais PUT usadas nos experimentos.

expressão facial, etnia, e outros.

Experimentos serão realizados (Capítulo 4) com todas as 200 classes de dados faciais (indivíduos) disponíveis na base de imagens faciais FEI usando as 10 poses (imagens faciais) indicadas na Figura 4.2, totalizando 2000 imagens faciais. Cada método citado anteriormente foi executado 10 vezes usando todo o subconjunto de teste especificado para a base FEI. Em cada execução, 9 poses (imagens faciais) em cada classe de dados faciais foram aleatoriamente selecionadas para treinamento e a pose restante (imagem facial) em cada classe de dados faciais foi selecionada para teste.



Figura 4.2: Poses disponíveis na base de imagens faciais FEI usadas nos experimentos.

A base de imagens faciais FERET (PHILLIPS et al., 1998) apresenta imagens faciais coloridas (RGB) em alta resolução (512×768 pixels), e foi criada para apoiar o desenvolvimento de novos algoritmos para localização automática de faces humanas e disponibiliza um grande conjunto de indivíduos e imagens faciais. Essa base apresenta imagens faciais com variação de gênero. As imagens faciais foram tiradas sob condições de iluminação controladas e é organizada em diversos subconjuntos com variações específicas de pose da cabeça, expressão, idade, entre outros.

Experimentos serão realizados (Capítulo 4) com as imagens faciais coloridas da base FERET pertencentes às primeiras 200 classes de dados faciais (indivíduos) que contém

os subconjuntos *fa*, *fb*, *hl*, *hr* e *rc*, incluindo todas as 6 poses de cabeça (imagens faciais), totalizando 1200 imagens faciais (6 imagens faciais por classe), como exemplificado na Figura 4.3. Cada método citado anteriormente foi executado 6 vezes usando o subconjunto de teste especificado para a base FERET. Em cada execução, 5 poses (imagens faciais) em cada classe de dados faciais foram aleatoriamente selecionadas para treinamento e a pose restante (imagem facial) em cada classe de dados faciais foi selecionada para teste.



Figura 4.3: Poses disponíveis na base de imagens faciais FERET usadas nos experimentos.

A base de imagens faciais Yale (BELHUMEUR; HESPANHA; KRIEGMAN, 1997) apresenta imagens faciais em escalas de cinza em baixa resolução (243×320 pixels), sendo tipicamente usada em métodos de reconhecimento facial. A base contém 165 imagens faciais pertencentes a 15 classes de dados faciais (indivíduos), com 11 imagens faciais por indivíduo com variação de gênero, expressões faciais e condições de iluminação, conforme Figura 4.4.

Experimentos serão realizados (Capítulo 4) com as imagens faciais em escala de cinza da base Yale pertencentes aos 15 indivíduos da base, de forma que as imagens faciais foram reduzidas ao tamanho de 100×100 pixels para reduzir a complexidade computacional. Cada método citado anteriormente foi executado 11 vezes usando toda a base de imagens Yale. Em cada execução, 10 poses (imagens faciais) em cada classe de dados faciais (indivíduos) foram aleatoriamente selecionadas para treinamento e a pose restante (imagem facial) em cada classe de dados faciais foi selecionada para teste.



Figura 4.4: Poses disponíveis na base de imagens faciais Yale usadas nos experimentos.

A base de imagens faciais ORL (SAMARIA; HARTER, 1994) apresenta imagens faciais em escalas de cinza em baixa resolução (112×92 pixels), e é comumente usada em reconhecimento facial. A base contém 400 imagens faciais de 40 indivíduos, com 10 imagens faciais por indivíduo (conforme Figura 4.5) com diferentes gêneros, variações na expressão facial, luz e detalhes da face, além disso, as imagens faciais contêm pequenas variações de pose (inclinação da cabeça) de até $\pm 20^\circ$. Todas as imagens foram tiradas contra um plano de fundo escuro homogêneo.

Experimentos serão realizados (Capítulo 4) com as imagens faciais em escalas de cinza da base ORL pertencente aos 40 indivíduos da base, usando a resolução original das imagens faciais. Cada método citado anteriormente foi executado 10 vezes usando

toda a base ORL. Em cada execução, 9 poses (imagens faciais) em cada classe de dados faciais (indivíduos) foram aleatoriamente selecionadas para treinamento e a pose restante (imagem facial) em cada classe de dados faciais foi selecionada para teste.



Figura 4.5: Poses disponíveis na base de imagens faciais ORL usadas nos experimentos.

A seguir, no Capítulo 4, são apresentados os resultados experimentais obtidos com o método proposto, assim como comparações com outros métodos representativos do estado da arte, geralmente usados em reconhecimento facial, usando as bases de imagens selecionadas.

4.2 Experimentos com Feições Faciais Esparsas

O método de extração de feições faciais esparsas proposto na Seção 3.2 foi aplicado nos experimentos e os resultados são relatados nas Tabelas 4.1, 4.2 e 4.3 usando as bases de imagens faciais PUT, FEI e FERET. A topologia de *landmarks* usada nos experimentos com as mesmas bases de imagens faciais é mostrada na Figura 3.2. Essa topologia foi criada com o objetivo de selecionar informação biométrica importante em imagens faciais em cores (RGB), e tem 6 *landmarks* manualmente demarcados incluindo 10 *landmarks* igualmente espaçados interpolados entre cada par de *landmarks* adjacentes, levando a um total de $C = 56$ *landmarks* por imagem facial. Usando janelas de tamanho 11×11 ($V = 11$) com três canais de cores, cada vetor de feições faciais de alta dimensionalidade x_{ab} tem $(3CV^2)$ 20328 dimensões (no caso de imagens coloridas) ou (CV^2) 6776 dimensões (no caso de imagens em escalas de cinza). A topologia escolhida, a quantidade de *landmarks* interpolados e o tamanho de janela V foram fixados nos experimentos das Tabelas 4.1, 4.2 e 4.3, sendo que os mesmos foram escolhidos experimentalmente através de testes comparativos usando outras topologias de *landmarks* que avaliam outros subconjuntos da informação facial total que existe nas imagens faciais, porém trabalhos futuros poderiam ampliar a investigação da influência da topologia e dos referidos parâmetros no reconhecimento facial.

Um conjunto de experimentos foi realizado usando as imagens faciais coloridas em alta resolução da base de imagens faciais PUT (ANDRZEJ Kasiński ANDRZEJ FLOREK, 2008) usando feições esparsas coloridas e em escalas de cinza (convertendo as imagens faciais para escalas de cinza) usando a estrutura de testes e seleção de imagens para treinamento e teste descritas na Seção 4.1. A Tabela 4.1 mostra a média das taxas de reconhecimento facial obtidas em 1000 experimentos usando o esquema proposto de extração de feições faciais esparsas descrito na Seção 3.2 aplicado a métodos representativos do estado da arte. Cada vetor de feições faciais obtido foi normalizado para o intervalo $[0, 1]$.

Um conjunto de experimentos foi realizado usando as imagens faciais coloridas em alta resolução da base de imagens faciais FEI (THOMAZ; GIRALDI, 2010) usando feições esparsas coloridas e em escalas de cinza (convertendo as imagens faciais para escalas de cinza) usando a estrutura de testes e seleção de imagens para treinamento e teste descritas na Seção 4.1. A Tabela 4.2 mostra a média das taxas de reconhecimento facial obtidas

Métodos × Taxas de Reconhecimento	RGB	Tons de Cinza
OLPP4+SVM ($d = 30, k = 10, r = 0.94$)	100%	99.0%
OLPP4+NN ($d = 30, k = 10, r = 0.94$)	100%	98.3%
OLPP3+SVM ($d = 52, k = 1, r = 0.54$)	99.2%	96.7%
OLPP3+NN ($d = 55, k = 1, r = 0.54$)	98.8%	93.6%
OLPP2+SVM ($d = 30, k = 10, t = 500, r = 0.94$)	100%	98.7%
OLPP2+NN ($d = 30, k = 10, t = 500, r = 0.94$)	100%	97.8%
OLPP1+SVM ($d = 65, k = 1, t = 500, r = 0.61$)	99.1%	96.7%
OLPP1+NN ($d = 65, k = 1, t = 500, r = 0.60$)	98.4%	93.6%
ASM Eficiente ($\alpha = 1$)	98.8%	91.5%
SVM ($C_{SVM} = 1$)	99.2%	96.5%
SRKDA ($\sigma = 20000$)	99.8%	96.8%
MvDA ($d = 90$)	97.0%	94.1%
<i>Eigenfaces</i> ($d = 50$)	97.0%	91.3%
<i>Fisherfaces</i> ($r = 0.98$)	99.7%	95.1%
<i>Laplacianfaces</i> ($d = 60, k = 10, t = 500, r = 0.34$)	99.0%	93.7%

Tabela 4.1: Comparação de métodos aplicados a reconhecimento facial (identificação facial) usando representações esparsas na base PUT.

em 2000 experimentos usando o esquema proposto de extração de feições faciais esparsas descrito na Seção 3.2 aplicado a métodos representativos do estado da arte. Cada vetor de feições faciais obtido foi normalizado para o intervalo $[0, 1]$.

Métodos × Taxas de Reconhecimento	RGB	Tons de Cinza
OLPP4+SVM ($d = 54, k = 10, r = 0.94$)	99.4%	91.8%
OLPP4+NN ($d = 55, k = 10, r = 0.94$)	99.3%	84.9%
OLPP3+SVM ($d = 51, k = 1, r = 0.39$)	92.0%	86.8%
OLPP3+NN ($d = 50, k = 1, r = 0.39$)	85.5%	79.8%
OLPP2+SVM ($d = 46, k = 10, t = 500, r = 0.89$)	99.3%	91.5%
OLPP2+NN ($d = 48, k = 10, t = 500, r = 0.90$)	98.9%	82.9%
OLPP1+SVM ($d = 54, k = 1, t = 500, r = 0.38$)	91.3%	86.8%
OLPP1+NN ($d = 54, k = 1, t = 500, r = 0.38$)	83.4%	79.8%
ASM Eficiente ($\alpha = 1$)	64.1%	61.1%
SVM ($C_{SVM} = 1$)	92.8%	87.9%
SRKDA ($\sigma = 20000$)	65.6%	60.1%
MvDA ($d = 100$)	79.4%	72.8%
<i>Eigenfaces</i> ($d = 50$)	82.8%	77.9%
<i>Fisherfaces</i> ($r = 0.95$)	95.6%	69.8%
<i>Laplacianfaces</i> ($d = 30, k = 6, t = 500, r = 0.41$)	83.1%	79.5%

Tabela 4.2: Comparação de métodos aplicados a reconhecimento facial (identificação facial) usando representações esparsas na base FEI.

Um conjunto de experimentos foi realizado usando as imagens faciais coloridas em alta resolução da base de imagens faciais FERET (PHILLIPS et al., 1998) usando feições esparsas coloridas e em escalas de cinza (convertendo as imagens faciais para escalas de cinza) usando a estrutura de testes e seleção de imagens para treinamento e teste descritas

na Seção 4.1. A Tabela 4.3 mostra a média das taxas de reconhecimento facial obtidas em 1200 experimentos usando o esquema proposto de extração de feições faciais esparsas descrito na Seção 3.2 aplicado a métodos representativos do estado da arte. Cada vetor de feições faciais obtido foi normalizado para o intervalo $[0, 1]$.

Métodos × Taxas de Reconhecimento	RGB	Tons de Cinza
OLPP4+SVM ($d = 54, k = 6, r = 0.81$)	93.8%	79.6%
OLPP4+NN ($d = 55, k = 6, r = 0.85$)	88.5%	73.8%
OLPP3+SVM ($d = 53, k = 1, r = 0.40$)	86.1%	80.8%
OLPP3+NN ($d = 57, k = 1, r = 0.41$)	75.7%	69.4%
OLPP2+SVM ($d = 54, k = 6, t = 500, r = 0.78$)	93.8%	79.6%
OLPP2+NN ($d = 54, k = 6, t = 500, r = 0.78$)	88.4%	66.6%
OLPP1+SVM ($d = 55, k = 3, t = 500, r = 0.36$)	83.6%	79.4%
OLPP1+NN ($d = 55, k = 1, t = 500, r = 0.37$)	69.0%	66.2%
ASM Eficiente ($\alpha = 1$)	72.5%	53.8%
SVM ($C_{SVM} = 1$)	86.3%	79.9%
SRKDA ($\sigma = 20000$)	64.9%	59.7%
MvDA ($d = 100$)	76.1%	71.4%
<i>Eigenfaces</i> ($d = 58$)	69.6%	64.1%
<i>Fisherfaces</i> ($r = 0.97$)	67.1%	65.7%
<i>Laplacianfaces</i> ($d = 60, k = 6, t = 500, r = 0.4$)	67.1%	65.5%

Tabela 4.3: Comparação de métodos aplicados a reconhecimento facial (identificação facial) usando representações esparsas na base FERET.

4.3 Experimentos com Feições Faciais Densas

Outro conjunto de experimentos foi realizado usando as mesmas imagens das bases de imagens faciais PUT, FEI e FERET selecionadas anteriormente nos experimentos das Tabelas 4.1, 4.2 e 4.3, porém foram utilizadas representações densas de feições faciais obtidas usando o método proposto de extração de feições faciais densas coloridas como descrito na Seção 3.3, incluindo também experimentos com feições densas em escalas de cinza conforme a abordagem baseada na aparência que comumente usa imagens em escalas de cinza (TURK; PENTLAND, 1991) (HE et al., 2005) (CAI et al., 2006), sendo os resultados apresentados nas Tabelas 4.4, 4.5 e 4.6.

Nos experimentos com representações de feições densas apresentados nas Tabelas 4.4, 4.5 e 4.6, as imagens faciais das bases de imagens faciais PUT, FEI e FERET foram reduzidas ao tamanho de 64×96 pixels para reduzir a complexidade computacional como comumente realizado na abordagem baseada na aparência (TURK; PENTLAND, 1991) (HE et al., 2005) (CAI et al., 2006). A resolução (64×96 pixels) usada nos experimentos das Tabelas 4.4, 4.5 e 4.6 é fixa e foi definida experimentalmente, porém trabalhos futuros poderiam investigar mais profundamente a influência da resolução no reconhecimento facial.

Nos experimentos reportados nas Tabelas 4.4, 4.5 e 4.6, foram obtidos vetores de feições com dimensionalidade de $3 \times 64 \times 96 = 18432$ no caso do uso do método proposto de extração de feições faciais densas coloridas como descrito na Seção 3.3 ou uma dimensionalidade de $64 \times 96 = 6144$ no caso do uso do método típico de extração de

feições faciais densas em escalas de cinza comumente aplicado na abordagem baseada na aparência (TURK; PENTLAND, 1991) (HE et al., 2005) (CAI et al., 2006).

Para a base de imagens faciais PUT, além dos experimentos com extração de feições faciais esparsas descritos na Tabela 4.1, foram realizados experimentos com extração de feições faciais densas, descritos na Tabela 4.4, sendo usadas as mesmas imagens faciais coloridas em alta resolução e a mesma estrutura de testes e seleção de imagens para treinamento e teste descritas na Seção 4.1. A Tabela 4.4 mostra a média das taxas de reconhecimento facial obtidas em 1000 experimentos usando o esquema proposto de extração de feições faciais densas descrito na Seção 3.2 usando imagens coloridas e em escalas de cinza aplicado a métodos representativos do estado da arte. Cada vetor de feições faciais obtido foi normalizado para o intervalo $[0, 1]$.

Métodos \times Taxas de Reconhecimento	RGB	Tons de Cinza
OLPP4+SVM ($d = 40, k = 10, r = 0.94$)	100%	100%
OLPP4+NN ($d = 40, k = 10, r = 0.94$)	100%	100%
OLPP3+SVM ($d = 30, k = 1, r = 0.50$)	100%	100%
OLPP3+NN ($d = 30, k = 1, r = 0.50$)	100%	100%
OLPP2+SVM ($d = 50, k = 10, t = 500, r = 0.94$)	100%	100%
OLPP2+NN ($d = 50, k = 10, t = 500, r = 0.94$)	100%	100%
OLPP1+SVM ($d = 50, k = 1, t = 500, r = 0.60$)	100%	100%
OLPP1+NN ($d = 50, k = 1, t = 500, r = 0.60$)	100%	100%
SVM ($C_{SVM} = 1$)	99.9%	99.9%
SRKDA ($\sigma = 20000$)	93.3%	99.3%
MvDA ($d = 100$)	99.7%	99.6%
<i>Eigenfaces</i> ($d = 50$)	99.9%	99.9%
<i>Fisherfaces</i> ($r = 0.98$)	99.9%	99.6%
<i>Laplacianfaces</i> ($d = 30, k = 10, t = 500, r = 0.37$)	99.9%	99.9%

Tabela 4.4: Comparação de métodos aplicados a reconhecimento facial (identificação facial) usando representações densas na base PUT.

Para a base de imagens faciais FEI, além dos experimentos com extração de feições faciais esparsas descritos na Tabela 4.2, foram realizados experimentos com extração de feições faciais densas, descritos na Tabela 4.5, sendo usadas as mesmas imagens faciais coloridas em alta resolução e a mesma estrutura de testes e seleção de imagens para treinamento e teste descritas na Seção 4.1. A Tabela 4.5 mostra a média das taxas de reconhecimento facial obtidas em 2000 experimentos usando o esquema proposto de extração de feições faciais densas descrito na Seção 3.2 usando imagens coloridas e em escalas de cinza aplicado a métodos representativos do estado da arte. Cada vetor de feições faciais obtido foi normalizado para o intervalo $[0, 1]$.

Para a base de imagens faciais FERET, além dos experimentos com extração de feições faciais esparsas descritos na Tabela 4.3, foram realizados experimentos com extração de feições faciais densas, descritos na Tabela 4.6, sendo usadas as mesmas imagens faciais coloridas em alta resolução e a mesma estrutura de testes e seleção de imagens para treinamento e teste descritas na Seção 4.1. A Tabela 4.6 mostra a média das taxas de reconhecimento facial obtidas em 1200 experimentos usando o esquema proposto de extração de feições faciais densas descrito na Seção 3.2 usando imagens coloridas e em escalas de cinza aplicado a métodos representativos do estado da arte. Cada vetor de feições faciais obtido foi normalizado para o intervalo $[0, 1]$.

Métodos × Taxas de Reconhecimento	RGB	Tons de Cinza
OLPP4+SVM ($d = 30, k = 10, r = 0.81$)	99.9%	99.6%
OLPP4+NN ($d = 40, k = 10, r = 0.71$)	99.8%	99.4%
OLPP3+SVM ($d = 46, k = 1, r = 0.72$)	99.5%	98.3%
OLPP3+NN ($d = 55, k = 1, r = 0.72$)	98.7%	96.6%
OLPP2+SVM ($d = 57, k = 10, t = 500, r = 0.80$)	99.9%	99.2%
OLPP2+NN ($d = 50, k = 10, t = 500, r = 0.70$)	99.8%	98.9%
OLPP1+SVM ($d = 57, k = 1, t = 500, r = 0.66$)	99.2%	97.9%
OLPP1+NN ($d = 55, k = 1, t = 500, r = 0.72$)	98.6%	96.3%
SVM ($C_{SVM} = 1$)	97.9%	96.1%
SRKDA ($\sigma = 20000$)	93.3%	93.3%
MvDA ($d = 100$)	94.2%	93.8%
<i>Eigenfaces</i> ($d = 50$)	92.9%	90.8%
<i>Fisherfaces</i> ($r = 0.95$)	99.3%	92.7%
<i>Laplacianfaces</i> ($d = 50, k = 10, t = 500, r = 0.42$)	95.8%	92.3%

Tabela 4.5: Comparação de métodos aplicados a reconhecimento facial (identificação facial) usando representações densas na base FEI.

Métodos × Taxas de Reconhecimento	RGB	Tons de Cinza
OLPP4+SVM ($d = 50, k = 6, r = 0.69$)	98.1%	95.2%
OLPP4+NN ($d = 57, k = 6, r = 0.74$)	98.0%	93.5%
OLPP3+SVM ($d = 49, k = 1, r = 0.34$)	87.8%	85.9%
OLPP3+NN ($d = 54, k = 1, r = 0.39$)	75.2%	69.8%
OLPP2+SVM ($d = 50, k = 6, t = 500, r = 0.69$)	98.0%	93.6%
OLPP2+NN ($d = 57, k = 6, t = 500, r = 0.73$)	98.0%	92.1%
OLPP1+SVM ($d = 50, k = 1, t = 500, r = 0.34$)	87.6%	82.5%
OLPP1+NN ($d = 57, k = 1, t = 500, r = 0.40$)	75.2%	69.6%
SVM ($C_{SVM} = 1$)	86.3%	83.0%
SRKDA ($\sigma = 20000$)	69.5%	68.9%
MvDA ($d = 100$)	64.8%	62.1%
<i>Eigenfaces</i> ($d = 50$)	73.7%	69.3%
<i>Fisherfaces</i> ($r = 0.6$)	87.3%	71.0%
<i>Laplacianfaces</i> ($d = 50, k = 6, t = 500, r = 0.75$)	70.7%	69.4%

Tabela 4.6: Comparação de métodos aplicados a reconhecimento facial (identificação facial) usando representações densas na base FERET.

4.4 Experimentos com Feições Faciais Densas em Imagens de Baixa Resolução

A fim de avaliar a generalidade do esquema proposto integrado para redução de dimensionalidade e classificação de imagens faciais, testes adicionais foram realizados de acordo com a abordagem baseada na aparência, que geralmente usa imagens de baixa resolução em escalas de cinza (TURK; PENTLAND, 1991) (HE et al., 2005) (CAI et al., 2006) para extrair feições densas de imagens faciais. Nesses experimentos, duas bases de imagens faciais típicas foram utilizadas, a Yale (BELHUMEUR; HESPANHA; KRIEGMAN, 1997) e a ORL (SAMARIA; HARTER, 1994), que apresentam imagens em baixa

resolução em escalas de cinza, sendo os resultados reportados nas Tabelas 4.7 e 4.8. É importante notar que não foram realizados experimentos com feições esparsas usando as bases Yale e ORL uma vez que imagens em baixa resolução não apresentam boa preservação de detalhes discriminativos (marcados por topologias de *landmarks*), o que pode implicar em imprecisões na representação de feições faciais. Concomitantemente, o método proposto de extração de feições faciais esparsas (Seção 3.2) é designado para ser aplicado a imagens faciais de alta resolução coloridas, que não é o caso das bases Yale e ORL.

Um conjunto de experimentos foi realizado usando as imagens faciais em escalas de cinza em baixa resolução da base de imagens faciais Yale (BELHUMEUR; HESPANHA; KRIEGMAN, 1997) usando a estrutura de testes e seleção de imagens para treinamento e teste descritas na Seção 4.1. A Tabela 4.7 mostra a média das taxas de reconhecimento facial obtidas em 165 experimentos usando o método típico de extração de feições faciais densas da abordagem baseada na aparência (TURK; PENTLAND, 1991) (HE et al., 2005) (CAI et al., 2006) aplicado a métodos representativos do estado da arte. Cada vetor de feições faciais obtido foi normalizado para o intervalo $[0, 1]$.

Métodos \times Taxas de Reconhecimento	Tons de Cinza
OLPP4+SVM ($d = 40, k = 11, r = 0.97$)	98.8%
OLPP4+NN ($d = 40, k = 11, r = 0.97$)	98.2%
OLPP3+SVM ($d = 49, k = 1, r = 0.75$)	89.7%
OLPP3+NN ($d = 44, k = 1, r = 0.78$)	78.2%
OLPP2+SVM ($d = 40, k = 10, t = 500, r = 0.97$)	98.8%
OLPP2+NN ($d = 40, k = 11, t = 500, r = 0.95$)	98.2%
OLPP1+SVM ($d = 55, k = 1, t = 500, r = 0.50$)	87.2%
OLPP1+NN ($d = 28, k = 1, t = 500, r = 0.60$)	78.2%
SVM ($C_{SVM} = 1$)	89.1%
SRKDA ($\sigma = 20000$)	98.2%
MvDA ($d = 80$)	71.7%
<i>Eigenfaces</i> ($d = 50$)	77.6%
<i>Fisherfaces</i> ($r = 0.98$)	98.1%
<i>Laplacianfaces</i> ($d = 40, k = 11, t = 500, r = 0.5$)	77.7%

Tabela 4.7: Comparação de métodos aplicados à identificação facial usando a base de imagens faciais Yale na abordagem clássica densa com imagens em escala de cinza em baixa resolução.

Um conjunto de experimentos foi realizado usando as imagens faciais em escalas de cinza em baixa resolução da base de imagens faciais ORL (SAMARIA; HARTER, 1994) usando a estrutura de testes e seleção de imagens para treinamento e teste descritas na Seção 4.1. A Tabela 4.8 mostra a média das taxas de reconhecimento facial obtidas em 165 experimentos usando o método típico de extração de feições faciais densas da abordagem baseada na aparência (TURK; PENTLAND, 1991) (HE et al., 2005) (CAI et al., 2006) aplicado a métodos representativos do estado da arte. Cada vetor de feições faciais obtido foi normalizado para o intervalo $[0, 1]$.

Métodos \times Taxas de Reconhecimento	Tons de Cinza
OLPP4+SVM ($d = 33, k = 10, r = 0.90$)	94.8%
OLPP4+NN ($d = 40, k = 10, r = 0.94$)	94.5%
OLPP3+SVM ($d = 45, k = 1, r = 0.45$)	94.0%
OLPP3+NN ($d = 40, k = 1, r = 0.45$)	93.3%
OLPP2+SVM ($d = 50, k = 10, t = 500, r = 0.95$)	94.0%
OLPP2+NN ($d = 50, k = 10, t = 500, r = 0.50$)	93.5%
OLPP1+SVM ($d = 55, k = 1, t = 500, r = 0.45$)	93.7%
OLPP1+NN ($d = 50, k = 1, t = 500, r = 0.48$)	93.0%
SVM ($C_{SVM} = 1$)	93.5%
SRKDA ($\sigma = 20000$)	89.0%
MvDA ($d = 80$)	78.1%
<i>Eigenfaces</i> ($d = 47$)	93.0%
<i>Fisherfaces</i> ($r = 0.99$)	91.7%
<i>Laplacianfaces</i> ($d = 52, k = 10, t = 500, r = 0.5$)	92.7%

Tabela 4.8: Comparação de métodos aplicados à identificação facial usando a base de imagens faciais ORL na abordagem clássica densa com imagens em escala de cinza em baixa resolução.

4.5 Discussão sobre os Resultados Experimentais

Conforme Tabelas 4.1, 4.2, 4.3, 4.4, 4.5, 4.6, 4.7 e 4.8, o método proposto usando tanto representações esparsas como densas de imagens faciais usando o esquema de definição de localidade OLPP4 (Seção 3.4) integrado à classificação com SVM usando margens suaves atinge taxas de reconhecimento maiores do que quando são aplicados os outros esquemas de definição de localidade OLPP3, OLPP2 e OLPP1, e inclusive apresentando maiores taxas de reconhecimento que métodos representativos do estado da arte e o método OLPP típico (CAI et al., 2006) (denotado por "OLPP1+NN"), mostrando que o método proposto consegue lidar melhor com problemas não separáveis linearmente que outros métodos. Os resultados indicam que o esquema de definição de localidade supervisionado OLPP4 aplicado a SVM com margens suaves ajuda a preservar a estrutura não-linear inerente às classes de dados faciais melhor do que os outros métodos mencionados. Nas mesmas tabelas, verifica-se que o método proposto atingiu as maiores taxas de reconhecimento em diferentes situações, tanto com imagens em alta resolução como imagens em baixa resolução, tanto com feições densas como esparsas, e tanto com imagens coloridas (RGB) como imagens em escalas de cinza.

Como observado nos experimentos das Tabelas 4.1, 4.2, 4.3, 4.4, 4.5, 4.6, 4.7 e 4.8, em todos os experimentos com representações faciais densas ou esparsas de imagens faciais, o esquema supervisionado OLPP4 atingiu taxas de reconhecimento maiores que o esquema não supervisionado OLPP3. Além disso, o esquema supervisionado OLPP2 atingiu taxas de reconhecimento maiores que o esquema não supervisionado OLPP1. Isso confirma uma tendência de que preservar a estrutura das classes de dados faciais no espaço de baixa dimensionalidade é mais discriminativo que preservar a estrutura dos dados faciais (SOLDERA; BEHAINE; SCHARCANSKI, 2015) (KOKIOPOULOU; SAAD, 2009), ou seja, a abordagem de reconhecimento facial supervisionada apresenta melhor discriminação de classes que a abordagem de reconhecimento facial não supervisionada.

Além disso, pode-se verificar que nas Tabelas 4.1, 4.2, 4.3, 4.4, 4.5, 4.6, 4.7 e 4.8,

o esquema supervisionado OLPP4 apresentou melhores resultados que o esquema supervisionado OLPP2 enquanto que o esquema não supervisionado OLPP3 apresentou melhores resultados que o esquema não supervisionado OLPP1, mostrando que definir o grafo de localidade através de afinidades com valores iguais (OLPP4 e OLPP3, com a Equação 3.2) é melhor que definir o grafo de localidade através de afinidades baseadas na estrutura Euclidiana (OLPP2 e OLPP1, com a Equação 3.1), ou seja, compactar as classes semelhantemente à LDA (BELHUMEUR; HESPANHA; KRIEGMAN, 1997) é mais discriminativo do que quando se preservam distâncias entre amostras faciais no espaço de baixa dimensionalidade.

Os resultados reportados nas Tabelas 4.1, 4.2, 4.3, 4.4, 4.5 e 4.6 que usam as bases de imagens faciais PUT, FEI e FERET, usando representações esparsas e densas de imagens faciais coloridas ou em escalas de cinza ajudam a confirmar uma tendência que a informação de cor ajuda a aumentar a discriminação das classes de dados faciais (indivíduos) no espaço de baixa dimensionalidade (BEHAINE; SCHARCANSKI, 2012) (SOLDERA; BEHAINE; SCHARCANSKI, 2015) (KOKIOPLOULOU; SAAD, 2009) em comparação a quando são usadas imagens em escalas de cinza, como no método OLPP típico (CAI et al., 2006).

Em comparação com a regra de classificação do vizinho mais próximo usada no método OLPP típico (CAI et al., 2006), a separação das classes de dados faciais tende a melhorar quando o esquema de classificação facial proposto com SVM é usado, provavelmente por que SVM com margens suaves é mais robusto a ruído e dados atípicos, levando a uma melhor performance em problemas não separáveis linearmente, como verificado nos resultados reportados nas Tabelas 4.1, 4.2, 4.3, 4.4, 4.5, 4.6, 4.7 e 4.8.

É importante mencionar que o método ASM Eficiente (BEHAINE; SCHARCANSKI, 2012) foi criado para receber feições faciais de entrada na forma de informação modal, mas na abordagem densa (com imagens inteiras), não existe informação modal disponível nos vetores de feições faciais obtidos, colocando o método ASM Eficiente em desvantagem, atingindo taxas de reconhecimento facial abaixo de 10%, portanto, o mesmo é omitido nas Tabelas 4.4, 4.5, 4.6, 4.7 e 4.8.

Nos experimentos apresentados nas Tabelas 4.1, 4.2, 4.3, 4.4, 4.5, 4.6, 4.7 e 4.8 com os esquemas de definição de localidade OLPP1, OLPP2, OLPP3 e OLPP4, nota-se que as maiores taxas de reconhecimento para os esquemas supervisionados OLPP4 e OLPP2 são atingidas quando o valor de k é máximo (que corresponde a definir a quantidade de vizinhos sendo igual ao número de imagens de treinamento utilizadas por classe) enquanto que as maiores taxas de reconhecimento para os esquemas não supervisionados OLPP3 e OLPP1 são atingidas quando o valor de k é mínimo (ou seja, o valor 1, que corresponde a somente um vizinho). Esse fato ocorre por que nos esquemas supervisionados, usar o máximo de vizinhos da mesma classe permite definir melhor as fronteiras de cada classe enquanto nos esquemas não supervisionados, usar o vizinho mais semelhante de qualquer classe implica em uma chance maior desse vizinho pertencer à mesma classe. Em outras palavras, na abordagem supervisionada, um número maior de vizinhos permite descrever melhor as classes, enquanto que na abordagem não supervisionada, quanto menor o k , menor a chance de se escolher vizinhos de outras classes. Essas observações são condizentes com a afirmação que a abordagem supervisionada apresenta melhor separação de classes do que a abordagem não supervisionada, e preservar a estrutura das classes é mais discriminativo que preservar a estrutura local dos dados (SOLDERA; BEHAINE; SCHARCANSKI, 2015) (KOKIOPLOULOU; SAAD, 2009).

Outro parâmetro que está presente nos experimentos apresentados nas Tabelas 4.1,

4.2, 4.3, 4.4, 4.5, 4.6, 4.7 e 4.8 é o t , que é usado pelo kernel do calor (Equação 3.1) no cálculo de afinidades entre imagens faciais nos métodos *Laplacianfaces* (HE et al., 2005) e OLPP (CAI et al., 2006). Como observado nos mesmos experimentos, o valor do t foi fixado em 500, correspondendo ao valor que causa uma boa preservação da estrutura dos dados descrita pela matriz S em todas as bases de imagens avaliadas nessa tese. O valor $t = 500$ foi determinado experimentalmente. Além disso, verificou-se que valores muito altos ou muito baixos para o parâmetro t causam distorção nos relacionamentos entre nodos vizinhos na construção do grafo de afinidades, conseqüentemente distorcendo o espaço de baixa dimensionalidade e a métrica de classificação.

Quanto aos parâmetros r (porcentagem de preservação dos dados no PCA) e d (dimensionalidade do espaço de baixa dimensionalidade) presentes em diversos métodos nos experimentos apresentados nas Tabelas 4.1, 4.2, 4.3, 4.4, 4.5, 4.6, 4.7 e 4.8, os valores dos mesmos que maximizam as taxas de reconhecimento foram determinados experimentalmente, como comentado anteriormente. Além disso, variações nos valores de outros parâmetros, como o k e o t (e no caso de dados esparsos, o tamanho da janela ao redor dos *landmarks* e a quantidade de *landmarks* interpolados) também influenciam os valores de r e d que maximizam a taxa de reconhecimento. As maiores taxas de reconhecimento tender a decorrer de valores altos ou medianos dos parâmetros r e d , porém não há uma indicação precisa sobre que valores de r e d que devem ser usados para maximizar a taxa de reconhecimento, já que eles precisam ser determinados experimentalmente, pois pequenas variações nos mesmos parâmetros influenciam sutilmente a separação de classes no espaço de baixa dimensionalidade obtido. Os parâmetros das Tabelas 4.1, 4.2, 4.3, 4.4, 4.5, 4.6, 4.7 e 4.8 foram determinados experimentalmente através de repetidos experimentos com refinamentos sucessivos, iniciando com valores de parâmetros de acordo com as sugestões explicadas anteriormente, tentando minimizar r e d para que fosse obtida uma boa performance computacional no treinamento e teste, ao mesmo tempo permitindo obter-se a maior taxa de reconhecimento nos mesmos experimentos.

Outra observação importante acerca dos valores dos parâmetros k , t , r e d que maximizam as taxas de reconhecimento nos experimentos das Tabelas 4.1, 4.2, 4.3, 4.4, 4.5, 4.6, 4.7 e 4.8 é que eles variam de acordo com as características das feições faciais extraídas das imagens faciais, ou seja, esses parâmetros mudam quando ocorre a mudança de base de imagem, tipo de método de extração de feições faciais (densas ou esparsas), tipo de imagem facial (colorida ou em escalas de cinza), que são fatores que alteram as feições faciais originais, inclusive causando considerável variação nos valores de parâmetros entre bases e métodos diferentes. Além disso, observa-se que esses parâmetros também podem variar conforme a definição de localidade empregada (OLPP4, OLPP3, OLPP2, OLPP1) e o classificador (NN ou SVM), pois valores diferentes desses parâmetros influenciam o espaço obtido de baixa dimensionalidade conjuntamente com o esquema de localidade e classificador empregados, conseqüentemente, influenciando a separação das classes no espaço de baixa dimensionalidade.

A abordagem de extração de feições faciais esparsa apresenta algumas vantagens em relação à abordagem densa, pois a abordagem esparsa pode usar imagens em alta resolução ao contrário da abordagem densa que necessita imagens em baixa resolução a fim de reduzir a complexidade computacional. Dessa forma, a abordagem esparsa preserva detalhes discriminativos das imagens faciais enquanto ignora informação ambígua e redundante. A influência do plano de fundo é minimizada na abordagem esparsa, pois somente as vizinhanças dos *landmarks* são selecionadas para criar vetores de feições, ao contrário da abordagem densa, que inclui o plano de fundo ao redor da face na criação de

vetores de feições faciais, o que pode causar influência relevante em aplicações práticas. Por outro lado, a abordagem esparsa possui uma desvantagem de requerer marcações de *landmarks* nas imagens faciais, sendo que as mesmas podem ser realizadas manualmente, porém, felizmente, existem diversos métodos confiáveis que detectam automaticamente *landmarks* (EFRATY et al., 2011) (ZHAO et al., 2011) (RAPP et al., 2011) (BURGOS-ARTIZZU; PERONA; DOLLAR, 2013).

Nos experimentos com representações faciais densas coloridas apresentados nas Tabelas 4.4, 4.5 e 4.6, foram obtidos melhores resultados do que os experimentos que usam as mesmas imagens, porém com representações faciais esparsas coloridas conforme os experimentos das Tabelas 4.1, 4.2 e 4.3. Entretanto, tratam-se de duas contribuições distintas ao estado da arte, sendo a primeira, a extração de feições esparsas coloridas (SOLDERA; BEHAINE; SCHARCANSKI, 2015), e a segunda, a extração de feições densas coloridas. Porém, os resultados com feições esparsas podem ser melhorados conforme descrito a seguir. Quanto aos experimentos com feições esparsas nas Tabelas 4.1, 4.2 e 4.3, o tamanho das janelas $V \times V$, a quantidade de *landmarks* interpolados e a topologia de *landmarks* empregada foram mantidas fixas em todos os experimentos a fim de que fosse simplificada a apresentação dos resultados para experimentos com feições faciais esparsas. Os mesmos foram determinados de forma experimental com o objetivo de se extrair feições faciais com discriminativas para todas as bases de imagens faciais testadas, porém mais testes podem ser realizados futuramente para avaliar o impacto da variação desses parâmetros no reconhecimento facial, o que pode potencialmente aumentar as taxas de reconhecimento na abordagem esparsa uma vez que o tamanho das janelas, a quantidade de *landmarks* interpolados e a topologia de *landmarks* utilizada influenciam a taxa de reconhecimento facial.

Para os experimentos realizados usando a base de imagens faciais PUT apresentados nas Tabelas 4.1 e 4.4, o método proposto com OLPP4 e SVM com margens suaves e o método OLPP típico apresentaram resultados ótimos usando feições densas coloridas, pois o subconjunto de teste da base PUT possui pouca variação de iluminação, expressão facial e pose da cabeça, facilitando o reconhecimento facial. Os experimentos descritos nas Tabelas 4.2 e 4.5 usando a base de imagens faciais FEI apresentaram excelentes resultados que foram afetados por maiores variações nas imagens faciais em decorrência de mudanças na iluminação, expressão facial, pose da cabeça, etnia, ruído e outros. Por outro lado, os resultados dos experimentos usando a base de imagens faciais FERET nas Tabelas 4.3 e 4.6, tiveram resultados inferiores às demais bases uma vez que essa base apresenta maiores variações de iluminação, pose e outros, representando um importante desafio em reconhecimento facial.

Para os experimentos reportados nas Tabelas 4.4, 4.5 e 4.6 usando o método proposto (OLPP4 e SVM com margens suaves) integrado ao esquema proposto de extração de feições faciais coloridas densas apresentado na Seção 3.3 e para os experimentos reportados nas Tabelas 4.7 e 4.8 usando o método proposto (OLPP4 e SVM com margens suaves) integrado ao esquema típico de extração de feições em escalas de cinza usado na abordagem baseada na aparência (CAI et al., 2006), os tempos médios de teste e treinamento são apresentados na Figura 4.6, onde (a), (b) e (c) indicam experimentos em cores (RGB) com as bases de imagens faciais PUT, FEI e FERET, respectivamente; (d), (e) e (f) indicam experimentos com as mesmas imagens faciais das bases PUT, FEI e FERET, porém convertidas para escalas de cinza, respectivamente; e (g) e (h) indicam experimentos com as imagens faciais em escalas de cinza das bases Yale e ORL, respectivamente.

Os tempos médios de processamento apresentados na Figura 4.6 obtidos com o mé-

tudo proposto (OLPP4 e SVM com margens suaves) sugerem que o método proposto potencialmente pode ser usado em aplicações de tempo real, uma vez que o tempo de processamento requerido para classificar cada imagem facial é menor que 0,25 segundos. Todos os testes foram realizados com um computador PC com processador Intel Core i5-3330 de 64 bits com processador de 3.3GHz, com 16GB de memória, usando um SSD de 480GB (*Solid-State Drive*), executando Matlab (que não é otimizado). Além disso, deve ser observado que a performance do sistema em modo de operação (classificação) requer um tempo de processamento menor que durante a fase de treinamento, uma vez que menos operações são requeridas para se classificar imagens faciais.

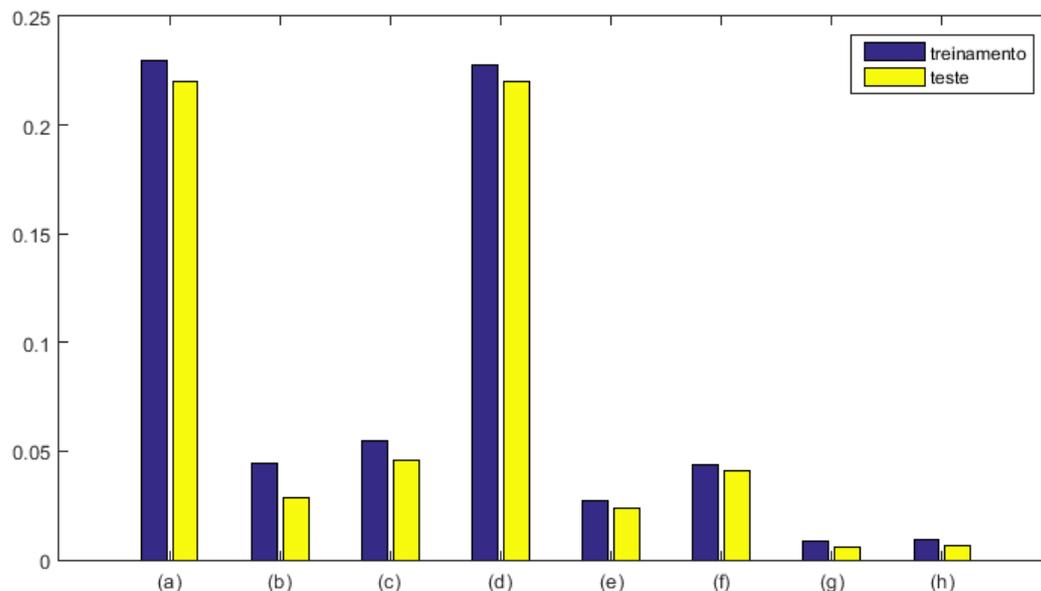


Figura 4.6: Tempos de processamento médios para treinamento e teste no método proposto (OLPP4 e SVM) em segundos.

Uma listagem mais ampla dos tempos de processamento do método proposto é apresentada nas Tabelas 4.9, 4.10, 4.11 e 4.12, apresentando tempos médios de processamento para treinamento e teste em segundos. Na Tabela 4.9, são apresentados os tempos médios de treinamento para o método proposto usando os esquemas de definição de localidade OLPP1, OLPP2, OLPP3 e OLPP4 com a classificação por SVM. Na Tabela 4.10, são apresentados os tempos médios de teste para o método proposto usando os esquemas de definição de localidade OLPP1, OLPP2, OLPP3 e OLPP4 com a classificação por SVM. Na Tabela 4.11, são apresentados os tempos médios de treinamento para o método proposto usando os esquemas de definição de localidade OLPP1, OLPP2, OLPP3 e OLPP4 com a classificação pela regra do vizinho mais próximo. E por fim, na Tabela 4.12, são apresentados os tempos médios de teste para o método proposto usando os esquemas de definição de localidade OLPP1, OLPP2, OLPP3 e OLPP4 com a classificação pela regra do vizinho mais próximo.

Algumas tendências podem ser verificadas nas Tabelas 4.9, 4.10, 4.11 e 4.12. Primeiramente, verifica-se que o tempo médio de treinamento por imagem é maior que o tempo médio de teste, devido a maior complexidade computacional apresentada no treinamento. Além disso, o tempo de treinamento e teste das imagens coloridas tende a ser maior que o tempo de treinamento e teste das imagens em escalas de cinza, respectivamente. Outra observação importante, é que os esquemas supervisionados OLPP4 e OLPP3 têm uma tendência de serem menores que os tempos de processamento dos esquemas OLPP2 e

OLPP1, respectivamente, pois apresentam menor complexidade computacional. Por fim, foram também realizadas medições de tempo de processamento usando feições esparsas, e foram obtidas aproximadamente as mesmas médias de tempo de processamento, portanto não havendo diferença significativa no tempo de processamento do método proposto usando feições densas ou esparsas.

Bases	OLPP4		OLPP3		OLPP2		OLPP1	
	RGB	Cinza	RGB	Cinza	RGB	Cinza	RGB	Cinza
PUT	0,2294	0,2276	0,2198	0,1739	0,2508	0,1852	0,1884	0,2263
FEI	0,0446	0,0273	0,0487	0,0293	0,0325	0,0545	0,0459	0,0284
FERET	0,0545	0,0439	0,0555	0,0443	0,0585	0,0489	0,0559	0,0448
ORL		0,0083		0,0088		0,0121		0,0090
Yale		0,0064		0,0080		0,0076		0,0062

Tabela 4.9: Tempos de treinamento por imagem do método proposto usando os esquemas OLPP1, OLPP2, OLPP3 e OLPP4 com SVM (em segundos). "RGB" indica experimentos com imagens coloridas e "Cinza" indica experimentos com imagens em tons de cinza.

Bases	OLPP4		OLPP3		OLPP2		OLPP1	
	RGB	Cinza	RGB	Cinza	RGB	Cinza	RGB	Cinza
PUT	0,2203	0,2201	0,2102	0,1708	0,2280	0,1774	0,1834	0,2159
FEI	0,0287	0,0240	0,0297	0,0251	0,0267	0,0326	0,0290	0,0244
FERET	0,0459	0,0409	0,0460	0,0412	0,0483	0,0452	0,0468	0,0418
ORL		0,0060		0,0060		0,0066		0,0061
Yale		0,0049		0,0049		0,0054		0,0050

Tabela 4.10: Tempos de teste por imagem do método proposto usando os esquemas OLPP1, OLPP2, OLPP3 e OLPP4 com SVM (em segundos). "RGB" indica experimentos com imagens coloridas e "Cinza" indica experimentos com imagens em tons de cinza.

Bases	OLPP4		OLPP3		OLPP2		OLPP1	
	RGB	Cinza	RGB	Cinza	RGB	Cinza	RGB	Cinza
PUT	0,2247	0,1773	0,2249	0,1785	0,2324	0,1843	0,2308	0,1777
FEI	0,0400	0,0279	0,0420	0,0283	0,0417	0,0301	0,0425	0,0290
FERET	0,0550	0,0445	0,0558	0,0446	0,0568	0,0453	0,0563	0,0446
ORL		0,0087		0,0093		0,0090		0,0092
Yale		0,0066		0,0083		0,0068		0,0067

Tabela 4.11: Tempos de treinamento por imagem do método proposto usando os esquemas OLPP1, OLPP2, OLPP3 e OLPP4 com o classificador do vizinho mais próximo (em segundos). "RGB" indica experimentos com imagens coloridas e "Cinza" indica experimentos com imagens em tons de cinza.

Bases	OLPP4		OLPP3		OLPP2		OLPP1	
	RGB	Cinza	RGB	Cinza	RGB	Cinza	RGB	Cinza
PUT	0,2174	0,1764	0,2171	0,1762	0,2173	0,1795	0,2215	0,1758
FEI	0,0310	0,0271	0,0310	0,0272	0,0311	0,0274	0,0310	0,0271
FERET	0,0487	0,0439	0,0487	0,0438	0,0487	0,0439	0,0490	0,0438
ORL		0,0068		0,0067		0,0068		0,0068
Yale		0,0053		0,0052		0,0052		0,0053

Tabela 4.12: Tempos de teste por imagem do método proposto usando os esquemas OLPP1, OLPP2, OLPP3 e OLPP4 com o classificador do vizinho mais próximo (em segundos). "RGB" indica experimentos com imagens coloridas e "Cinza" indica experimentos com imagens em tons de cinza.

5 CONCLUSÕES E TRABALHOS FUTUROS

Um novo método de reconhecimento facial foi apresentado nessa tese, que é apto a usar tanto representações esparsas como densas de imagens faciais, coloridas ou em escalas de cinza, apresentando melhores resultados que métodos representativos do estado da arte. Além disso, o método proposto nessa tese realiza diversas melhorias em relação ao método OLPP típico (CAI et al., 2006) em aspectos tais como extração de feições faciais, redução de dimensionalidade e reconhecimento facial. Conforme experimentos apresentados no Capítulo 4, o método proposto usando OLPP4 e SVM com margens suaves apresenta maiores taxas de reconhecimento que o método OLPP típico e métodos representativos do estado da arte, perdendo um pouco em performance computacional em relação ao método OLPP típico conforme Seção 4.5.

Um novo método de extração de feições faciais esparsas coloridas foi apresentado na Seção 3.2 de forma a preservar detalhes discriminativos de imagens faciais coloridas em alta resolução (SOLDERA; BEHAINE; SCHARCANSKI, 2015). Essa nova representação esparsa adiciona *landmarks* interpolados a topologias de *landmarks* para ajudar a compensar por imprecisões no posicionamento de *landmarks* nas imagens faciais durante a extração de feições faciais, ajudando a minimizar o impacto da oclusão de feições faciais. Conforme os experimentos apresentados nas Tabelas 4.1, 4.2 e 4.3 no Capítulo 4, o método proposto utilizando feições faciais esparsas coloridas supera o método OLPP típico (CAI et al., 2006) e métodos de reconhecimento facial representativos do estado da arte usando as mesmas feições faciais esparsas extraídas.

Também essa tese contribui com a proposta de um novo método de extração de feições faciais densas coloridas (conforme Seção 3.3), que tem o objetivo de preservar a informação de cor nas imagens faciais durante a extração de feições faciais densas. Conforme os experimentos apresentados nas Tabelas 4.4, 4.5 e 4.6 no Capítulo 4, o método proposto utilizando feições faciais densas coloridas supera o método OLPP típico (CAI et al., 2006) e métodos de reconhecimento facial representativos do estado da arte usando as mesmas feições faciais densas extraídas.

Independentemente da aplicação de representações faciais esparsas ou densas, são obtidos vetores de feições faciais de alta dimensionalidade como resultado do processo de extração de feições faciais. Como as feições faciais originais de alta dimensionalidade não apresentam boa discriminação de classes de dados faciais (indivíduos), o método proposto projeta feições faciais de alta dimensionalidade em espaços de feições de baixa dimensionalidade que preservam informações discriminativas das classes de dados faciais. Esses espaços de feições de baixa dimensionalidade são obtidos com uma modificação proposta nessa tese do método *Projeções Ortogonais Preservadoras de Localidade* (*Orthogonal Locality Preserving Projections* - OLPP) que utiliza novos esquemas de definição de localidade (que é quando duas representações faciais são vizinhas no espaço de feições),

propostos nessa tese como detalhado nas Seções 3.4 e 3.5.

Quatro opções para customizar a localidade dos dados faciais foram propostas: OLPP1 (CAI et al., 2006), OLPP2 (SOLDERA; BEHAINE; SCHARCANSKI, 2015), OLPP3 e OLPP4. Os resultados experimentais mostram que a estrutura das classes de dados faciais é melhor preservada usando o esquema supervisionado OLPP4 ao invés do esquema não supervisionado OLPP3. De forma semelhante, é mostrado também que o esquema supervisionado OLPP2 leva a uma maior separabilidade de classes de dados faciais do que o esquema não supervisionado OLPP1 (que é usado pelo método OLPP típico (CAI et al., 2006)). Isso sugere que preservar a estrutura das classes de dados faciais no espaço de baixa dimensionalidade é mais útil que preservar a estrutura dos dados faciais (SOLDERA; BEHAINE; SCHARCANSKI, 2015) (KOKIOPOULOU; SAAD, 2009).

Neste trabalho também foram investigadas técnicas de extração de feições faciais esparsas de imagens faciais com o objetivo de criar topologias de *landmarks* para serem usadas em reconhecimento facial como alternativa a outros métodos de extração de feições faciais disponíveis na literatura. Uma vez que se tenham feições faciais de melhor qualidade mapeadas para um espaço de baixa dimensionalidade obtém-se uma maior discriminação das classes de dados faciais que outros métodos (como o OLPP típico (CAI et al., 2006)), facilitando classificar novas imagens faciais de forma robusta a ruído e dados atípicos pelo uso de SVM com margens suaves (SOLDERA; BEHAINE; SCHARCANSKI, 2015).

Além disso, o método de reconhecimento facial proposto usando o esquema supervisionado de localidade OLPP4 e SVM com margens suaves atinge taxas de reconhecimento maiores do que quando são usados os outros esquemas de localidade tais como OLPP3, OLPP2 e OLPP1, e inclusive apresentando melhores resultados que métodos representativos do estado da arte e o método OLPP típico (CAI et al., 2006). Isto sugere que o método proposto consegue lidar melhor com problemas não separáveis linearmente. Os resultados também sugerem que o esquema de definição de localidade supervisionado OLPP4 (Seção 3.4), e SVM com margens suaves ajuda a preservar a estrutura não-linear inerente aos dados faciais melhor do que os outros métodos mencionados.

Deve ser observado que em todos os experimentos, o esquema supervisionado de definição de localidade OLPP4 apresentou taxas de reconhecimento maiores que o esquema não supervisionado OLPP3, e que o esquema de definição de localidade supervisionado OLPP2 apresentou taxas de reconhecimento maiores que o esquema não supervisionado OLPP1. Os resultados sugerem que preservar a estrutura das classes de dados faciais no espaço de baixa dimensionalidade (abordagem supervisionada) é mais discriminativo que preservar a estrutura dos dados faciais (abordagem não supervisionada). Além disso, o esquema de definição de localidade supervisionado OLPP4 obteve taxas de reconhecimento maiores que o esquema de definição de localidade supervisionado OLPP2 e o esquema de definição de localidade não supervisionado OLPP3 ofereceu taxas de reconhecimento maiores que o esquema de definição de localidade não supervisionado OLPP1, sugerindo que usando pesos iguais (afinidades) na definição da localidade dos dados ajuda a preservar melhor a estrutura das classes de dados faciais, causando um impacto positivo nas taxas de reconhecimento facial, tanto na abordagem supervisionada como na não supervisionada.

O método proposto de reconhecimento facial usando tanto representações esparsas como densas de imagens faciais integrado ao esquema proposto de redução de dimensionalidade é robusto a ruído e dados atípicos, pois novas imagens faciais são classificadas usando SVM com margens suaves no espaço de baixa dimensionalidade obtido, supe-

rando o classificador do vizinho mais próximo comumente usado na abordagem baseada na aparência (CAI et al., 2006) (TURK; PENTLAND, 1991) (HE et al., 2005), conforme demonstrado pelos experimentos no Capítulo 4.

Os resultados experimentais obtidos são baseados em 5 bases de imagens faciais públicas, a PUT, a FEI, a FERET, a Yale e a ORL. O método proposto foi aplicado a representações faciais esparsas e densas de imagens faciais (em escalas de cinza e coloridas), sob várias condições encontradas na prática, envolvendo mudanças na pose da cabeça, expressão facial, iluminação e incluindo oclusões. O método proposto usando o esquema de definição de localidade OLPP4 e SVM treinado com margens suaves pode potencialmente chegar a taxas de reconhecimento maiores que o método OLPP típico (CAI et al., 2006) e outros métodos representativos do estado da arte disponíveis na literatura.

Tanto para representações de imagens faciais densas como para esparsas, verificou-se que a extração de feições faciais a partir de imagens coloridas levou a melhores resultados em comparação a quando são usadas as mesmas imagens faciais, porém convertidas para escalas de cinza, ajudando a confirmar uma tendência de que a informação de cor ajuda a aumentar a discriminação das classes de dados faciais (indivíduos) no espaço de baixa dimensionalidade (BEHAINE; SCHARCANSKI, 2012) (SOLDERA; BEHAINE; SCHARCANSKI, 2015) (KOKIOPOULOU; SAAD, 2009).

Uma parte do método proposto desenvolvido nessa tese de doutorado foi publicada no *journal* internacional *IEEE Transactions on Instrumentation and Measurement*, correspondendo à abordagem de reconhecimento facial proposta usando feições esparsas, e a parte restante referente à abordagem de reconhecimento facial proposta usando feições densas foi submetida para o *journal Engineering Applications of Artificial Intelligence*, além da criação de um relatório técnico.

Futuramente, temos a intenção de executar mais testes com o *framework* proposto, investigando novas métricas de similaridades de faces e como elas podem melhorar o reconhecimento facial no nosso esquema de projeção de espaços discriminativos de baixa dimensionalidade. É possível investigar mais a representação esparsa para determinar o tamanho de janela e quantidade de *landmarks* interpolados automaticamente de acordo com a resolução apresentada pelas imagens faciais. Podem ser estudadas técnicas de redução de dimensionalidade não-linear a fim de preservar a estrutura não-linear inerente aos dados faciais. Além disso, os estudos apresentados podem ser estendidos a outras bases de imagens faciais. O site "<http://www.face-rec.org/databases/>" oferece uma listagem de bases adicionais que poderiam ser usadas em trabalhos futuros. Prova de vida é outra área que pode ser estudada a fim de detectar tentativas de burlar o método proposto.

REFERÊNCIAS

- ABSIL, P. A.; MAHONY, R.; SEPULCHRE, R. Riemannian Geometry of Grassmann Manifolds with a View on Algorithmic Computation. **Acta Applicandae Mathematicae**, [S.l.], v.80, n.2, p.199–220, Janeiro 2004.
- ANDRZEJ Kasiński ANDRZEJ FLOREK, A. S. The PUT face database. **Image Processing & Communication**, [S.l.], v.13, n.3, p.59–64, 2008.
- BAUDAT, G.; ANOUAR, F. Generalized discriminant analysis using a kernel approach. **Neural Computation**, [S.l.], v.12, n.10, p.2385–2404, Outubro 2000.
- BEHAINE, C. A. R.; SCHARCANSKI, J. Enhancing the performance of active shape models in face recognition applications. **IEEE Transactions on Instrumentation and Measurement**, [S.l.], v.61, n.8, p.2330–2333, Agosto 2012.
- BELHUMEUR, P. N.; HESPANHA, J. a. P.; KRIEGMAN, D. J. Eigenfaces vs. Fisherfaces: recognition using class specific linear projection. **IEEE transactions on Pattern Analysis and Machine Intelligence**, [S.l.], v.19, n.7, p.711–720, Julho 1997.
- BETTA, G.; CAPRIGLIONE, D.; CORVINO, M.; LIGUORI, C.; PAOLILLO, A. Face based recognition algorithms: a first step toward a metrological characterization. **IEEE Transactions on Instrumentation and Measurement**, [S.l.], v.62, n.5, p.1008–1016, Maio 2013.
- BOSER, B. E.; GUYON, I. M.; VAPNIK, V. N. A Training Algorithm for Optimal Margin Classifiers. In: ANNUAL ACM WORKSHOP ON COMPUTATIONAL LEARNING THEORY, 5., 1992. **Proceedings...** ACM Press, 1992. p.144–152.
- BURGOS-ARTIZZU, X. P.; PERONA, P.; DOLLAR, P. Robust Face Landmark Estimation under Occlusion. In: THE IEEE INTERNATIONAL CONFERENCE ON COMPUTER VISION (ICCV), 2013. **Anais...** [S.l.: s.n.], 2013. p.1513–1520.
- CAI, D.; HE, X.; HAN, J. Speed up kernel discriminant analysis. **The VLDB Journal**, [S.l.], v.20, n.1, p.21–33, Fevereiro 2011.
- CAI, D.; HE, X.; HAN, J.; ZHANG, H.-J. Orthogonal laplacianfaces for face recognition. **IEEE Transactions on Image Processing**, [S.l.], v.15, n.11, p.3608–3614, Novembro 2006.
- CHANG, C.-C.; LIN, C.-J. LIBSVM: a library for support vector machines. **ACM Transactions on Intelligent Systems and Technology**, [S.l.], v.2, p.27:1–27:27, 2011.

CHOU, C.-T.; SHIH, S.-W.; CHEN, W.-S.; CHENG, V.; CHEN, D.-Y. Non-orthogonal view iris recognition system. **IEEE Transactions on Circuits and Systems for Video Technology**, [S.l.], v.20, n.3, p.417–430, Março 2010.

CHUNG, F. R. K. **Spectral Graph Theory**. [S.l.]: American Mathematical Society, 1997.

COOTES, T. F.; TAYLOR, C. J.; COOPER, D. H.; GRAHAM, J. Active shape models - Their training and application. **Computer Vision and Image Understanding**, [S.l.], v.61, n.1, p.38–59, Janeiro 1995.

CORTES, C.; VAPNIK, V. Support-vector network. **Machine Learning**, [S.l.], n.20, p.273–297, 1995.

COVER, T.; HART, P. Nearest neighbor pattern classification. **IEEE Transactions on Information Theory**, [S.l.], v.13, n.1, p.21–27, Janeiro 1967.

DARDAS, N. H.; GEORGANAS, N. D. Real-time hand gesture detection and recognition using bag-of-features and support vector machine techniques. **IEEE Transactions on Instrumentation and Measurement**, [S.l.], v.60, n.11, p.3592–3607, Novembro 2011.

DEVROYE, L.; GYÖRFI, L.; LUGOSI, G. **A Probabilistic theory of pattern recognition**. [S.l.]: Springer, 1996.

EFRATY, B.; PAPADAKIS, M.; PROFITT, A.; SHAH, S.; KAKADIARIS, I. Pose invariant facial component-landmark detection. In: IEEE INTERNATIONAL CONFERENCE ON IMAGE PROCESSING (ICIP), 18., 2011. **Anais...** [S.l.: s.n.], 2011. p.569–572.

FAN, R.-E.; CHEN, P.-H.; LIN, C.-J. Working Set selection using second order information for training support vector machines. **The Journal of Machine Learning Research**, [S.l.], v.6, p.1889–1918, Dezembro 2005.

GONZALEZ, R. C.; WOODS, R. E. **Digital Image Processing**. 2nd.ed. Boston, MA, USA: Addison-Wesley Longman Publishing Co., Inc., 2001.

GUO, G.; MU, G.; FU, Y.; DYER, C.; HUANG, T. A study on automatic age estimation using a large database. In: IEEE 12TH INTERNATIONAL CONFERENCE ON COMPUTER VISION, ICCV 2009, 2009. **Anais...** [S.l.: s.n.], 2009. p.1986–1991.

GUO, G.; WANG, X. Kinship measurement on salient facial features. **IEEE Transactions on Instrumentation and Measurement**, [S.l.], v.61, n.8, p.2322–2325, Agosto 2012.

HE, X.; YAN, S.; HU, Y.; NIYOGI, P.; ZHANG, H. jiang. Face recognition using laplacianfaces. **IEEE transactions on Pattern Analysis and Machine Intelligence**, [S.l.], v.27, n.3, p.328–340, Março 2005.

HE, Z.; SHEN, Y.; WANG, Q.; WANG, Y. Optimized ensemble EMD-based spectral features for hyperspectral image classification. **IEEE Transactions on Instrumentation and Measurement**, [S.l.], v.63, n.5, p.1041–1056, Maio 2014.

JAFRI, R.; ARABNIA, H. R. A survey of face recognition techniques. **Journal of Information Processing Systems**, [S.l.], v.5, n.2, p.41–68, Junho 2009.

JIA, H.; MARTINEZ, A. Support Vector Machines in face recognition with occlusions. In: IEEE CONFERENCE ON COMPUTER VISION AND PATTERN RECOGNITION (CVPR), 2009. **Anais...** [S.l.: s.n.], 2009. p.136–141.

KAN, M.; SHAN, S.; ZHANG, H.; LAO, S.; CHEN, X. Multi-view Discriminant Analysis. In: EUROPEAN CONFERENCE ON COMPUTER VISION, 12., 2012. **Anais...** [S.l.: s.n.], 2012. v.7572, p.808–821.

KARIMI, B.; DEVROYE, L. A Study on Significance of Color in Face Recognition using Several Eigenface Algorithms. In: CANADIAN CONFERENCE ON ELECTRICAL AND COMPUTER ENGINEERING, 2007. **Anais...** [S.l.: s.n.], 2007. p.1309–1312.

KOKIOPOULOU, E.; SAAD, Y. Orthogonal neighborhood preserving projections: a projection-based dimensionality reduction technique. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, [S.l.], v.29, n.12, p.2143–2156, Dezembro 2007.

KOKIOPOULOU, E.; SAAD, Y. Enhanced graph-based dimensionality reduction with repulsion Laplaceans. **Pattern Recognition**, [S.l.], v.42, n.11, p.2392–2402, 2009.

KRESSEL, U. H.-G. Pairwise Classification and Support Vector Machines. In: **Advances in Kernel Methods**. Cambridge, MA, USA: MIT Press, 1999. p.255–268.

LANGNER, O.; DOTSCHE, R.; BIJLSTRA, G.; WIGBOLDUS, D. H. J.; HAWK, S. T.; KNIPPENBERG, A. van. Presentation and validation of the Radboud Faces Database. **Cognition & Emotion**, [S.l.], v.24, p.1377–1388, 2010.

LEE, D.; CHOI, K.; CHOI, H.; KIM, J. Recognizable-image selection for fingerprint recognition with a mobile-device camera. **IEEE Transactions on Systems, Man, and Cybernetics, Part B**, [S.l.], v.38, n.1, p.233–243, Fevereiro 2008.

LU, J.; HU, J.; ZHOU, X.; SHANG, Y.; TAN, Y.-P.; WANG, G. Neighborhood repulsed metric learning for kinship verification. In: IEEE CONFERENCE ON COMPUTER VISION AND PATTERN RECOGNITION (CVPR), 2012. **Anais...** [S.l.: s.n.], 2012. p.2594–2601.

MIKA, S.; RATSCH, G.; WESTON, J.; SCHOLKOPF, B.; MULLER, K. Fisher discriminant analysis with kernels. In: IEEE SIGNAL PROCESSING SOCIETY WORKSHOP. NEURAL NETWORKS FOR SIGNAL PROCESSING IX, 1999., 1999. **Proceedings...** [S.l.: s.n.], 1999. p.41–48.

NASTAR, C.; MITSCHKE, M. Real-time face recognition using feature combination. In: IEEE INTERNATIONAL CONFERENCE ON AUTOMATIC FACE AND GESTURE RECOGNITION, 1998. **Anais...** [S.l.: s.n.], 1998. p.312–317.

PARK, B. U.; MARRON, J. S. Comparison of Data-Driven Bandwidth Selectors. **Journal of the American Statistical Association**, [S.l.], v.85, n.409, p.66–72, Março 1990.

PHAM, D.-S.; VENKATESH, S. Supervised subspace learning with multi-class lagrangian SVM on the grassmann manifold. In: AUSTRALASIAN CONFERENCE ON ARTIFICIAL INTELLIGENCE, 2011. **Anais...** [S.l.: s.n.], 2011. v.7106, p.241–250.

PHILLIPS, P. J.; WECHSLER, H.; HUANG, J.; RAUSS, P. The FERET database and evaluation procedure for face recognition algorithms. **Image and Vision Computing**, [S.l.], v.16, n.5, p.295–306, Abril 1998.

RAHULKAR, A.; HOLAMBE, R. Half-iris feature extraction and recognition using a new class of biorthogonal triplet half-band filter bank and flexible k-out-of-n: a postclassifier. **IEEE Transactions on Information Forensics and Security**, [S.l.], v.7, n.1, p.230–240, Fevereiro 2012.

RAPP, V.; SENECHAL, T.; BAILLY, K.; PREVOST, L. Multiple kernel learning SVM and statistical validation for facial landmark detection. In: IEEE INTERNATIONAL CONFERENCE ON AUTOMATIC FACE AND GESTURE RECOGNITION (FG2011), 9., 2011. **Anais...** [S.l.: s.n.], 2011. p.265–271.

REN, Z.; YUAN, J.; MENG, J.; ZHANG, Z. Robust part-based hand gesture recognition using kinect sensor. **IEEE Transactions on Multimedia**, [S.l.], v.15, n.5, p.1110–1120, Agosto 2013.

SAMARIA, F.; HARTEK, A. Parameterisation of a stochastic model for human face identification. In: IEEE WORKSHOP ON APPLICATION OF COMPUTER VISION, 1994. **Proceedings...** [S.l.: s.n.], 1994. p.138–142.

SCHARCANSKI, J.; PROENÇA, H.; DU, E. **Signal and Image Processing for Biometrics**. [S.l.]: Springer, 2014. (Lecture Notes in Electrical Engineering).

SCHOLKOPF, B.; SMOLA, A. J. **Learning with kernels: support vector machines, regularization, optimization, and beyond**. Cambridge, MA, USA: MIT Press, 2001.

SHAOGANG GONG, S. M.; PSARROU, A. (Ed.). **Dynamic Vision; From Images to Face Recognition**. [S.l.]: Imperial College Press, 2000.

SIROVICH, L.; KIRBY, M. Low-Dimensional Procedure for the Characterization of Human Faces. **Journal of the Optical Society of America A**, [S.l.], v.4, n.3, p.519–524, 1987.

SOLDERA, J.; BEHAINE, C. A. R.; SCHARCANSKI, J. Customized orthogonal locality preserving projections with soft margin maximization for face recognition. **IEEE Transactions on Instrumentation and Measurement**, [S.l.], v.64, n.9, p.2417–2426, Setembro 2015.

SUKNO, F.; ORDAS, S.; BUTAKOFF, C.; CRUZ, S.; FRANGI, A. Active shape models with invariant optimal features: application to facial analysis. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, [S.l.], v.29, n.7, p.1105–1117, Julho 2007.

THOMAZ, C. E.; GIRALDI, G. A. A new ranking method for principal components analysis and its application to face image analysis. **Image and Vision Computing**, [S.l.], v.28, p.902–913, Junho 2010.

TIPPING, M. E. Sparse bayesian learning and the relevance vector machine. **The Journal of Machine Learning Research**, [S.l.], v.1, p.211–244, Setembro 2001.

TURK, M.; PENTLAND, A. Eigenfaces for recognition. **Journal of Cognitive Neuroscience**, [S.l.], v.3, n.1, p.71–86, Janeiro 1991.

VALIANT, L. G. A Theory of the Learnable. **Communications of the ACM**, [S.l.], v.27, n.11, p.1134–1142, Novembro 1984.

VAPNIK, V. N. **Statistical Learning Theory**. Ney York: Wiley-Interscience, 1998.

VERDOOLAEGE, G.; SOLDERA, J.; MACEDO, T.; SCHARCANSKI, J. Data and Information Dimensionality in Non-cooperative Face Recognition. In: SCHARCANSKI, J.; PROENÇA, H.; DU, E. (Ed.). **Signal and Image Processing for Biometrics**. [S.l.]: Springer, 2014. p.1–35. (Lecture Notes in Electrical Engineering, v.292).

VINH, N. X.; EPPS, J.; BAILEY, J. Information theoretic measures for clusterings comparison: variants, properties, normalization and correction for chance. **The Journal of Machine Learning Research**, [S.l.], v.11, p.2837–2854, Dezembro 2010.

WU, T.; TURAGA, P.; CHELLAPPA, R. Age estimation and face verification across aging using landmarks. **IEEE Transactions on Information Forensics and Security**, [S.l.], v.7, n.6, p.1780–1788, Dezembro 2012.

ZHANG, D.; LIU, F.; ZHAO, Q.; LU, G.; LUO, N. Selecting a reference high resolution for fingerprint recognition using minutiae and pores. **IEEE Transactions on Instrumentation and Measurement**, [S.l.], v.60, n.3, p.863–871, Março 2011.

ZHAO, X.; CHAI, X.; NIU, Z.; HENG, C.; SHAN, S. Context constrained facial landmark localization based on discontinuous Haar-like feature. In: IEEE INTERNATIONAL CONFERENCE ON AUTOMATIC FACE AND GESTURE RECOGNITION (FG2011), 9., 2011. **Anais...** [S.l.: s.n.], 2011. p.673–678.