

UNIVERSIDADE FEDERAL DE SÃO CARLOS

CENTRO DE CIÊNCIAS EXATAS E DE TECNOLOGIA

PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA COMPUTAÇÃO

DEPARTAMENTO DE COMPUTAÇÃO

COMBINAÇÃO DE MÚLTIPLOS CLASSIFICADORES
PARA RECONHECIMENTO DE FACE HUMANA

DENIS HENRIQUE PINHEIRO SALVADEO

SÃO CARLOS

JULHO / 2009

**Ficha catalográfica elaborada pelo DePT da
Biblioteca Comunitária da UFSCar**

S182cm

Salvadeo, Denis Henrique Pinheiro.
Combinação de múltiplos classificadores para
reconhecimento de face humana / Denis Henrique Pinheiro
Salvadeo. -- São Carlos : UFSCar, 2009.
134 f.

Dissertação (Mestrado) -- Universidade Federal de São
Carlos, 2009.

1. Reconhecimento de padrões. 2. Reconhecimento da
face humana (Ciência da computação). 3. Combinação de
classificadores. 4. Extração de Atributos. 5. Regularização
de matriz de covariância. 6. Análise de componentes
principais. I. Título.

CDD: 006.4 (20^a)

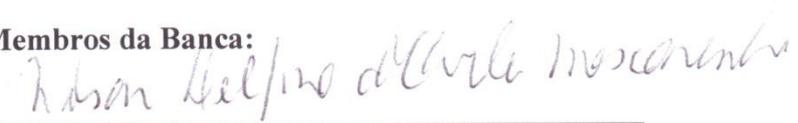
Universidade Federal de São Carlos
Centro de Ciências Exatas e de Tecnologia
Programa de Pós-Graduação em Ciência da Computação

“Combinação de Múltiplos Classificadores para
Reconhecimento de Face Humana”

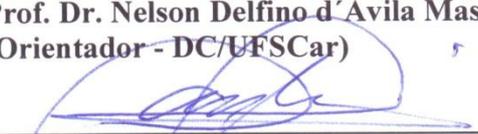
DENIS HENRIQUE PINHEIRO SALVADEO

Dissertação de Mestrado apresentada ao
Programa de Pós-Graduação em Ciência da
Computação da Universidade Federal de São
Carlos, como parte dos requisitos para a
obtenção do título de Mestre em Ciência da
Computação

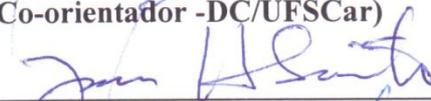
Membros da Banca:



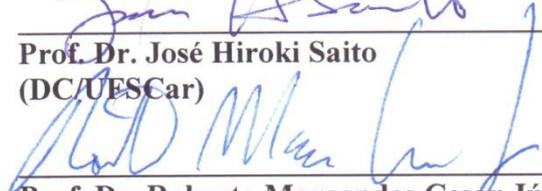
Prof. Dr. Nelson Delfino d'Ávila Mascarenhas
(Orientador - DC/UFSCar)



Prof. Dr. Jander Moreira
(Co-orientador -DC/UFSCar)



Prof. Dr. José Hiroki Saito
(DC/UFSCar)



Prof. Dr. Roberto Marcondes Cesar Júnior
(IME/USP)

São Carlos
Julho/2009

AGRADECIMENTOS

A Deus por permitir que esse projeto fosse concluído, fornecendo a luz necessária nos momentos de escuridão.

Aos meus pais e ao meu irmão pelo amor irrestrito, atenção, apoio, incentivo e formação durante toda a jornada da minha vida pessoal e profissional.

Aos meus avós maternos, Alice (em memória) e Osório, pelos momentos de sabedoria, humildade e riqueza de união familiar.

À minha noiva Juliana pelos momentos de carinho, paixão, amor, inspiração, compreensão e felicidade.

Aos meus orientadores, prof. Nelson e prof. Jander, pelas valiosas discussões, amizade, orientação, confiança, incentivo, acessibilidade e partilha do conhecimento e experiência adquiridos durante suas vidas.

Aos membros do GAPIS, especialmente ao Alexandre, Débora, Michelle, Ana Luísa, Eduardo, Moacir, Zorzan, João Fernando, Leonardo, Talita, Murillo, Gustavo e Márcia pelas colaborações neste e em outros projetos, mas principalmente, por serem verdadeiros e grandes amigos.

À professora Sandra Abib pela sua dedicação à UFSCar e principalmente, pela amizade, incentivo e oportunidades a mim proporcionadas e confiadas.

Aos meus amigos da república, Luciano, Sílvio, Leonardo, Francis, Fábio, Chohfi e Lucas pelos momentos de descontração e convivência.

Aos docentes e funcionários do DC e demais alunos do PPG-CC e graduação, pelo incentivo, aprendizado e momentos compartilhados.

Ao CNPq pela ajuda financeira tão importante e necessária durante o desenvolvimento deste trabalho.

Enfim, a todas as pessoas que de alguma forma contribuíram para a realização e conclusão do meu mestrado.

*“We judge ourselves by what we feel capable of doing, while
others judge us by what we have already done.”*

(Henry Wadsworth Longfellow)

RESUMO

Nas últimas décadas, o objeto face humana tem sido muito explorado graças ao advento dos sistemas envolvendo biometria, especialmente para aplicações nas áreas de segurança. Uma das aplicações mais desafiadoras é o problema de reconhecimento de face humana, que consiste em determinar a qual indivíduo em um banco de dados de pessoas conhecidas uma imagem de face corresponde. O processo de reconhecimento de face consiste basicamente em duas etapas: extração de atributos e classificação. Na literatura de reconhecimento de faces, diversas técnicas foram utilizadas, podendo ser divididas em técnicas holísticas (extração de atributos implícita), técnicas baseadas em características (extração de atributos explícita) e técnicas híbridas (envolvem as duas anteriores). Em maior número na literatura, as técnicas holísticas demonstraram ser mais eficientes e envolvem em geral, métodos de reconhecimento de padrões estatísticos como *Principal Component Analysis* (PCA), *Independent Component Analysis* (ICA), *Linear Discriminant Analysis* (LDA), *Support Vector Machines* (SVM), além de Redes Neurais. Para problemas como o reconhecimento de face humana em imagens digitais, um ponto determinante é a capacidade de generalização. Por esse problema ser complexo de ser resolvido, devido à alta dimensionalidade dos dados e ao número pequeno de amostras por pessoa, utilizar um único classificador tornaria reduzida a capacidade de reconhecimento, já que é difícil projetar um único classificador nestas condições que capture todas as variações que formam o espaço de faces humanas. Dessa forma, este trabalho se propôs à investigação de combinação de múltiplos classificadores aplicados ao problema do reconhecimento de face, obtendo um novo esquema para solução do mesmo, variando a extração de atributos com PCA e algumas variantes e LDA, classificadores K Vizinhos Mais Próximos e Máxima Verossimilhança e diversos combinadores treináveis ou não. Ainda, para amenizar o problema de poucas amostras, uma técnica de regularização de matrizes de covariância foi utilizada. Finalmente, para avaliação de desempenho utilizou-se *Holdout* e *Resubstitution* para particionar o conjunto de dados, assim como o coeficiente Kappa e testes de significância Z e T foram utilizados para medir o desempenho do esquema proposto. Dos experimentos concluiu-se que os melhores subesquemas foram o RBPCA/MaxVer-PCA/NN-KL5/NN → Votação por Maioria para o banco de dados ORL e o RLDA+RPCA/MaxVer-KL4/NN-KL5/NN → Soma para o banco de dados AR com coeficientes Kappa 0,956 (médio) e 0,839, respectivamente. Além disso, foi determinado que estes subesquemas são robustos à pose (ORL), iluminação e à pequenas variações de expressão facial, mas sofreram influência de oclusões (AR).

ABSTRACT

Lately, the human face object has been exploited by the advent of systems involving biometrics, especially for applications in security. One of the most challenging applications is the problem of human face recognition, which consists of determining the correspondence between an input face and an individual from a database of known persons. The process of face recognition consists of two steps: feature extraction and classification. In the literature of face recognition, different techniques have been used, and they can be divided into holistic techniques (implicit feature extraction), feature-based techniques (explicit feature extraction) and hybrid techniques (involving the two previous). In many articles, holistic techniques have proved to be most efficient and generally they involve methods of statistical pattern recognition as Principal Component Analysis (PCA), Independent Component Analysis (ICA), Linear Discriminant Analysis (LDA), Support Vector Machines (SVM), and Neural Networks. For problems such as human face recognition in digital images, a crucial point is the ability to generalize. The solution for this problem is complex due to the high dimensionality of data and the small number of samples per person. Using a single classifier would reduce the ability of recognition, since it is difficult to design a single classifier in these conditions that capture all variability that span the human faces spaces. Thus, this work proposes to investigate the combination of multiple classifiers applied to the problem of face recognition, defining a new scheme to resolve this problem, varying the feature extraction with PCA and some its variants and LDA, K-Nearest Neighbor (K-NN) and Maximum Likelihood (MaxLike) classifiers and several trainable or not trainable methods for combining classifiers. Still, to mitigate the problem of small sample size (SSS), a technique for regularizing the covariance matrix was used. Finally, to assess the classification performance, Holdout and Resubstitution methods were used to partition the data set and the Kappa coefficient and Z and T statistics were used to measure the performance of the proposed scheme. From the experiments it was concluded that the best sub-schemes were the RBPCA/MaxLike-PCA/NN-KL5/NN classifiers combined by the Majorite Vote Rule for the ORL database and the RLDA+RPCA/MaxLike-KL4/NN-KL5/NN classifiers combined by the Sum Rule for the AR database, obtaining Kappa coefficients of 0.956 (mean) and 0.839, respectively. Besides that, it has been determined that these sub-schemes are robust to pose (ORL database), illumination and small change of the facial expression, but they were affected by occlusions (AR database).

LISTA DE FIGURAS

Figura 2.1. Divisão da Abordagem de Reconhecimento de Padrões Estatísticos. Adaptado de (JAIN, DUIN <i>et al.</i> , 2000).....	18
Figura 2.2. Limitação da Transformação de Karhunen-Loève quanto à separabilidade entre classes. Retirado de (THEODORIDIS e KOUTROUMBAS, 2003).	23
Figura 2.3. Ilustração da Transformada de Karhunen-Loève sobre o caso Gaussiano Bidimensional. Retirado de (FUKUNAGA, 1990).	24
Figura 2.4. Ilustração do processo aplicado no <i>Block-based PCA</i> . Nesse exemplo, a imagem final contém o mesmo número de pixels quanto o número de blocos da imagem, o que implica que apenas o componente de maior variância foi selecionado.	27
Figura 2.5. Exemplo de imagem real com aplicação do extrator de atributos <i>Block-based PCA</i> . Neste exemplo foi considerado blocos 8x8. No lado esquerdo da seta temos a imagem original, enquanto no lado direito temos a imagem reconstruída. Somente para facilitar a visualização da imagem reconstruída ela foi ampliada ao seu lado.	27
Figura 2.6. Abordagens para a construção de <i>ensembles</i> em combinação de classificadores. Adaptado de (KUNCHEVA, 2004).....	31
Figura 2.7. Tendência das probabilidades de erro dos métodos Leave-One-Out e Resubstitution no caso assintótico. Retirado de (THEODORIDIS e KOUTROUMBAS, 2003).	40
Figura 2.8. Ilustração de uma matriz de confusão. Adaptado de (CONGALTON, 1999).....	41
Figura 3.1. Sistema Completo de Reconhecimento de Faces. Adaptado de (LI e JAIN, 2005).	48
Figura 3.2. O subespaço de faces: a) num contexto de espaço de imagens e b) a sua divisão interna com subespaços de indivíduos diferentes. Adaptado de (LI e JAIN, 2005).....	49
Figura 3.3. Desafios em sistemas de reconhecimento de faces do ponto de vista de subespaços de faces: a) Distância euclidiana não é capaz de diferenciar entre dois indivíduos, pois algumas distâncias interpessoais são menores que intrapessoais e b) O subespaço aprendido não é suficiente para cobrir todo o subespaço de um indivíduo, ou seja, representa a incapacidade de generalização. Adaptado de (LI e JAIN, 2005).	49

Figura 3.4. As possíveis orientações de uma face que caracterizam mudanças na pose. Adaptado de (JENG, LIAO <i>et al.</i> , 1998).....	50
Figura 3.5. Algumas amostras de faces frontais do banco de dados FERET: Fa, Fb, Fc, Duplicate I e Duplicate II são nomes de algumas das principais categorias de divisão do FERET. Retirado de (GROSS, 2005).....	52
Figura 3.6. Algumas amostras de um sujeito do banco de dados AR. As duas primeiras linhas representam a primeira seção e as restantes a segunda seção. Retirado de (MARTÍNEZ e BENAVENTE, 1998).....	53
Figura 3.7. Algumas amostras de um sujeito do banco de dados ORL. Retirado de (ZUO, WANG <i>et al.</i> , 2007).	53
Figura 3.8. Algumas amostras de um sujeito do banco de dados CMU PIE. Essas amostras foram recortadas a fim de eliminar o fundo e conter somente a face do indivíduo. Retirado de (HUANG, YUEN <i>et al.</i> , 2006).....	54
Figura 3.9. Algumas amostras de um sujeito do banco de dados XM2VTS. Retirado de (MESSER, MATAS <i>et al.</i> , 1999).....	55
Figura 3.10. Algumas amostras de um sujeito do banco de dados Yale. Retirado de < http://cvc.yale.edu/projects/yalefaces/yalefaces.html >.....	55
Figura 3.11. Algumas amostras de um indivíduo do banco de dados YaleB. Retirado de (GEORGHIADES, BELHUMEUR <i>et al.</i> , 2001).....	56
Figura 3.12. Algumas amostras de um sujeito do banco de dados do MIT. Retirado de (MARCIALIS e ROLI, 2004).	57
Figura 3.13. Exemplos de Eigenfaces sobre o banco de dados ORL. Na primeira linha temos a imagem média, na segunda os 7 autovetores correspondendo aos 7 maiores autovalores e na última linha os 7 autovetores correspondendo aos 7 menores autovalores. Retirado de (LU, 2003).....	60
Figura 3.14. Exemplos dos primeiros Fisherfaces sobre o banco de dados ORL. Retirado de (LU, 2003).	60
Figura 3.15. Exemplo de alguns vetores base ICA gerados sobre o banco de dados ORL. Retirado de (LU, 2003).....	61
Figura 3.16. Kernels de autocorrelação de zero a segunda ordem. Adaptado de (ARBUCKLE, LANGE <i>et al.</i> , 1995).	63

Figura 3.17. Exemplos de imagens no banco de dados: a) indivíduos diferentes; b) mesmo indivíduo. Extraído de (ARBUCKLE, LANGE <i>et al.</i> , 1995).....	66
Figura 3.18. Modelo Esquerda-Direita de Face Humana Linear, onde aij representa a probabilidade de transição do estado i ao j e bi representa a probabilidade de emissão de saída num estado i . Extraído de (ACHERMANN e BUNKE, 1996a).....	67
Figura 3.19. Exemplos de um indivíduo no banco de dados. Adaptado de (ACHERMANN e BUNKE, 1996a).	68
Figura 3.20. Exemplo de faces geradas artificialmente usando uma rede LVQ. Retirado de (TOLBA e ABU-REZQ, 2000).	69
Figura 3.21. Metodologia de fusão proposta por Marcialis e Roli. Adaptado de (MARCIALIS e ROLI, 2002).....	70
Figura 3.22. Exemplo de distribuição do conjunto de treinamento e geração de amostras virtuais. Retirado de (RYU e OH, 2002).	72
Figura 3.23. Arquitetura da ERN. Retirado de (RYU e OH, 2002).	73
Figura 3.24. Algumas amostras de um indivíduo do banco de dados NUST603. Retirado de (JING e ZHANG, 2003).	75
Figura 3.25. Exemplos de imagens faciais na base de dados: a) ORL, b) Yale, c) AR e d) NLPR+MSU. Retirado de (LU, WANG <i>et al.</i> , 2003).	75
Figura 3.26. Classificação de face usando o LLDA. A distribuição original dos dados e a distribuição transformada são exibidas do lado esquerdo e direito, respectivamente. Retirado de (KIM e KITTLER, 2004).....	76
Figura 3.27. Geração de Visão Virtual Rotacionada usando LUT 3D. Adaptado de (KIM e KITTLER, 2004).	77
Figura 3.28. Uma amostra do banco de dados JDL e o pré-processamento executado: a) Imagem Original; b) Imagem Recortada; c) Histograma Equalizado e d) Máscara para Remoção de Cabelo aplicada. Retirado de (ZHANG, SHAN <i>et al.</i> , 2004).....	78
Figura 3.29. Algumas amostras do banco de dados da Universidade de Notre Dame (ND): a) FA-LF; b) FA-LM; c) FB-LF e d) FB-LM. Onde FA e FB indicam expressões neutra e sorrindo, respectivamente e LF indica que a fonte de iluminação é lateral (ambos os lados) e LM que é, além de lateral, também central. Retirado de (CHAWLA e BOWYER, 2005).	79

Figura 3.30. Imagem do banco de dados ORL a) antes e b) depois da etapa de pré-processamento. Retirado de (LUMINI e NANNI, 2005).	81
Figura 3.31. Exemplo dos componentes faciais extraídos: a) <i>bounding box</i> ; b, c) sobrancelhas; d) ponte do nariz; e, f) olhos; g) nariz; h, i) narinas e j) boca. Retirado de (IVANOV, HEISELE <i>et al.</i> , 2004).....	87
Figura 3.32. Exemplos do conjunto de dados usado por Ivanov e seus colegas. Os exemplos das duas primeiras linhas são do conjunto de treinamento e as restantes, do conjunto de teste. Retirado de (IVANOV, HEISELE <i>et al.</i> , 2004).	88
Figura 3.33. O processo de divisão da imagem recortada em 5 segmentos: imagem original, imagem recortada e imagem dividida. Retirado de (TOYGAR e ACAN, 2004a).....	88
Figura 3.34. Algumas amostras de um indivíduo do banco de dados CNL. Retirado de (MU, WATTA <i>et al.</i> , 2005).....	89
Figura 3.35. Ilustração do processo de amostragem de n-tuplas contínuas. Retirado de (LUCAS, 1997b).	91
Figura 3.36. Pontos importantes na face para o reconhecimento pelo método Wavelet de Gabor Local: a) 15 marcos e b) 23 marcos. Retirado de (HUANG, YUEN <i>et al.</i> , 2002; 2004), respectivamente.	92
Figura 3.37. Componentes Faciais extraídos manualmente. Retirado de (KHUWAJA, 2002).	94
Figura 3.38. Alguns exemplos da segmentação inicial para o HMM. Retirado de (LEMIEUX e PARIZEAU, 2003).	95
Figura 3.39. Imagem média e os 4 primeiros <i>Eigenobjects</i> para a) Olho esquerdo e b) o nariz. Retirado de (LEMIEUX e PARIZEAU, 2003).	96
Figura 3.40. Pontos de referência usados para o aprendizado dos componentes. Retirado de (HEISELE, SERRE <i>et al.</i> , 2007).....	98
Figura 3.41. Esquema de Quantização de Características Aleatória (Bit-plane aleatório). Adaptado de (SIRLANTZIS, HOQUE <i>et al.</i> , 2008).	99
Figura 3.42. Efeito da quantização em escala de nível de cinza: a) Imagem Original; b) Quantização Ordenada (Bit-Plane) e c) Quantização de Camada Aleatória (Bit-Plane Aleatório). Adaptado de (SIRLANTZIS, HOQUE <i>et al.</i> , 2008).....	100
Figura 4.1. Diagrama de blocos do sistema inicial proposto.....	112

Figura 4.2. Taxas de Acerto por variação na face dos classificadores individuais usando Holdout sobre o banco de dados AR.....	124
Figura 4.3. Taxas de Acerto por variação na face do melhor método individual, das Eigenfaces e Fisherfaces e dos dois melhores esquemas de combinação usando Holdout sobre o banco de dados AR.....	125

LISTA DE TABELAS

Tabela 2.1. Interpretação subjetiva do coeficiente Kappa sobre o desempenho de um classificador. Adaptado de (LANDIS e KOCH, 1977).	42
Tabela 3.1. Resumo dos métodos de reconhecimento de faces humanas usando combinação de classificadores.	101
Tabela 4.1. Coeficientes Kappa dos classificadores individuais sobre o banco de dados ORL usando Holdout com diferentes conjuntos de treinamento e teste.....	115
Tabela 4.2. Coeficientes Kappa do esquema proposto inicialmente com diferentes combinadores sobre o banco de dados ORL usando Holdout com diferentes conjuntos de treinamento e teste.	116
Tabela 4.3. Coeficientes Kappa dos melhores subesquemas para cada combinador sobre o banco de dados ORL usando Holdout com diferentes conjuntos de treinamento e teste. Para facilitar a visualização consideramos 1 = RBPCA/MaxVer, 2 = RPCA+RLDA/MaxVer, 3 = PCA/NN, 4 = KL2/NN, 5 = KL3/NN, 6 = KL4/NN e 7 = KL5/NN.	116
Tabela 4.4. Coeficientes Kappa do melhor método individual, das Eigenfaces e Fisherfaces e dos dois melhores esquemas de combinação sobre o banco de dados ORL usando Holdout com diferentes conjuntos de treinamento e teste.	117
Tabela 4.5. Variâncias dos coeficientes Kappa dos classificadores individuais sobre o banco de dados ORL usando Holdout com diferentes conjuntos de treinamento e teste.	117
Tabela 4.6. Variâncias dos coeficientes Kappa do esquema proposto inicialmente com diferentes combinadores sobre o banco de dados ORL usando Holdout com diferentes conjuntos de treinamento e teste.	118
Tabela 4.7. Variâncias dos coeficientes Kappa dos melhores subesquemas para cada combinador sobre o banco de dados ORL usando Holdout com diferentes conjuntos de treinamento e teste. Para facilitar a visualização consideramos 1 = RBPCA/MaxVer, 2 = RPCA+RLDA/MaxVer, 3 = PCA/NN, 4 = KL2/NN, 5 = KL3/NN, 6 = KL4/NN e 7 = KL5/NN.	118
Tabela 4.8. Variâncias dos coeficientes Kappa do melhor método individual, das Eigenfaces e Fisherfaces e dos dois melhores esquemas de combinação sobre o banco de dados ORL usando Holdout com diferentes conjuntos de treinamento e teste.....	119

Tabela 4.9. Teste de Significância usando Estatística T sobre o banco de dados ORL.	119
Tabela 4.10. Resultados gerais dos classificadores individuais usando Holdout e Resubstitution sobre o banco de dados AR, onde η é a Taxa de Acerto e σ^2 é a variância do Kappa.....	121
Tabela 4.11. Resultados gerais do esquema proposto inicialmente com diferentes combinadores usando Holdout e Resubstitution sobre o banco de dados AR, onde η é a Taxa de Acerto e σ^2 é a variância do Kappa.	122
Tabela 4.12. Resultados gerais dos melhores subsquemas para cada combinador usando Holdout e Resubstitution sobre o banco de dados AR, onde η é a Taxa de Acerto e σ^2 é a variância do Kappa. Para facilitar a visualização consideramos 1 = RBPCA/MaxVer, 2 = RPCA+RLDA/MaxVer, 3 = PCA/NN, 4 = KL2/NN, 5 = KL3/NN, 6 = KL4/NN e 7 = KL5/NN.....	122
Tabela 4.13. Resultados gerais do melhor método individual, das Eigenfaces e Fisherfaces e dos dois melhores esquemas de combinação usando Holdout e Resubstitution sobre o banco de dados AR, onde η é a Taxa de Acerto e σ^2 é a variância do Kappa.	123
Tabela 4.14. Teste de Significância usando Estatística Z sobre o banco de dados AR.....	123
Tabela 4.15. Tempos aproximados de cada extrator de atributos na etapa de treinamento sobre os bancos de dados de faces AR e ORL.	126

SUMÁRIO

1	INTRODUÇÃO	13
	1.1 Motivação e Objetivos	13
	1.2 Organização do Texto	14
2	RECONHECIMENTO DE PADRÕES.....	16
	2.1 Extração de Atributos	18
	2.1.1 Maldição da Dimensionalidade	19
	2.1.2 Problema do Small Sample Size.....	19
	2.1.2.1 Regularização de Matrizes de Covariância.....	20
	2.1.3 PCA e suas variantes	22
	2.1.3.1 KL2	24
	2.1.3.2 KL3	25
	2.1.3.3 KL4	25
	2.1.3.4 KL5	26
	2.1.4 Block-Based PCA	26
	2.1.5 LDA	27
	2.2 Classificação	28
	2.2.1 Máxima Verossimilhança (MaxVer)	28
	2.2.2 K-Vizinhos Mais Próximos (KNN)	29
	2.3 Combinação de Classificadores	30
	2.3.1 Combinadores Não-Treináveis	32
	2.3.1.1 Regra do Produto	33
	2.3.1.2 Regra da Soma	33
	2.3.1.3 Regra da Soma Ponderada	34
	2.3.1.4 Derivações de Outras Regras de Combinação a partir das Regras da Soma e do Produto.....	34
	2.3.1.5 Regra do Máximo	35

2.3.1.6	Regra do Mínimo	35
2.3.1.7	Regra da Mediana	35
2.3.1.8	Votação por Maioria	35
2.3.2	Combinadores Treináveis	36
2.3.2.1	Decision Templates	36
2.3.2.2	Dempster-Shafer	37
2.4	Validação	38
2.4.1	Técnicas de partição dos dados e Estimação de Erros.....	38
2.4.2	Medidas de Avaliação do Desempenho.....	39
2.4.2.1	Coeficiente Kappa	40
2.4.2.2	Taxa de Reconhecimento Correto	42
2.4.3	Teste de Significância.....	43
2.4.3.1	Estatística Z	43
2.4.3.2	Estatística T	43
2.5	Considerações finais	44
3	RECONHECIMENTO DE FACES.....	46
3.1	Definição do Problema e Sistema Completo de Reconhecimento de Faces.....	47
3.2	Bancos de Dados de Faces.....	51
3.2.1	FERET	51
3.2.2	AR.....	52
3.2.3	ORL	53
3.2.4	CMU PIE	54
3.2.5	XM2VTS	54
3.2.6	Yale.....	55
3.2.7	YaleB	55
3.2.8	MIT	57
3.3	Metodologias de Avaliação FERET e FRVT 2006	57

3.4 Principais Métodos	59
3.4.1 Eigenfaces (PCA)	59
3.4.2 Fisherfaces (LDA)	60
3.4.3 Análise de Componentes Independentes (ICA).....	61
3.4.4 Máquinas de Vetores de Suporte (SVM).....	61
3.5 Combinando Classificadores para Reconhecimento de Faces Frontais Bidimensionais: Uma Visão Geral	62
3.5.1 Técnicas Holísticas	63
3.5.2 Técnicas Baseadas em Componentes	85
3.5.3 Técnicas Híbridas	89
3.5.4 Discussões.....	101
4 DESENVOLVIMENTO DO PROJETO.....	110
4.1 Metodologia	110
4.2 Materiais e Bancos de Dados Utilizados	111
4.3 Avaliação dos Resultados	113
4.4 Experimentos e Resultados.....	113
4.4.1 ORL	113
4.4.2 AR.....	119
4.4.2.1 Experimento 1 – Resultados Gerais.....	120
4.4.2.2 Experimento 2 – Avaliação de Robustez à Iluminação, Oclusão e Expressão.....	123
4.4.3 Custo Computacional.....	124
4.5 Considerações finais	126
5 CONCLUSÕES E TRABALHOS FUTUROS.....	127
5.1 Conclusões	127
5.2 Trabalhos Futuros	128
REFERÊNCIAS	129

1 INTRODUÇÃO

Nas últimas décadas, o objeto face humana tem sido muito explorado graças ao advento dos sistemas envolvendo biometria, especialmente para aplicações nas áreas de segurança. Uma das aplicações mais desafiadoras é o problema de reconhecimento de face humana, que consiste em determinar com qual indivíduo em um banco de dados de pessoas conhecidas uma imagem de face corresponde.

Dessa forma, o presente trabalho discute esses desafios e propõe um novo sistema para reconhecimento de faces usando combinação de classificadores.

Dentre as principais contribuições deste trabalho destacamos um *survey* para reconhecimento de faces usando combinação de classificadores, a utilização de uma nova técnica de regularização de matrizes de covariância das classes e a extração de atributos por outras variantes de PCA não encontrados na literatura de sistemas de reconhecimento de faces e o projeto de dois novos sistemas de classificação denominados RBPCA/MaxVer e RPCA_RLDA/MaxVer. RBPCA/MaxVer consiste em extrair os atributos com Block-based PCA, gerar as matrizes de covariância de cada classe nesse espaço e regularizá-las a fim de potencializar o desempenho de um classificador de Máxima Verossimilhança. Por sua vez, RPCA_RLDA/MaxVer extrai os atributos com PCA, gera as matrizes de covariância de cada classe nesse espaço e as regulariza, e em seguida, estas matrizes de covariância regularizadas são utilizadas para gerar o espaço LDA e finalmente, obter matrizes de covariância de cada classe regularizadas no espaço LDA que serão utilizadas na classificação por Máxima Verossimilhança.

No restante desse capítulo, ainda são apresentadas as motivações para esse trabalho, os objetivos a serem atingidos e a organização deste documento.

1.1 Motivação e Objetivos

O reconhecimento de faces humanas é de vital importância para aplicações em segurança e videoconferência, por exemplo. A face humana, por ser um objeto complexo, com grande espalhamento em um espaço não linear entre faces do mesmo indivíduo que, por sua

vez, está num espaço reduzido (também não linear) que engloba o conjunto de faces de todos os indivíduos (JAIN e LI, 2005), torna inviável a utilização de um único classificador. Portanto, faz-se necessária a utilização de múltiplos classificadores a fim de aumentar o desempenho sem aumentar a complexidade (KUNCHEVA, 2004). Deve-se observar que uma abordagem adequada na combinação de múltiplos classificadores melhora o desempenho em relação a um único classificador individual.

Dessa forma, este trabalho propõe a investigação de combinação de múltiplos classificadores a fim de definir um esquema selecionando classificadores, regras de combinação e técnicas de extração de atributos mais adequadas ao problema de reconhecimento de face através do cálculo de medidas de desempenho como o coeficiente Kappa e testes de significância.

1.2 Organização do Texto

O restante deste documento está dividido em 4 capítulos de forma a permitir o entendimento da área de Reconhecimento de Padrões, da Aplicação de Reconhecimento de Faces Humanas e suas dificuldades, além do trabalho proposto e os seus resultados obtidos.

No Capítulo 2, que aborda a área de Reconhecimento de Padrões, serão apresentados conceitos básicos, técnicas de extração de atributos e alguns dos seus problemas enfrentados devido, principalmente, ao número pequeno e limitado de amostras de treinamento e alta dimensionalidade dos dados. Também serão abordados alguns classificadores e a subárea de combinação de classificadores explicando algumas das técnicas e o que considerar a fim de otimizar o desempenho da combinação.

No Capítulo 3, a aplicação de Reconhecimento de Faces é abordada e apresentada em grandes detalhes. O espaço de faces e as etapas de um sistema completo de reconhecimento de faces são explicados, bem como os bancos de dados mais comuns, os principais métodos para reconhecimento de faces e metodologias para avaliação de desempenho. Encerrando o capítulo, um *survey* de técnicas de combinação de classificadores para reconhecimento de faces é apresentado.

No Capítulo 4, a metodologia e um esquema inicial para reconhecimento de faces usando combinação de classificadores são fornecidos, assim como os softwares utilizados para

a implementação dos algoritmos e o método de avaliação dos resultados. Encerrando o capítulo, os resultados obtidos pelo sistema proposto sobre os bancos de dados de faces ORL e AR são relatados, analisados e discutidos.

Por fim, o Capítulo 5 apresenta as conclusões e alguns trabalhos futuros.

2 RECONHECIMENTO DE PADRÕES

Ler, falar, reconhecer objetos, identificar pessoas através da face, timbre de voz ou mesmo forma de andar, distinguir se alguém está feliz ou triste são todas tarefas realizadas diariamente por seres humanos que utilizam um complexo processo de reconhecimento de padrões. Apesar de complexa, essa tarefa costuma ser bastante intuitiva e não necessita de grande esforço por parte dos seres humanos. Entretanto, quando se considera ensinar uma máquina para realizar esses processos de reconhecimento de padrões ainda é um grande desafio.

De uma maneira geral, podemos definir reconhecimento de padrões da seguinte forma: “Reconhecimento de padrões é o estudo de como máquinas podem observar o ambiente, aprender a distinguir padrões de interesse do seu fundo, e tomar decisões confiáveis e razoáveis sobre as categorias dos padrões” (JAIN, DUIN *et al.*, 2000). Ou seja, de um modo simplista, reconhecer padrões consiste em atribuir um dado bruto ou objeto a uma categoria ou classe.

Basicamente, um processo de reconhecimento de padrões envolve várias etapas que são pré-processamento, extração de atributos e classificação. Pré-processamento consiste em aplicar aos dados brutos vindo dos sensores, operações como normalização e segmentação de forma a tornar a informação mais simples e adequada para as próximas etapas (DUDA, HART *et al.*, 2000). Extração de atributos pretende determinar um subespaço apropriado sobre o espaço original de forma que esse novo subespaço contenha dimensionalidade no máximo igual ao espaço original (JAIN, DUIN *et al.*, 2000). Esses atributos são os responsáveis por descrever um padrão e serão utilizados como entrada para um classificador que tomará uma decisão, ou seja, atribuirá um rótulo ou nome de uma classe para esse padrão. Essa etapa de classificação pode ser feita de dois modos: supervisionado e não-supervisionado. No modo supervisionado, as classes são previamente definidas por um Supervisor ou alguém que conhece o domínio do problema. Já no modo não-supervisionado, aglomerados de dados determinarão as classes.

Para projetar um sistema de reconhecimento de padrões, primeiramente, precisamos entender e definir o domínio do problema e a definição das classes. Dessa forma, podemos determinar como será o processo de aquisição (com a definição dos sensores), o pré-processamento, a representação dos dados no espaço original, o esquema de extração ou sele-

ção de atributos, o modelo que será o responsável por tomar a decisão (classificação) e como avaliar e validar os resultados de desempenho.

Em reconhecimento de padrões, as quatro melhores abordagens são *template matching*, estatística, sintática ou estrutural e redes neurais. Com um embasamento teórico-matemático consistente, fundamentado na Teoria de Decisão Bayesiana, a abordagem estatística se sobressai sobre as outras. A Teoria de Decisão Bayesiana consiste em encontrar o particionamento ótimo do espaço de atributos de forma a minimizar o risco de Bayes ou a probabilidade de erro se considerada uma matriz de perdas do tipo $[0,1]$. Os limites de decisão responsáveis pelo particionamento do espaço são determinados “pelas distribuições de probabilidade dos padrões pertencentes a cada classe, que deve ser especificado ou aprendido” (JAIN, DUIN *et al.*, 2000). Dessa forma, o conhecimento ou não dessas distribuições ou densidades condicionais das classes, o tipo de aprendizado dessas distribuições (supervisionado ou não-supervisionado) e a quantidade de informação disponível para o projetista do sistema definem uma divisão das técnicas de reconhecimento de padrões estatístico como mostrado pela estrutura de árvore da Figura 2.1. Nessa estrutura, quanto mais próximo das folhas e mais à direita, menos informação está disponível para o projetista, o que implica em aumento da dificuldade de classificação. Vale ressaltar que a divisão entre técnicas paramétricas e não-paramétricas é devida ao conhecimento da forma das densidades condicionais das classes (caso paramétrico) ou não (caso não-paramétrico).

Um padrão em reconhecimento de padrões estatísticos é representado por um vetor de atributos de d dimensões num espaço de atributos \mathcal{H}^d , ou seja, pode ser interpretado como um ponto nesse espaço. “O objetivo é escolher atributos que permitam aos padrões pertencentes a categorias diferentes ocuparem regiões compactas e disjuntas no espaço de atributos” (JAIN, DUIN *et al.*, 2000). Dessa forma, podemos perceber que a extração de atributos ou o espaço de atributos definido são os principais responsáveis pela separabilidade das classes. É justamente essa separabilidade (ou o quanto bem as classes podem ser separadas) que define a eficiência do espaço de atributos. Mais informações sobre reconhecimento de padrões estatísticos podem ser encontradas em (DEVIJVER e KITTLER, 1982; FUKUNAGA, 1990; DUDA, HART *et al.*, 2000; WEBB, 2002; THEODORIDIS e KOUTROUMBAS, 2003).

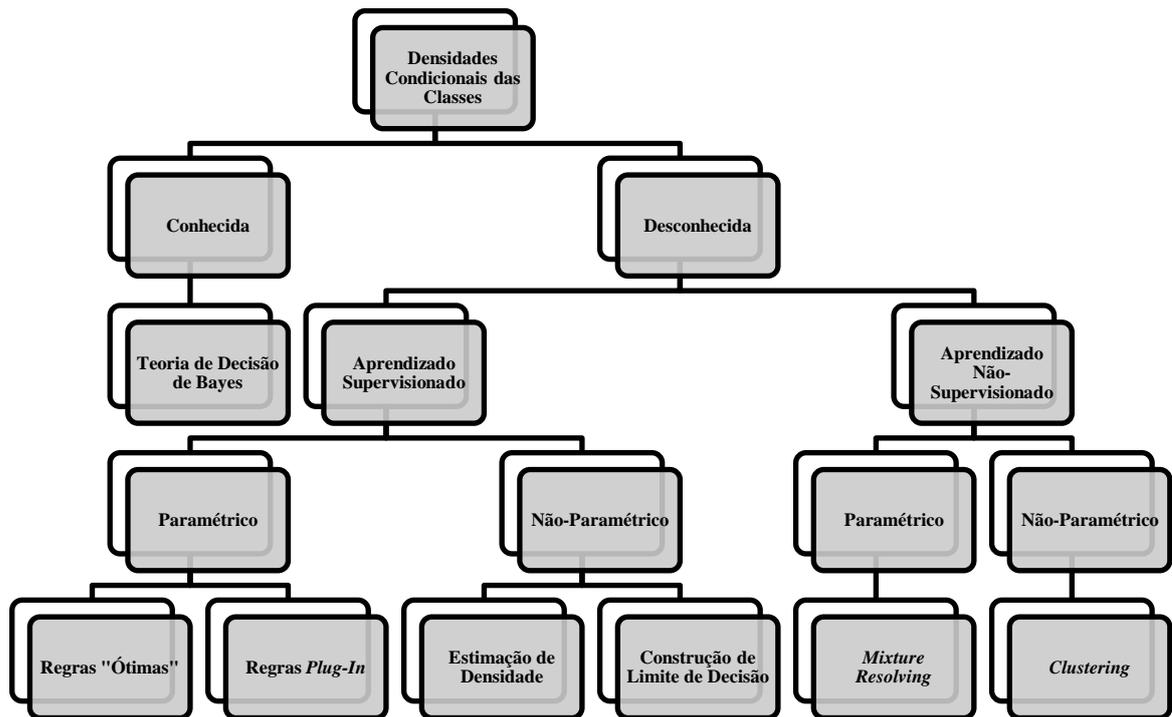


Figura 2.1. Divisão da Abordagem de Reconhecimento de Padrões Estatísticos. Adaptado de (JAIN, DUIN *et al.*, 2000).

No restante deste capítulo, são expostas algumas técnicas de extração de atributos e os seus problemas inerentes (Seção 2.1) e alguns classificadores (Seção 2.2). Na Seção 2.3, combinação de classificadores é explorada em maiores detalhes. Finalmente, na Seção 2.4, técnicas de partição dos dados para a validação de algoritmos de reconhecimento de padrões, além de medidas de desempenho e forma de medir se esse mesmo desempenho quando comparado com outros algoritmos é significativo são apresentadas.

2.1 Extração de Atributos

A redução da dimensionalidade é importante para a redução de custo e aumento da precisão de classificação (JAIN, DUIN *et al.*, 2000). A justificativa para a redução de custo se deve à menor quantidade de informação que precisa ser armazenada em memória e processada. O aumento da precisão de classificação está relacionado com a Maldição da Dimensionalidade e também ao problema de Tamanho Pequeno da Amostra (*Small Sample Size*), que serão apresentados nas Subseções 2.1.1 e 2.1.2.

Basicamente, duas abordagens foram propostas para reduzir a dimensão: extração e seleção de atributos. Enquanto seleção de atributos procura selecionar um subconjunto do espaço original que melhor represente os padrões e suas classes, extração de atributos cria novos atributos a partir de transformações ou combinações do conjunto original.

Matematicamente, extração de atributos consiste em determinar um subespaço de dimensionalidade m de modo linear ou não-linear no espaço de atributos original de dimensionalidade d , onde $m \leq d$. Aqui, apresentaremos a Análise de Componentes Principais (PCA) e algumas de suas variantes (Seção 2.1.3), o *Block-Based PCA* (2.1.4) e a Análise de Discriminantes Lineares (LDA) (Seção 2.1.5) como técnicas de extração de atributos.

2.1.1 Maldição da Dimensionalidade

“O desempenho de um classificador depende da relação entre tamanho da amostra, número de atributos e complexidade do classificador” (JAIN, DUIN *et al.*, 2000). A Teoria de Decisão Bayesiana considera número infinito de amostras, tal que aumentando-se a dimensão dos dados, a probabilidade de erro não é incrementada. Entretanto, na prática, temos número finito e pequeno de amostras se comparado com a dimensionalidade dos dados. Dessa forma, foi demonstrado que se aumentarmos a dimensionalidade dos dados, o desempenho do classificador será degradado. Esse problema é denominado a Maldição da Dimensionalidade.

Na prática, um número de amostras por classe igual a 10 vezes o número de atributos é geralmente aceito como um bom tamanho de amostra no projeto de classificadores (JAIN, DUIN *et al.*, 2000). Além disso, para resolver ou atenuar o problema da dimensionalidade, técnicas de extração ou seleção de atributos podem ser aplicadas. Outro problema relacionado com a Maldição da Dimensionalidade é o *Small Sample Size* que será apresentado na próxima Subseção.

2.1.2 Problema do Small Sample Size

O problema do Tamanho Pequeno da Amostra (*Small Sample Size*) é também chamado de Problema da Singularidade. Em tarefas de classificação de alta dimensionalidade como reconhecimento de faces, geralmente o número de amostras é menor do que a dimensionalidade dessas amostras. Com isso, no caso da estimação das matrizes de covariância intraclas-

ses como as utilizadas pelo LDA (que veremos na Seção 2.1.5) e pelo classificador de Máxima Verossimilhança (Seção 2.2.2) obteremos matrizes singulares, ou seja, que não admitem inversa. Dessa forma, o desempenho de um sistema que utiliza essas técnicas será degradado.

Entretanto, aumentar o número de amostras pode ser inviável e caro. Assim, existem algumas formas de atenuar esse problema. Na próxima Subseção, uma técnica de Regularização de Matrizes de Covariância será apresentada a fim de reduzir esse problema.

2.1.2.1 Regularização de Matrizes de Covariância

Uma técnica de regularização *Mixed-LOOC2* (Estimador de Covariância *Leave-One-Out* Misturado) foi proposta por (KUO e LANDGREBE, 2002) para ultrapassar o problema da singularidade de matrizes de covariância. Ela parte de dois princípios: 1) no caso de poucas amostras, a Matriz de Covariância Comum ou da Mistura pode substituir a Matriz de Covariância intraclasses já que apesar de ainda ser singular contém menor erro de estimação e 2) inserindo uma pequena perturbação na diagonal principal da matriz, pode torná-la não singular.

O *Mixed-LOOC2* propõe a seguinte forma das matrizes de covariância de cada classe i regularizadas:

$$\widehat{\Sigma}_i(\alpha_i) = \alpha_i A + (1 - \alpha_i) B,$$

onde $A = \frac{tr(S_i)}{p} I, diag(S_i), S_i, \frac{tr(S_p)}{p} I, diag(S_p)$ ou S_p , $B = S_i$ ou $diag(S_p)$, S_i é a matriz de covariância da classe i , S_p é a matriz de covariância comum (*pooled*), p é o número de atributos, tr é o traço da matriz, $diag$ é uma matriz diagonal com somente a diagonal principal mantida e o restante é zero e α_i é um parâmetro que deve ser definido com valor próximo a 1 e deve ser estimado para cada classe seguindo um critério de seleção de máximo logaritmo da verossimilhança de *Leave-One-Out* (LOOL) dado por

$$LOOL_i = \frac{1}{N_i} \sum_{k=1}^{N_i} \ln \left[f \left(x_k | m_{i/k}, \widehat{\Sigma}_{i/k}(\alpha_i) \right) \right],$$

onde N_i é o número de amostras por classe, $m_{i/k}$ é a média da classe sem a amostra k e $\widehat{\Sigma}_{i/k}$ é a matriz de covariância da classe estimada sem a amostra k . Como a estimativa de $\widehat{\Sigma}_{i/k}$ é a

plicada como entrada para o LOOL, matrizes de covariância intraclasses $S_{i/k}$ e comum $S_{p_{i/k}}$ serão calculadas removendo uma amostra (*Leave-One-Out*).

As matrizes de covariância das classes e média das classes sem a amostra k , considerando L classes, são dadas, respectivamente, por

$$S_{i/k} = \frac{1}{N_i - 2} \sum_{\substack{j=1 \\ j \neq k}}^{N_i} (x_{i,j} - m_{i/k})(x_{i,j} - m_{i/k})^T \quad e$$

$$m_{i/k} = \frac{1}{N_i - 1} \sum_{\substack{j=1 \\ j \neq k}}^{N_i} x_{i,j} .$$

Por sua vez, a matriz de covariância comum (*pooled*) S_p é definida pela equação

$$S_p = \left[\sum_{i=1}^L \frac{f_i}{f_i + t - p - 1} \right]^{-1} \sum_{i=1}^L \frac{f_i S_i}{f_i + t - p - 1},$$

onde p é o número de atributos, L é o número de classes, S_i é a matriz de covariância da classe i e f_i e t correspondem às seguintes equações, respectivamente

$$f_i = N_i - 1 \quad e$$

$$t = \frac{(\alpha_i - 1)f_i - \alpha_i(p + 1)}{2 - \alpha_i}.$$

Onde N_i é o número de amostras por classes. Durante a etapa de estimação dos parâmetros α_i , A e B , a matriz de covariância comum (*pooled*) deverá considerar no seu cálculo a remoção de uma amostra como proposto pelo *Leave-One-Out*. Esta matriz é denotada por $S_{p_{i/k}}$ e é definida como

$$S_{p_{i/k}} = \left[\left(\sum_{\substack{j=1 \\ j \neq i}}^L \frac{f_j}{f_j + t - p - 1} \right) + \frac{f'_i}{f'_i + t_{i/k} - p - 1} \right]^{-1} \left[\left(\sum_{\substack{j=1 \\ j \neq i}}^L \frac{f_j S_j}{f_j + t - p - 1} \right) + \frac{f'_i S_{i/k}}{f'_i + t_{i/k} - p - 1} \right]$$

$$f'_i = N_i - 2 \quad e$$

$$t_{i/k} = \frac{(\alpha_i - 1)f'_i - \alpha_i(p + 1)}{2 - \alpha_i}.$$

Dessa forma, precisamos determinar qual o valor de α_i , e os tipos A e B para cada classe que maximiza o LOOL para obter uma matriz de covariância não-singular e que melhor represente a classe. Vale ressaltar que a matriz de covariância estimada final para cada classe considera todas as amostras da classe.

Ainda, é importante notar que esta técnica de regularização exigirá no mínimo 3 amostras por classe no treinamento, pois durante a etapa das estimativas dos parâmetros α_i , A e B consideramos o critério *Leave-One-Out* e portanto, teremos uma amostra a menos para estimar as matrizes de covariância, sendo que para se calcular a matriz de covariância de uma dada classe é necessária ao menos 2 amostras desta classe.

2.1.3 PCA e suas variantes

Análise de Componentes Principais (PCA), ou Expansão de Karhunen-Loève, é a técnica de extração de atributos mais utilizada em reconhecimento de padrões estatístico. O PCA é linear e não supervisionado e determina um espaço onde os atributos são mutuamente descorrelacionados.

Uma limitação do PCA é que ele não garante a separabilidade das classes, já que não foi projetado com essa finalidade, mas com o propósito de encontrar uma representação comprimida ou compacta da mistura removendo informações redundantes. A Figura 2.2 ilustra essa limitação do PCA do ponto de vista de separabilidade. No exemplo ilustrado, podemos notar que usando a projeção sobre o autovetor a_1 associado ao maior autovalor as classes co-

incidem. Entretanto, se fosse utilizada a projeção sobre o outro autovetor as classes continuariam separadas.

PCA procura uma transformação ortogonal A que projeta um vetor d -dimensional x centralizado pela subtração da média global μ num vetor m -dimensional y , sendo $m < d$ e é dada por

$$y = A^T(x - \mu).$$

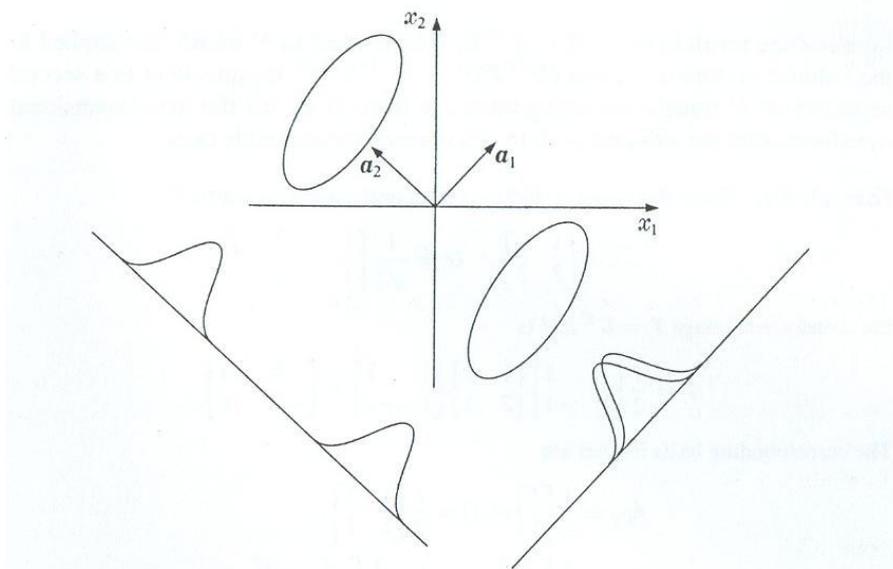


Figura 2.2. Limitação da Transformação de Karhunen-Loève quanto à separabilidade entre classes. Retirado de (THEODORIDIS e KOUTROUMBAS, 2003).

Foi identificado por meio da otimização da representação considerando alguns critérios como a maximização da variância em poucos componentes ou mínimo erro quadrático que essa transformação é dada pelos m autovetores a_i (denominados aqui componentes principais) associados aos m maiores autovalores λ_i em ordem decrescente extraídos da matriz de covariância da mistura Σ , representada por

$$\Sigma a_i = \lambda_i a_i.$$

Ou seja, geometricamente, temos uma rotação do sistema de coordenadas. A Figura 2.3 mostra um exemplo para um caso gaussiano multivariado. Como podemos ver, os autovetores da matriz de covariância φ_1 e φ_2 correspondem aos componentes principais da distribuição definida pelo hiperelipsóide e os seus autovalores λ_1 e λ_2 são as variâncias nos respecti-

vos componentes principais. Além disso, foi determinado que a soma dos autovalores corresponde à variância total.

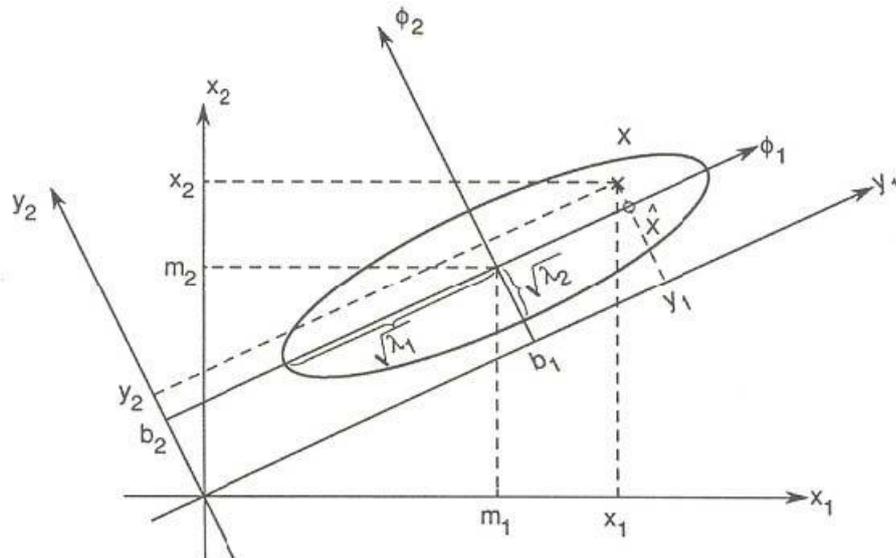


Figura 2.3. Ilustração da Transformada de Karhunen-Loève sobre o caso Gaussiano Bidimensional. Retirado de (FUKUNAGA, 1990).

A seguir, algumas variantes de PCA serão abordadas. Essas variantes são denominadas KL1-KL5, sendo KL1 o PCA original. O que as difere do PCA é que elas utilizam informação da classe e critérios diferentes para a seleção dos atributos que formarão o novo sistema de coordenadas e são técnicas supervisionadas.

2.1.3.1 KL2

KL2 foi proposto por (CHIEN e KING-SUN, 1967) com o propósito de utilizar a informação contida dentro da classe. A diferença para o KL1 é que os autovetores são calculados da média das matrizes de covariância de cada classe S_W .

2.1.3.2 KL3

KL3 foi proposto para utilizar a informação discriminante contida nas médias das classes para criar um critério de seleção dos autovetores criados como no KL2 (DEVIJVER e KITTTLER, 1982; WEBB, 2002). O critério de ordenação e seleção dos autovetores é dado por

$$J_j = \frac{a_j^T S_B a_j}{\lambda_j},$$

onde a_j são os autovetores de S_W , λ_j são os autovalores de S_W e S_B é a matriz de espalhamento entre classes. Os autovetores serão ordenados em ordem decrescente do critério J_j .

2.1.3.3 KL4

Outra forma de ordenar os autovetores extraídos como no KL2 é usar a informação comprimida nas variâncias já que nem sempre a informação da média é suficiente para separar as classes (WEBB, 2002). Essa técnica é denominada KL4 e foi proposta por (KITTTLER e YOUNG, 1973). O critério para ordenação utilizado é uma medida baseada em função de entropia logarítmica, onde os autovetores serão ordenados em ordem crescente da medida dada por

$$H_j = - \sum_{i=1}^C \frac{\lambda_{ij}}{\lambda_j} \log \left(\frac{\lambda_{ij}}{\lambda_j} \right)$$

$$\lambda_j = \sum_{i=1}^C \lambda_{ij}$$

$$\lambda_{ij} = p(w_i) a_j^T \widehat{\Sigma}_i a_j,$$

onde $p(w_i)$ é a probabilidade a priori da classe w_i , $\widehat{\Sigma}_i$ é a matriz de covariância da classe i , λ_{ij} é a variância do atributo j na classe i , a_j são os autovetores de S_W , C é o número de classes.

2.1.3.4 KL5

KL3 é expandido na variação KL5 proposta por (KITTLER e YOUNG, 1973) de forma a comprimir a informação discriminatória contida na média das classes. Essa técnica tem uma interpretação geométrica próxima ao do LDA que será apresentado na Seção 2.1.5. Basicamente, o KL5 é um processo de dois estágios: 1) encontrar uma transformação que torne S_W diagonal (no caso, levando à matriz identidade) e 2) comprimir a informação da média das classes encontrando uma nova transformação diagonal que reduza a dimensão dos vetores média das classes.

No primeiro estágio, queremos encontrar uma transformação $B = U_r A_r^{-\frac{1}{2}}$, onde U_r é a matriz composta pelos r primeiros autovetores associados aos r maiores autovalores de S_W e A_r é a matriz diagonal composta pelos r maiores autovalores de S_W , tal que obtemos nesse novo espaço matrizes S'_W iguais à identidade, ou seja,

$$S'_W = B^T S_W B = I.$$

No segundo estágio, precisaremos calcular as novas matrizes de covariância entre classes S'_B nesse novo espaço e calcular os seus autovetores e autovalores. S'_B é calculado aplicando

$$S'_B = B^T S_B B.$$

Dessa forma, poderemos obter o extrator de características ótimo final A^T dado por

$$A^T = V_v^T B^T,$$

onde v é o posto da matriz S'_B , V_v consiste na matriz composta pelos v autovetores de S'_B associados aos v maiores autovalores. No máximo, teremos $(C - 1)$ autovalores não nulos.

2.1.4 *Block-Based PCA*

Block-based PCA é uma técnica aplicada diretamente sobre uma única amostra ao invés de considerar a mistura. A idéia aqui é reduzir a dimensionalidade sem perder informações relevantes da imagem. Para isso, primeiramente, a imagem é dividida em $k \times k$ blocos de mesmo tamanho. Esses blocos em forma lexicográfica serão como amostras da imagem e representariam a mistura não entre classes, mas entre partes de uma mesma face. Em seguida,

aplicaremos PCA sobre a matriz formada pelos blocos em forma lexicográfica (cada bloco formará uma linha da matriz) para obtermos os principais componentes. Depois, cada bloco é projetado sobre os m componentes principais encontrados e a imagem pode assim ser reconstruída com menos componentes. A Figura 2.4 ilustra essa técnica, enquanto a Figura 2.5 apresenta um exemplo com imagem real usando bloco 8×8 .

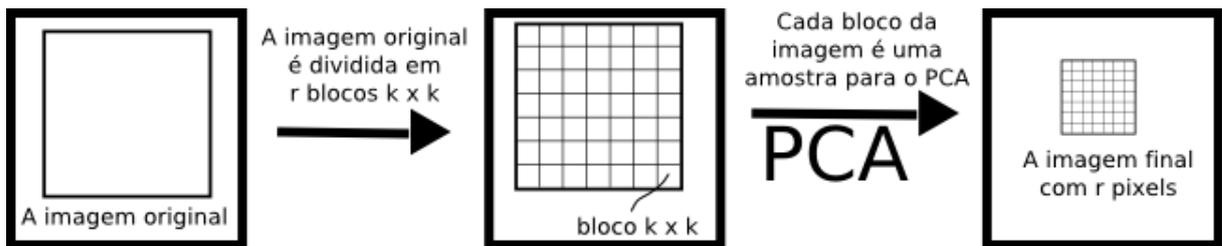


Figura 2.4. Ilustração do processo aplicado no *Block-based PCA*. Nesse exemplo, a imagem final contém o mesmo número de pixels quanto o número de blocos da imagem, o que implica que apenas o componente de maior variância foi selecionado.

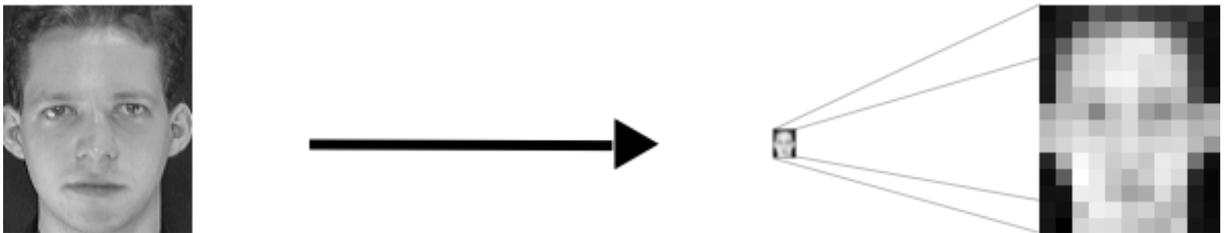


Figura 2.5. Exemplo de imagem real com aplicação do extrator de atributos *Block-based PCA*. Neste exemplo foi considerado blocos 8×8 . No lado esquerdo da seta temos a imagem original, enquanto no lado direito temos a imagem reconstruída. Somente para facilitar a visualização da imagem reconstruída ela foi ampliada ao seu lado.

2.1.5 LDA

Análise de Discriminante Linear (LDA) visa obter uma transformação linear que maximiza o espalhamento entre classes S_B (separabilidade) enquanto minimiza o espalhamento intraclasses S_W (variabilidade). Baseado numa generalização do critério de Fisher para o problema multiclases, até $C - 1$ bases ortogonais podem ser obtidas tal que o critério de Fisher é maximizado. Esses vetores base são determinados por um problema de autovalores de $S_W^{-1}S_B$, onde no máximo $C - 1$ autovetores associados aos $C - 1$ maiores autovalores podem ser selecionados. O critério de Fisher, e as matrizes S_W e S_B podem ser definidos, respectivamente, como

$$J_F(a) = \frac{a_j^T S_B a_j}{a_j^T S_W a_j},$$

$$S_W = \sum_{i=1}^C \frac{n_i}{n} \widehat{\Sigma}_i,$$

$$S_B = \sum_{i=1}^C \frac{n_i}{n} (m_i - m)(m_i - m)^T,$$

onde n_i é o número de amostras da classe i , C é o número de classes, n é o número total de amostras, m_i é a média da classe i , m é a média global, a_j são autovetores que maximizam o critério de Fisher, $\widehat{\Sigma}_i$ são as matrizes de covariância da classe i . Como é necessário o cálculo da inversa de S_W , caso haja número insuficiente de amostras, S_W poderá ser singular e sofrer com o problema do *Small Sample Size*. Para resolver esse problema, uma forma muito utilizada é aplicar o PCA antes do LDA. Entretanto, informação discriminante poderá estar contida no espaço nulo do PCA e, portanto, ser descartada.

2.2 Classificação

Um classificador é o responsável por realizar uma decisão e pode ser entendido como uma função D que associa um vetor de padrões de entrada $x \in R^n$ a uma classe $w \in \Omega$ particionando um espaço de atributos, ou seja,

$$D: R^n \rightarrow \Omega.$$

Nesta Seção, abordaremos dois classificadores comuns que serão utilizados nesse trabalho: Máxima Verossimilhança (Seção 2.2.1) e KNN (Seção 2.2.2).

2.2.1 Máxima Verossimilhança (MaxVer)

O classificador de Máxima Verossimilhança (MaxVer) é baseado diretamente na Regra de Decisão de Bayes e é considerada uma técnica Plug-in Rule se os parâmetros das densidades foram estimados e injetados nessas densidades. A regra de decisão do classificador consiste em associar o vetor de padrões x à classe w_i , tal que

$$p(x|w_i) \geq p(x|w_j), \text{ para todo } j = 1, \dots, C,$$

onde C é o número de classes e $p(x|w_j)$ é a função de verossimilhança que pode representada por uma função de densidade de probabilidade com distribuição normal ou Gaussiana Multivariada que é parametrizada pela média μ e matriz de covariância Σ da classe, ou seja, $p(x|w_j) \sim N(\mu, \Sigma)$. A distribuição Gaussiana Multivariada é dada por

$$p(x|w_j) = \frac{1}{(2\pi)^{d/2} |\Sigma|^{1/2}} \exp \left[-\frac{1}{2} (x - \mu)^T \Sigma^{-1} (x - \mu) \right],$$

onde d é a dimensionalidade do espaço, $|\Sigma|$ é o determinante da matriz de covariância da classe, Σ^{-1} é a inversa da matriz de covariância da classe e \exp representa o exponencial.

2.2.2 K-Vizinhos Mais Próximos (KNN)

O classificador de K-Vizinhos Mais Próximos (KNN) é uma abordagem de construção de limite de decisão. De uma maneira prática, a regra de decisão do KNN consiste em, dados um conjunto de protótipos rotulados $V = \{v_1, \dots, v_V\}$ e o vetor de padrões x a ser classificado, x será associado à classe mais representativa entre os k vizinhos mais próximos, com o k a ser definido inicialmente. A probabilidade condicional da classe pode ser estimada considerando uma hipersfera R (no caso euclidiano) composta pelos k protótipos rotulados mais próximos à x . Ela pode ser aproximada por

$$p(x|w_i) \approx \frac{k_i}{N_i V_R},$$

onde k_i é o número de ocorrências da classe w_i em R , N_i é o número de protótipos da classe w_i em V , V_R é uma variável que define o volume da hipersfera e o denominador pode ser entendido como um fator de normalização. Com isso, podemos obter uma probabilidade a posteriori definida como

$$P(w_i|x) \approx \frac{k_i}{k}.$$

Um caso específico do KNN é a denominada Regra do Vizinho Mais Próximo (NN) e ocorre quando um $k = 1$ é fixado. Para determinar a proximidade entre uma amostra x e um protótipo v_i geralmente a distância Euclidiana é utilizada.

2.3 Combinação de Classificadores

“A idéia de combinar entradas múltiplas para inferir informações sobre o ambiente real é muito natural. Ela é feita por humanos diariamente, já que estamos o tempo todo combinando informações acústicas, visuais, táteis, olfatórias e térmicas para reagir sobre o mundo à nossa volta. Sistemas biológicos parecem adotar tais esquemas de combinação muito frequentemente para obter conhecimento mais confiável sobre o ambiente. Algumas vezes não é mesmo possível derivar a informação necessária para uma tarefa particular de um único sensor, mas num esforço conjunto isso pode ser feito. Além do mais, não há sensor perfeito, então é razoável usar as propriedades favoráveis de um sensor e suprimir as suas desvantagens aplicando um esquema de combinação inteligente” (ACHERMANN e BUNKE, 1996a).

Baseado nessa inspiração biológica, fusão de sensores e combinação de classificadores têm sido explorados como uma alternativa para obter melhores desempenhos em tarefas de reconhecimento de padrões sem precisar aumentar a complexidade. Apesar disso, essa área que está crescendo rapidamente ainda não tem uma formulação matemática consolidada e muitas questões ainda estão abertas. Consequentemente, ainda há muito de empirismo nessa área.

Entretanto, foi identificado que as saídas dos classificadores individuais devem ser fortemente descorrelacionadas nos erros de classificação. “Isto é, os classificadores não devem concordar um com o outro quando erram a classificação de uma amostra, ou ao menos não devem associar uma amostra à mesma classe incorreta” (KITTLER, HATEF *et al.*, 1998). É justamente essa complementariedade entre os classificadores que faz com que uma fusão possa obter bons desempenhos. A fim de obter essa complementariedade um *ensemble* pode ser estabelecido. Para isso, podemos variar os combinadores (nível de Combinação), os classificadores individuais ou especialistas (nível de Classificador), os atributos determinados por técnicas de seleção ou extração de atributos diferentes (nível de Características) e os subconjuntos de dados (nível de Dados). Isso é ilustrado pela Figura 2.6.

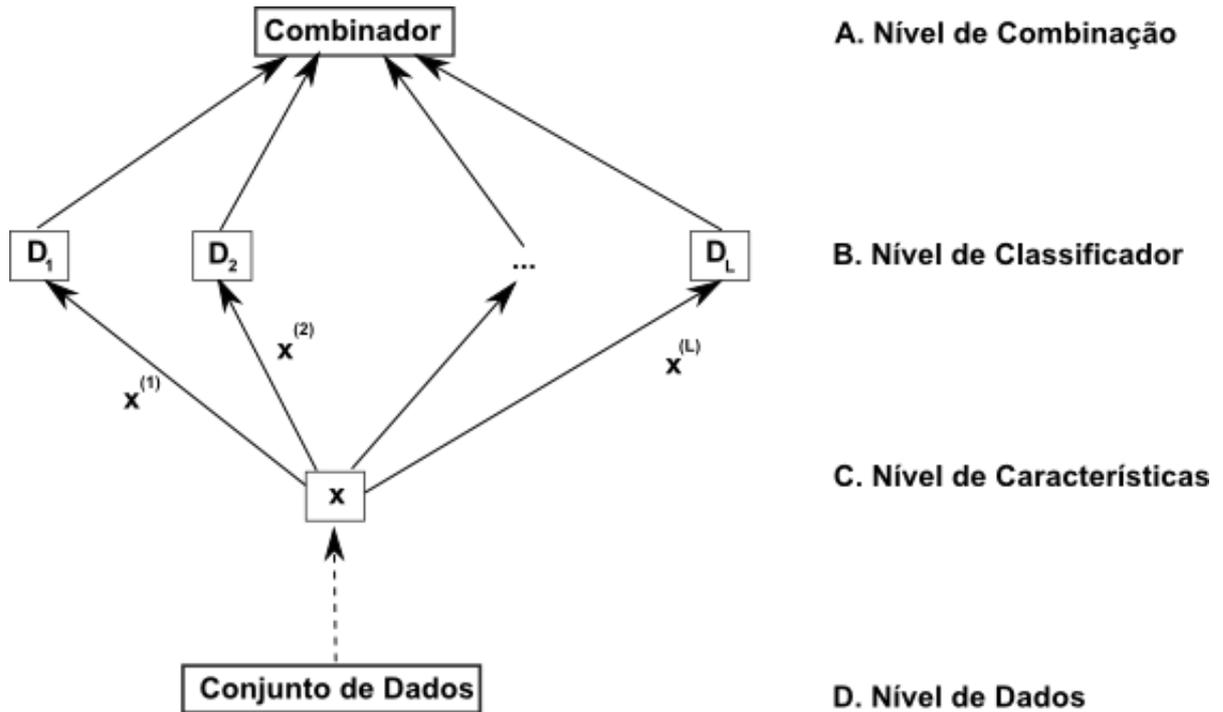


Figura 2.6. Abordagens para a construção de *ensembles* em combinação de classificadores. Adaptado de (KUNCHEVA, 2004).

Um combinador de classificadores nada mais é do que um novo classificador, que toma como entrada a resposta de R classificadores e como saída uma decisão final sobre eles. Essas respostas dos classificadores podem se apresentar em 3 níveis, como proposto por (XU, KRZYZAK *et al.*, 1992):

1. **Abstração (*menos informação para a tomada de decisão*):** apenas a informação do rótulo da classe associada ao padrão é fornecida por cada classificador.
2. **Ranking:** uma lista contendo a ordem das classes é fornecida pelos classificadores individuais, onde a classe mais provável está no topo da lista e recebe *rank*-1.
3. **Medida (*mais informação disponível para a tomada de decisão*):** uma medida que pode ser interpretada como o grau de confiança de determinada classe ser a correta é fornecida por cada especialista.

Portanto, antes de combinar a saída de diversos especialistas, essas saídas precisam ser colocadas num mesmo domínio e normalizadas. Isso implica que o tipo de informação gerada pelos classificadores individuais influenciam a forma de combinar. O domínio de probabilidades a posteriori é um normalmente considerado.

No restante dessa Seção, apresentaremos alguns dos principais combinadores em nível de decisão. Esses combinadores são divididos em não-treináveis (fixos) e em treináveis (necessitam de uma etapa de treinamento).

2.3.1 Combinadores Não-Treináveis

Combinadores não treináveis ou fixos são aqueles que não necessitam de um passo de treinamento ou ajuste. Esses combinadores são baseados no *framework* desenvolvido em (KITTLER, HATEF *et al.*, 1998), onde foram definidas as Regras do Produto e da Soma e, a partir dos mesmos, outros combinadores foram deduzidos.

Essas regras partem da Regra de Decisão Bayesiana considerando probabilidades conjuntas. Dado a saída x_i de R classificadores, com $i = 1, \dots, R$, um padrão Z deve ser associado à classe w_j de m possíveis classes (w_1, \dots, w_m) que maximiza a probabilidade a posteriori, ou seja, associar $Z \rightarrow w_j$ se

$$P(w_j | x_1, \dots, x_R) = \max_k P(w_k | x_1, \dots, x_R)$$

$$P(w_k | x_1, \dots, x_R) = \frac{p(x_1, \dots, x_R | w_k)P(w_k)}{p(x_1, \dots, x_R)},$$

onde $P(w_k)$ é a probabilidade a priori de ocorrência da classe w_k , $p(x_1, \dots, x_R | w_k)$ é uma função de densidade de probabilidade conjunta e $p(x_1, \dots, x_R)$ é a densidade de probabilidade conjunta incondicional e é dada por

$$p(x_1, \dots, x_R) = \sum_{j=1}^m p(x_1, \dots, x_R | w_j)P(w_j).$$

Entretanto, muitas vezes pode ser difícil determinar a densidade de probabilidade conjunta $p(x_1, \dots, x_R | w_k)$. Assim, as regras básicas da Soma e do Produto levam em conta algumas outras limitações ou situações para conseguir defini-la.

Nas próximas subseções, os combinadores básicos definidos pelas Regras do Produto e da Soma e outras decorrentes desses serão apresentados.

2.3.1.1 Regra do Produto

A Regra do Produto supõe que as representações dos dados são estatisticamente independentes. Essa é considerada uma regra severa, já que em aplicações reais essa condição pode ser difícil de ser atendida. Entretanto, algumas aproximações podem ser satisfeitas variando o espaço de atributos ou utilizando classificadores com características diferentes. Como consequência dessa suposição, podemos definir a densidade de probabilidade conjunta como

$$p(x_1, \dots, x_R | w_k) = \prod_{i=1}^R p(x_i | w_k).$$

Após algumas deduções, podemos obter a Regra do Produto tal que associamos $Z \rightarrow w_j$ se

$$P(w_j) \prod_{i=1}^R P(w_j | x_i) = \max_{1 \leq k \leq n} P(w_k) \prod_{i=1}^R P(w_k | x_i).$$

Ou seja, considerando classes equiprováveis, a probabilidade de uma hipótese é calculada pelo produto das probabilidades a posteriori obtidas de cada classificador.

2.3.1.2 Regra da Soma

A Regra da Soma advém da Regra do Produto e assume que as probabilidades a posteriori obtidas pelos classificadores não se distanciam muito das probabilidades a priori. Segundo (KITTLER, HATEF *et al.*, 1998), “essa suposição é rígida, mas pode ser satisfeita quando a informação discriminatória disponível é altamente ambígua devido a altos níveis de ruído”. Podemos expressar essa condição, matematicamente, como

$$P(w_k | x_i) = P(w_k)(1 + \delta_{ki}),$$

onde δ_{ki} satisfaz $\delta_{ki} \ll 1$. Utilizando essa condição e após algumas deduções, podemos obter a Regra da Soma tal que associamos $Z \rightarrow w_j$ se

$$P(w_j) + \sum_{i=1}^R P(w_j | x_i) = \max_{1 \leq k \leq n} \left[P(w_k) + \sum_{i=1}^R P(w_k | x_i) \right].$$

2.3.1.3 Regra da Soma Ponderada

A Regra da Soma Ponderada origina-se da Regra da Soma, dando um peso ρ_i para a saída de cada classificador, ou seja, classificadores com maiores desempenhos segundo algum critério seriam mais considerados do que outros. Esse critério poderia ser a taxa de classificação correta sobre o conjunto de treinamento ou ainda algum conjunto de avaliação. Considerando probabilidades a priori iguais, podemos definir a Regra da Soma Ponderada tal que associamos $Z \rightarrow w_j$ se

$$\sum_{i=1}^R \rho_i P(w_j | x_i) = \max_{1 \leq k \leq m} \sum_{i=1}^R \rho_i P(w_k | x_i).$$

2.3.1.4 Derivações de Outras Regras de Combinação a partir das Regras da Soma e do Produto

A partir das Regras da Soma e do Produto, podemos definir uma relação entre elas que será condição básica para definir as outras regras apresentadas abaixo. Essa relação é

$$\prod_{i=1}^R P(w_k | x_i) \leq \min_{1 \leq i \leq R} P(w_k | x_i) \leq \frac{1}{R} \sum_{i=1}^R P(w_k | x_i) \leq \max_{1 \leq i \leq R} P(w_k | x_i).$$

Além disso, podemos produzir a partir das probabilidades a posteriori dos classificadores, uma saída no nível abstrato que será utilizada posteriormente no combinador de Votação por Maioria. Para fazer isso, a seguinte função é utilizada

$$\Delta_{ki} = \begin{cases} 1, & \text{se } P(w_k | x_i) = \max_{1 \leq j \leq m} P(w_j | x_i) \\ 0, & \text{caso contrário} \end{cases}$$

Considerando a relação e a função de binarização definidas acima, além das Regras do Produto e da Soma, podemos definir os próximos combinadores.

2.3.1.5 Regra do Máximo

Considerando probabilidades a priori iguais, partindo da Regra da Soma e substituindo a soma pelas máximas probabilidades a posteriori, podemos definir a Regra do Máximo tal que associamos $Z \rightarrow w_j$ se

$$\max_{1 \leq i \leq R} P(w_j | x_i) = \max_{1 \leq k \leq m} \max_{1 \leq i \leq R} P(w_k | x_i).$$

2.3.1.6 Regra do Mínimo

Considerando probabilidades a priori iguais, partindo da Regra do Produto, limitando o produto das probabilidades a posteriori superiormente pelo seu mínimo, podemos definir a Regra do Mínimo tal que associamos $Z \rightarrow w_j$ se

$$\min_{1 \leq i \leq R} P(w_j | x_i) = \max_{1 \leq k \leq m} \min_{1 \leq i \leq R} P(w_k | x_i).$$

2.3.1.7 Regra da Mediana

Considerando probabilidades a priori iguais, a Regra da Soma pode ser vista como uma Regra da Média. Entretanto, a média pode ser afetada por *outliers*. Dessa forma, é proposta a utilização da mediana ao invés da média, o que define a Regra da Mediana tal que associamos $Z \rightarrow w_j$ se

$$\text{med}_{1 \leq i \leq R} P(w_j | x_i) = \max_{1 \leq k \leq m} \text{med}_{1 \leq i \leq R} P(w_k | x_i).$$

2.3.1.8 Votação por Maioria

Considerando probabilidades a priori iguais, partindo da Regra da Soma e usando função de binarização da saída, podemos definir a Regra da Votação por Maioria tal que associamos $Z \rightarrow w_j$ se

$$\sum_{i=1}^R \Delta_{ji} = \max_{1 \leq k \leq m} \sum_{i=1}^R \Delta_{ki}.$$

Basicamente, essa regra pode ser entendida como uma simples contagem de votos, onde a classe que obtiver mais votos será a associada ao padrão de entrada.

2.3.2 *Combinadores Treináveis*

Combinadores treináveis são aqueles que necessitam de uma etapa de treinamento, seja para construção de *templates* ou ajuste de parâmetros. Aqui, apresentaremos dois combinadores treináveis: Dempster-Shafer e *Decision Templates* que utilizam perfis de decisão como atributos num espaço intermediário para a tomada de decisão que será realizada por algum outro classificador. Por isso, são denominados inconscientes da classe (KUNCHEVA, 2004).

Um perfil de decisão $DP(x)$ consiste numa matriz composta pelas saídas (graus de suporte) para c classes de L classificadores individuais D_i considerando um padrão de entrada x . Denotamos $d_{i,j}(x)$ como o suporte para hipótese de x ser da classe w_j dada por um classificador D_i . Portanto, $DP(x)$ pode ser construído como

$$DP(x) = \begin{bmatrix} d_{1,1}(x) & \cdots & d_{1,j}(x) & \cdots & d_{1,c}(x) \\ \vdots & & \vdots & & \vdots \\ d_{i,1}(x) & \cdots & d_{i,j}(x) & \cdots & d_{i,c}(x) \\ \vdots & & \vdots & & \vdots \\ d_{L,1}(x) & \cdots & d_{L,j}(x) & \cdots & d_{L,c}(x) \end{bmatrix}.$$

A seguir, os combinadores de Decision Templates e Dempster-Shafer serão apresentados.

2.3.2.1 Decision Templates

O Decision Templates (DT) (KUNCHEVA, 2004) “aprende” qual o perfil de decisão mais comum para cada classe w_j , denotado por DT_j . Para fazer isso, uma etapa de treinamento é necessária. Para calcular cada DT_j , utilizamos todos os perfis de decisão $DP(z_k)$ tal que z_k é uma amostra pertencente a w_j do conjunto de treinamento Z . DT_j será a média dos perfis de decisão $DP(z_k)$, ou seja,

$$DT_j = \frac{1}{N_j} \sum_{\substack{z_k \in w_j \\ z_k \in Z}} DP(z_k),$$

onde N_j é o número de amostras de w_j em Z .

Por fim, para realizar a tomada de decisão, um valor $\mu_j(x)$ de uma medida de similaridade S entre o perfil de decisão $DP(x)$ do padrão a ser classificado e todo DT_j é calculado, ou seja,

$$\mu_j(x) = S(DP(x), DT_j), \quad j = 1, \dots, c.$$

Essa medida de similaridade pode ser uma distância Euclidiana quadrada, definida por

$$S(DP(x), DT_j) = 1 - \frac{1}{L \times c} \sum_{i=1}^L \sum_{k=1}^c [DT_j(i, k) - d_{i,k}(x)]^2.$$

Finalmente, x será associado à classe que fornecer o maior valor de $\mu_j(x)$.

2.3.2.2 Dempster-Shafer

O combinador Dempster-Shafer é inspirado na Teoria da Evidência proposta por Dempster e Shafer (KUNCHEVA, 2004). Ao invés de calcular a similaridade entre os perfis e templates de decisão, uma medida de crença é calculada baseada também em DPs e DTs. Dempster-Shafer utiliza o mesmo passo de treinamento de Decision Templates, mudando apenas a regra de decisão que agora considera a informação de proximidade entre o template de decisão e a saída dos classificadores. Essa regra de decisão é dividida em 3 passos: 1) Cálculo de uma medida de proximidade entre os templates de decisão e a saída dos classificadores individuais; 2) Cálculo de medidas de crença para toda classe e todo classificador e 3) Determinação de graus de suporte finais e tomada de decisão.

Para calcular uma medida de proximidade Φ entre DT_j^i (i -ésima linha do template de decisão da classe w_j) e a saída de um classificador D_i (dada pela i -ésima linha de um perfil de decisão $DP(x)$), usamos a seguinte equação

$$\Phi_{j,i}(x) = \frac{(1 + \|DT_j^i - D_i(x)\|^2)^{-1}}{\sum_{k=1}^c (1 + \|DT_k^i - D_i(x)\|^2)^{-1}},$$

onde $\| \cdot \|$ representa a norma da matriz.

Na segunda etapa, utiliza-se as medidas de proximidade obtidas na etapa anterior, para calcular medidas de crença $b_j(D_i(x))$ para toda classe $j = 1, \dots, c$ e todo classificador $i = 1, \dots, L$, usando

$$b_j(D_i(x)) = \frac{\Phi_{j,i}(x) \prod_{k \neq j} (1 - \Phi_{k,i}(x))}{1 - \Phi_{j,i}(x) [1 - \prod_{k \neq j} (1 - \Phi_{k,i}(x))]}.$$

Finalmente, graus de suporte $\mu_j(x)$ fornecerão a decisão final. Um vetor de padrão x será associado à classe que obtiver o maior grau de suporte. A equação que define $\mu_j(x)$ é

$$\mu_j(x) = K \prod_{i=1}^L b_j(D_i(x)), \quad j = 1, \dots, c,$$

onde K é uma constante de normalização.

2.4 Validação

Para avaliar o desempenho de um classificador, alguns métodos para partição dos dados, estimação de erros e avaliação de significância no caso de comparação entre o desempenho de dois classificadores são necessários.

Nessa seção, serão apresentadas as principais técnicas para partição do conjunto de dados para treinamento e teste dos classificadores e a estimação de erro. Além disso, outras medidas de avaliação de desempenho como o coeficiente Kappa e a taxa de reconhecimento correto serão abordadas. O Kappa será também base para métodos de teste de significância para avaliar se o classificador proposto tem desempenho superior significativo se comparado a outro classificador.

2.4.1 Técnicas de partição dos dados e Estimação de Erros

Métodos de Estimação de Erros são essenciais para avaliar o desempenho de um classificador. Eles definem, basicamente, como o conjunto de dados será particionado entre conjuntos de treinamento e teste e, a partir disso, como determinar a estimativa da probabilidade de erro. Nessa seção, apresentaremos quatro dos principais métodos de estimação de erros:

Resubstitution, *Holdout*, *Leave-One-Out* e Método da Rotação. A probabilidade de erro dessas técnicas pode ser estimada a partir da contabilidade dos erros e acertos.

O *Resubstitution* utiliza todo o conjunto de dados disponível para treinamento e teste (JAIN, DUIN *et al.*, 2000). Dessa forma, é considerado um estimador viesado otimista.

Holdout divide o conjunto de dados em dois subconjuntos disjuntos: um para treinamento e outro para teste. Geralmente, o conjunto de dados é dividido ao meio. Assim, esse método tem a principal desvantagem de reduzir ambos os conjuntos de treinamento e teste (THEODORIDIS e KOUTROUMBAS, 2003). Ainda, é considerado um estimador viesado pessimista.

O *Leave-One-Out*, como o próprio nome diz, devemos treinar um classificador com $N - 1$ amostras e testar com a restante. Isso é feito N vezes, variando o conjunto de treinamento. Por isso, essa técnica tem a desvantagem de um alto custo computacional. Sua estimativa é não viesada, mas tem grande variância.

O Método da Rotação ou *n-Fold Cross Validation* é um meio termo entre *Holdout* e *Leave-One-Out*. Ele divide o conjunto de dados em P subconjuntos disjuntos. Para treinar o classificador usamos $P - 1$ subconjuntos e o restante para teste. Isso é feito P vezes, variando o conjunto de teste utilizado e o de treinamento correspondente. Sua estimativa tem menor viés que o *Holdout* e tem menos custo computacional do que o *Leave-One-Out* (JAIN, DUIN *et al.*, 2000).

Finalmente, a probabilidade de erro estimada tende ao erro assintótico médio quanto mais amostras são utilizadas, ou seja, quando $N \rightarrow \infty$. Além disso, *Leave-One-Out* e *Resubstitution* são considerados limitantes superior e inferior, respectivamente, da probabilidade de erro médio no caso assintótico. Essas afirmações são ilustradas pela Figura 2.7, onde P_e é o erro assintótico médio.

2.4.2 Medidas de Avaliação do Desempenho

Para mensurar o desempenho de um sistema de reconhecimento de padrões, medidas de avaliação que permitam a comparação objetiva entre diversas técnicas devem ser usadas. Nesta seção, duas dessas medidas serão apresentadas: o coeficiente Kappa, que permite tam-

bém uma interpretação subjetiva e a medida simples e direta da Taxa de Reconhecimento Correto.

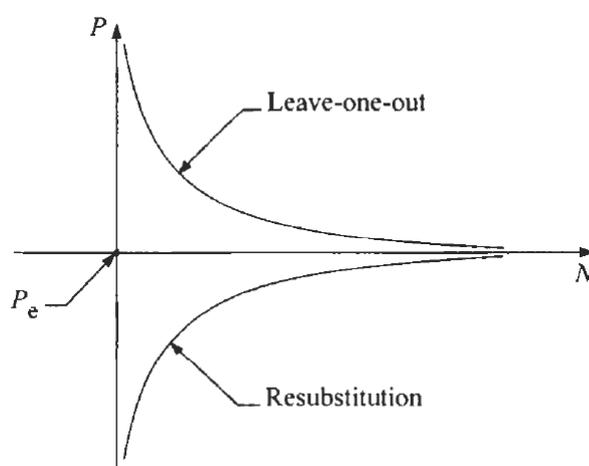


Figura 2.7. Tendência das probabilidades de erro dos métodos Leave-One-Out e Resubstitution no caso assintótico. Retirado de (THEODORIDIS e KOUTROUMBAS, 2003).

2.4.2.1 Coefficiente Kappa

O coeficiente Kappa ou estatística KHAT foi proposto originalmente por Cohen (COHEN, 1960). Esse coeficiente é uma medida de concordância baseada na diferença entre as concordâncias real e provável de um classificador e os dados de referência. Ele é resultado de uma análise Kappa que consiste numa “técnica multivariada discreta usada na avaliação de precisão para determinar estatisticamente se uma matriz de erro é significativamente diferente de outra” (CONGALTON, 1999). Portanto, para gerar o coeficiente Kappa precisamos gerar uma Matriz de Confusão ou Erro, sendo que a diagonal principal representará a concordância real e os totais das linhas e colunas (marginais) indicarão a concordância provável.

Uma Matriz de Confusão é uma matriz que contém k^2 células, nas quais n amostras são distribuídas, onde k é o número de classes. Normalmente, nas suas linhas são representados os dados classificados e nas colunas os dados de referência. Cada célula n_{ij} representa o número de amostras classificadas como a classe i quando a classe verdadeira é j , com $i, j = 1, \dots, k$. Isso implica que na diagonal principal teremos as amostras associadas à classe correta, enquanto fora dela estarão os erros de classificação (confusão). A Figura 2.8 ilustra uma matriz de confusão.

		j = Colunas			Total
		(Dados de			da
		Referência)			Linha
		1	2	k	n_{i+}
i = Linhas (Classificação)	1	n_{11}	n_{12}	n_{1k}	n_{1+}
	2	n_{21}	n_{22}	n_{2k}	n_{2+}
	k	n_{k1}	n_{k2}	n_{kk}	n_{k+}
Total da Coluna	n_{+i}	n_{+1}	n_{+2}	n_{+k}	n

Figura 2.8. Ilustração de uma matriz de confusão. Adaptado de (CONGALTON, 1999).

Tendo a matriz de confusão, o coeficiente Kappa \hat{K} pode ser determinado pela seguinte equação

$$\hat{K} = \frac{n \sum_{i=1}^k n_{ii} - \sum_{i=1}^k n_{i+} n_{+i}}{n^2 - \sum_{i=1}^k n_{i+} n_{+i}},$$

onde n_{i+} é a soma das colunas e n_{+i} é a soma das linhas, dadas por

$$n_{i+} = \sum_{j=1}^k n_{ij},$$

$$n_{+j} = \sum_{i=1}^k n_{ij}.$$

Como Kappa é normalmente distribuído, podemos calcular a sua variância $v\hat{a}r(\hat{K})$ definida pelas equações

$$v\hat{a}r(\hat{K}) = \frac{1}{n} \left\{ \frac{\theta_1(1-\theta_1)}{(1-\theta_2)^2} + \frac{2(1-\theta_1)(2\theta_1\theta_2 - \theta_3)}{(1-\theta_2)^3} + \frac{(1-\theta_1)^2(\theta_4 - 4\theta_2^2)}{(1-\theta_2)^4} \right\}$$

$$\theta_1 = \frac{1}{n} \sum_{i=1}^k n_{ii}$$

$$\theta_2 = \frac{1}{n^2} \sum_{i=1}^k n_{i+} n_{+i}$$

$$\theta_3 = \frac{1}{n^2} \sum_{i=1}^k n_{ii} (n_{i+} + n_{+i})$$

$$\theta_4 = \frac{1}{n^3} \sum_{i=1}^k \sum_{j=1}^k n_{ij} (n_{j+} + n_{+i})^2.$$

Há ainda uma interpretação subjetiva do desempenho de um classificador sobre o coeficiente Kappa que foi proposta por (LANDIS e KOCH, 1977) e é ilustrada na Tabela 2.1.

Tabela 2.1. Interpretação subjetiva do coeficiente Kappa sobre o desempenho de um classificador. Adaptado de (LANDIS e KOCH, 1977).

Kappa (\hat{K})	Interpretação sobre o desempenho do Classificador
< 0,0	Péssimo
0,0 – 0,2	Ruim
0,2 – 0,4	Razoável
0,4 – 0,6	Bom
0,6 – 0,8	Muito Bom
0,8 – 1,0	Excelente

2.4.2.2 Taxa de Reconhecimento Correto

A Taxa de Reconhecimento Correto η é uma medida simples e direta, dada por

$$\eta = \frac{\Psi}{N},$$

onde Ψ é o número de amostras de teste classificadas corretamente e N é o número de amostras no conjunto de testes.

2.4.3 Teste de Significância

Até agora, foi visto como particionar os dados a fim de obter uma estimativa de erro e como calcular uma medida objetiva que permite avaliar e comparar o desempenho de vários sistemas de reconhecimento de faces. Entretanto, para uma comparação mais completa e para determinar objetivamente se a superioridade de desempenho de um classificador sobre outro é significativa, testes de significância são utilizados.

Nesta seção, apresentaremos dois métodos para teste de significância, que utilizam dados normalmente distribuídos, como é o caso do Kappa. Esses métodos são as Estatísticas Z e T.

2.4.3.1 Estatística Z

A estatística Z (CONGALTON, 1999) se propõe a verificar se há diferenças significativas entre dois valores de coeficientes de Kappa \hat{K}_1 e \hat{K}_2 . Para isso, consideramos as seguintes hipóteses

$$H_0: \hat{K}_1 = \hat{K}_2$$

$$H_1: \hat{K}_1 \neq \hat{K}_2$$

para calcular a estatística Z, dada por

$$Z = \frac{|\hat{K}_1 - \hat{K}_2|}{\sqrt{\hat{v}ar(\hat{K}_1) + \hat{v}ar(\hat{K}_2)}}.$$

Considerando um nível de confiança de 95%, podemos afirmar que a diferença entre os coeficientes \hat{K}_1 e \hat{K}_2 será significativa se $Z > 1,96$. Dessa forma, a hipótese nula (H_0) será rejeitada.

2.4.3.2 Estatística T

A estatística T (MONTGOMERY e RUNGER, 2002), que apresenta uma análise global, verifica se há diferenças significativas entre duas médias de distribuições normais. Para isso, imagine que queremos comparar dois classificadores, sendo que para cada classificador,

Holdout com 7 tamanhos diferentes para conjuntos de treinamento e teste foram utilizados e calculados os respectivos coeficientes Kappa. Dessa forma, obteremos 7 coeficientes Kappa para cada classificador. A partir das médias \bar{K}_1 e \bar{K}_2 dos coeficientes Kappa de cada classificador, respectivamente, definimos $\bar{K} = \bar{K}_1 - \bar{K}_2$ e consideramos as seguintes hipóteses

$$H_0: \bar{K} = 0$$

$$H_1: \bar{K} < 0$$

para calcular a estatística T, dada por

$$T = \frac{\bar{K}}{\left(\sigma_d / \sqrt{n}\right)},$$

onde n é o número de coeficientes utilizados no cálculo das médias (no exemplo, $n = 7$) e σ_d é o desvio padrão das diferenças pontuais ($K_1^{(i)} - K_2^{(i)}$, onde $K_1^{(i)}$ e $K_2^{(i)}$ são o i -ésimo coeficiente Kappa dos respectivos classificadores). Considerando um nível de confiança de 95% ($\alpha = 0,05$), podemos afirmar que a diferença entre os coeficientes \bar{K}_1 e \bar{K}_2 será significativa se T for menor que um limiar, dado por uma distribuição T de *Student* considerando $(n - 1)$ graus de liberdade. Dessa forma, a hipótese nula (H_0) será rejeitada.

2.5 Considerações finais

Neste capítulo foi apresentada toda a base teórica de reconhecimento de padrões que será utilizada no restante do trabalho, especialmente, no sistema de reconhecimento de faces proposto no Capítulo 4.

Os diferentes métodos apresentados aqui formam o alicerce deste trabalho. Os métodos de extração de atributos como o PCA e suas variantes, o Block-Based PCA e o LDA, além da técnica de Regularização das Matrizes de Covariância das Classes para funcionar em conjunto com o classificador de Máxima Verossimilhança e ainda o classificador KNN, visam fornecer uma menor redundância dos dados e maior diversidade de fontes de informação para potencializar a combinação de classificadores. Assim, esses métodos permitem projetar um sistema de reconhecimento de faces com, possivelmente, desempenho superior e significativo comparado aos classificadores individuais que correspondem às técnicas mais usadas na lite-

ratura, como os *Eigenfaces* e *Fisherfaces*. Isso pode ser medido por meio dos métodos de estimação de erros e testes de significância também apresentados neste capítulo.

No próximo capítulo, é fornecida uma revisão bibliográfica sobre reconhecimento de faces englobando a definição do espaço de faces e os desafios dos sistemas de reconhecimento de faces, os principais bancos de dados, metodologias de avaliação e métodos encontrados na literatura e finalmente, um levantamento dos sistemas de reconhecimento de faces que utilizam combinação de classificadores.

3 RECONHECIMENTO DE FACES

Iniciada nos anos 70 e intensificada nos últimos 20 anos, a pesquisa em reconhecimento de faces tornou-se extremamente ativa, excitante e tem recebido grande atenção por parte dos pesquisadores, principalmente, graças ao baixo custo de dispositivos de hardware para captura e alto desempenho dos equipamentos para processar os dados brutos o que tornou esse tipo de aplicação viável e também devido à preocupação com segurança, especialmente no que diz respeito aos ataques terroristas e controle em transações e acessos.

Para um ser humano, reconhecer faces é uma tarefa fácil, sem muito esforço e importante para a vida em sociedade. Segundo (MIRHOSSEINI, YAN *et al.*, 1998), “pessoas podem reconhecer faces sobre iluminação pobre, cena complexa, de vários ângulos de visão e com variações de perspectivas”. Entretanto, para o computador essa ainda é uma tarefa desafiadora e complexa.

Um sistema de Reconhecimento de Faces consiste em um sistema biométrico, pois ele é um sistema de reconhecimento de padrões baseado em características fisiológicas ou comportamentais dos seres humanos (LUCAS, 1998; MARCIALIS e ROLI, 2004; ABATE, NAPPI *et al.*, 2007). Outros sistemas biométricos são, por exemplo, reconhecimento de íris e impressões digitais, sendo que reconhecimento de faces tem a vantagem em relação a outros sistemas biométricos, pois é discreto (LUCAS, 1998), não-intrusivo, não exige a colaboração do sujeito alvo, além do processo de aquisição ser mais simples e barato (MARCIALIS e ROLI, 2002). Entretanto, algoritmos de reconhecimento de faces são fortemente afetados pela alta variabilidade das condições ambientes durante a aquisição (MARCIALIS e ROLI, 2004).

Sistemas de reconhecimento de faces podem servir como base ou interface para muitas aplicações, principalmente, nas áreas de entretenimento (interações homem-máquina, realidade virtual, videogames, programas de treinamento), segurança (passaportes, controles de acesso, vigilância, rastreamento e investigação de suspeitos), telecomunicações, multimídia e outras aplicações de biometria (ZHAO, CHELLAPPA *et al.*, 2003).

Esse capítulo foi dividido em seções que transmitam uma idéia geral de sistemas de reconhecimento de faces. Para isso, na Seção 3.1 o espaço de faces é definido e os componentes de um sistema completo de reconhecimento de faces e os desafios são apresentados. Nas

seções seguintes são apresentados alguns dos bancos de dados de faces mais utilizados (Seção 3.2), metodologias de avaliação de desempenho para sistemas de reconhecimento de face e sua importância para a área (Seção 3.3), um conjunto das principais e mais comuns abordagens para reconhecimento de faces utilizando especialistas individuais (Seção 3.4) e finalmente, um *survey* sobre combinação de classificadores para reconhecimento de faces frontais é apresentado na Seção 3.5.

3.1 Definição do Problema e Sistema Completo de Reconhecimento de Faces

Segundo (MU, WATTA *et al.*, 2005), “um problema de reconhecimento de faces pode ser formulado do ponto de vista de reconhecimento de padrões, sendo que cada pessoa no banco de dados pode ser considerada como uma classe e a tarefa é projetar um sistema para mapear uma imagem de entrada para uma das classes (ou então rejeitar a entrada)”. Entretanto, essa é apenas uma parte de um sistema completo de reconhecimento de faces. Como mostrado na Figura 3.1, um sistema completo de reconhecimento de faces é composto por um módulo de detecção ou rastreamento de faces de uma imagem ou vídeo de entrada, respectivamente. Caso exista alguma face, a sua localização, pose e escala devem ser determinadas e fornecidas para que numa etapa de alinhamento sejam detectados os componentes faciais como olhos, nariz, boca e contorno facial para permitir à face detectada ser normalizada de forma mais fina com relação à escala e pose usando transformações geométricas. Ainda nessa etapa, geralmente, há um pré-processamento para normalização com relação a medidas fotométricas como iluminação e nível de cinza. Com a face alinhada, uma etapa de extração de atributos pode ser feita com a finalidade de reduzir a dimensionalidade e manter apenas informações importantes para que numa etapa de *matching* permita-se determinar a classe ou pessoa à qual pertence aquela face com confiança suficiente ou então rejeitá-la utilizando um banco de dados de faces conhecidas. A saída da etapa de *matching* é a saída do sistema. A maioria dos trabalhos sobre reconhecimento de face encontrados na literatura concentra-se nas etapas de extração de atributos e *matching*, ou seja, considera que as faces já foram detectadas e normalizadas (LI e JAIN, 2005).

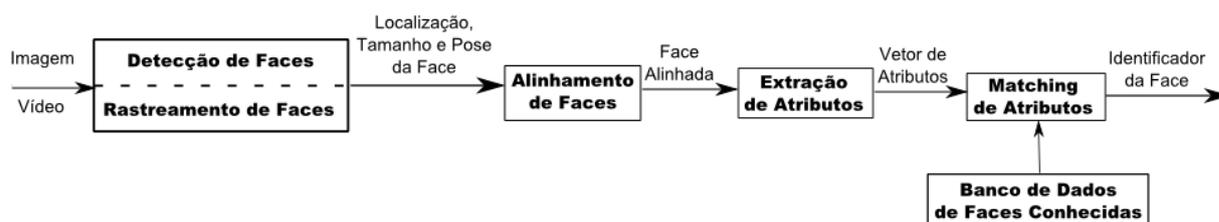


Figura 3.1. Sistema Completo de Reconhecimento de Faces. Adaptado de (LI e JAIN, 2005).

As aplicações de reconhecimento de faces podem ser divididas em dois tipos: autenticação e identificação. Em aplicações de autenticação temos um problema de um-para-um onde uma pessoa fornece a sua imagem facial e declara a sua identidade. O objetivo consiste em determinar se uma imagem facial corresponde ou não à identidade sendo declarada. Para isso, uma medida de similaridade é calculada entre a imagem de entrada e o modelo da classe declarada que está armazenado no banco de dados de faces conhecidas e verifica se ultrapassa ou não um limiar. Caso a similaridade esteja dentro do limiar, a face é declarada sendo genuína. Senão, é uma impostora. Geralmente, para esse tipo de aplicação a imagem facial foi adquirida em ambientes controlados. Para as aplicações de identificação, temos um problema de um-para-muitos onde uma dada face desconhecida é fornecida ao sistema que determinará a classe ou identidade correspondente analisando todas as faces armazenadas no banco de dados de faces conhecidas e selecionando as mais próximas segundo algum critério (MARCIALIS e ROLI, 2004).

Para conhecer melhor a distribuição ou subespaço de faces, Li e Jain (LI e JAIN, 2005) utilizaram PCA sobre um conjunto de faces. O espaço de imagens pode ser dividido em subespaço de todas as faces humanas que considera as variações em aparência e o subespaço de não-faces que engloba o restante do espaço de imagens. Foi identificado que esses subespaços são altamente não-lineares e não-convexos. A Figura 3.2a, ilustra a divisão do espaço de imagens, enquanto a Figura 3.2b ilustra os subespaços de dois indivíduos diferentes dentro do subespaço de faces. Basicamente, a Figura 3.2a e a Figura 3.2b representam as tarefas da detecção e reconhecimento de faces, respectivamente. Devido à complexidade desses espaços, o uso de técnicas lineares para a extração de atributos pode não ser eficiente, pois não preserva as variações não-convexas que são necessárias para diferenciar entre indivíduos. Dessa forma, do ponto de vista euclidiano, uma distância interpessoal pode ser menor do que uma intrapessoal, como pode ser visto pela Figura 3.3a.

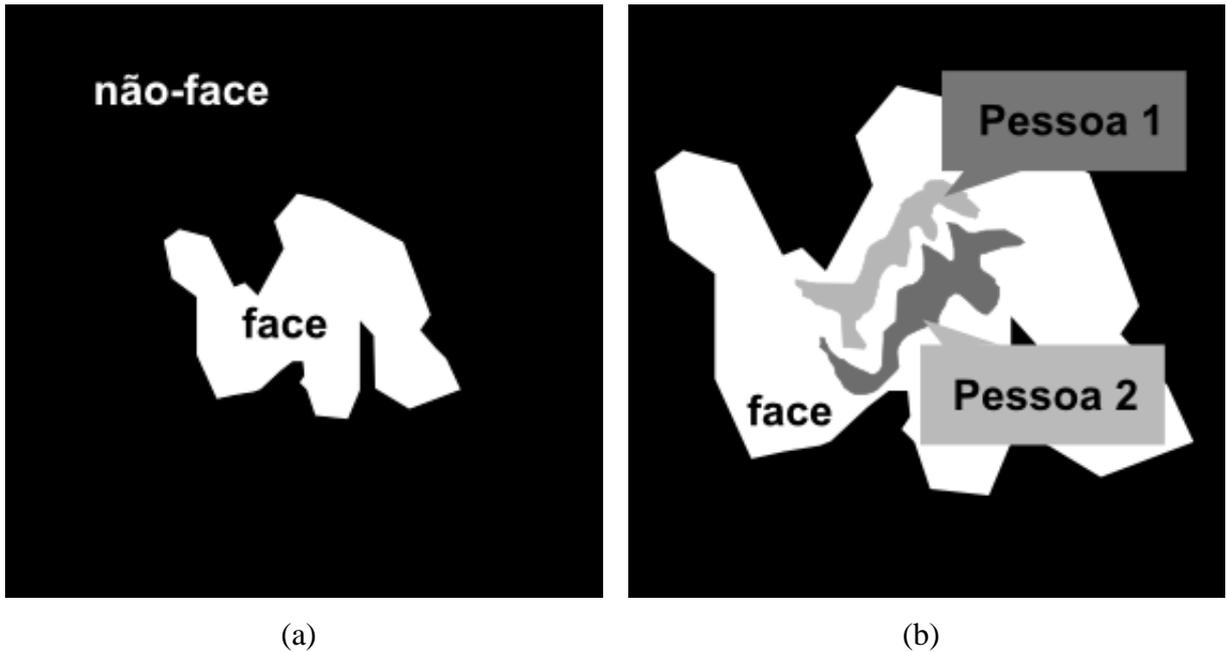


Figura 3.2. O subespaço de faces: a) num contexto de espaço de imagens e b) a sua divisão interna com subespaços de indivíduos diferentes. Adaptado de (LI e JAIN, 2005).

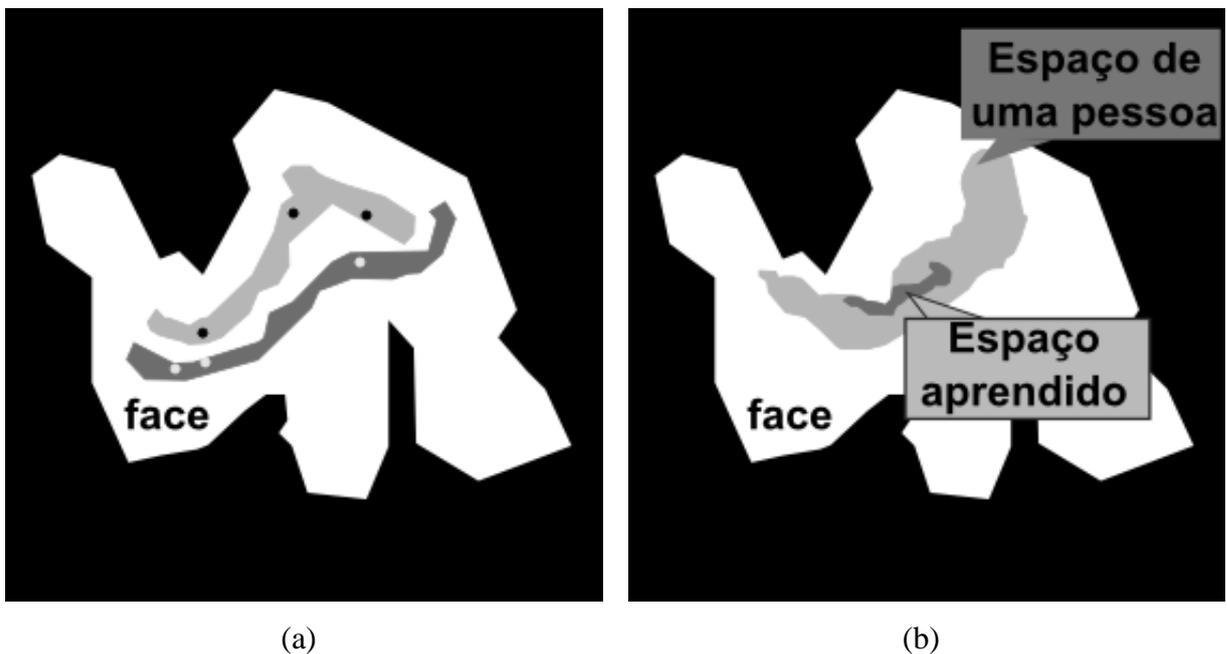


Figura 3.3. Desafios em sistemas de reconhecimento de faces do ponto de vista de subespaços de faces: a) Distância euclidiana não é capaz de diferenciar entre dois indivíduos, pois algumas distâncias interpessoais são menores que intrapessoais e b) O subespaço aprendido não é suficiente para cobrir todo o subespaço de um indivíduo, ou seja, representa a incapacidade de generalização. Adaptado de (LI e JAIN, 2005).

Além da dificuldade inerente do subespaço de faces, os maiores desafios enfrentados no projeto de sistemas de reconhecimento de faces consistem no grande número de fatores

que afetam o sistema, principalmente, devido à sua variabilidade ou complexidade. Caso as condições de aquisição não sejam controladas, ou seja, não existe um controle com relação às condições ambientais ou comportamentais do indivíduo, o que se traduz em maior proximidade com o que se encontra no mundo real e torna esse tipo de aplicação mais complexo. Os fatores-chave a serem considerados são:

- **Iluminação,**
- **Pose,**
- **Expressão,**
- **Envelhecimento e**
- **Oclusões.**

A variação na iluminação é gerada por propriedades de refletância da pele e controle interno da câmera (ABATE, NAPPI *et al.*, 2007). Pose é devido ao angulo de visão que pode variar com relação às 3 dimensões como apresentado na Figura 3.4. Mudanças na pose podem gerar deformações projetivas e auto-oclusões (ABATE, NAPPI *et al.*, 2007) como pode ser percebido com faces próximas à visão de perfil, onde um dos olhos não aparece na imagem. Expressão altera a geometria da face, reduzindo o número de características invariantes (LI, MORI *et al.*, 2006). O problema do envelhecimento ocorre pela tendência que as faces apresentam de mudar durante o tempo e de forma não-linear (ABATE, NAPPI *et al.*, 2007). E finalmente, oclusões podem afetar características essenciais para identificação de faces, como a região dos olhos, nariz e boca e podem ser geradas também por acessórios como óculos.

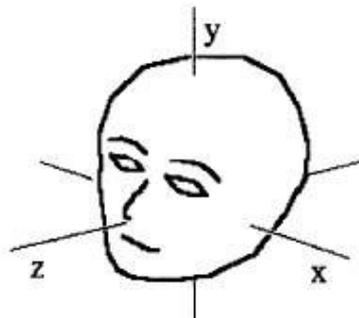


Figura 3.4. As possíveis orientações de uma face que caracterizam mudanças na pose. Adaptado de (JENG, LIAO *et al.*, 1998).

Outros fatores a se considerar são fundo complexo, alta dimensionalidade dos dados e poucas amostras de treinamento por indivíduo. Fundos complexos podem prejudicar a detecção facial ou seus componentes enquanto que a alta dimensionalidade dos dados e o número insuficiente de amostras de treinamento prejudicam a capacidade de generalização de um sistema de reconhecimento de faces. Considerando a grande variação nas condições de aquisição das faces e a alta dimensionalidade e complexidade do subespaço de faces, um grande número de amostras de treinamento por indivíduo se faz necessário para uma boa generalização e robustez do sistema. Entretanto, os conjuntos de treinamento apresentam número finito e insuficiente de amostras por pessoa (normalmente, menos de 10), correspondendo a apenas uma pequena porção do subespaço daquela pessoa como mostrado na Figura 3.3b, gerando outros problemas como os descritos na seção 2.2.2. Esta situação poderá ser verificada na próxima seção sobre bancos de dados de faces que são utilizados para a avaliação do desempenho dos sistemas de reconhecimento de faces.

3.2 Bancos de Dados de Faces

Devido à grande variação na aparência e nos fatores que afetam os sistemas de reconhecimento de faces, necessitamos de bancos de dados que forneçam exemplos suficientes dessas variações a fim de permitir o desenvolvimento de algoritmos robustos e invariantes a esses fatores, além do que a existência de bancos de dados públicos de faces permite comparar os diversos algoritmos projetados.

Nessa seção, apresentaremos os principais e mais comuns bancos de dados para reconhecimento de faces disponíveis (FERET, ORL, AR, Yale, YaleB, CMU PIE) (ABATE, NAPPI *et al.*, 2007), sendo que alguns serão utilizados nesse trabalho, além de outros que foram citados nesse documento, em especial na revisão realizada na Seção 3.5. Ao todo, 8 bancos de dados serão apresentados aqui.

3.2.1 FERET

O banco de dados FERET foi desenvolvido como parte do programa de Tecnologia de Reconhecimento Facial (FERET) que surgiu de um programa patrocinado pelo Departamento de Defesa americano através do Programa de Tecnologias Anti Drogas. Ele contém um total

de 14.051 faces em tons de cinza de 1.199 indivíduos com resolução 256x384 pixels. Essas imagens foram adquiridas com câmeras de 35 mm. As amostras para cada indivíduo são distribuídas em variações de expressão facial, iluminação, pose e idade com 2, 2, 9-20 e 2 condições, respectivamente. Alguns subconjuntos do FERET são geralmente utilizados. Fa e Fb correspondem a duas visões frontais, mas para Fb existe variação na expressão facial. Fc contém imagens tomadas com uma câmera e condições de iluminação diferentes. Duplicate I e Duplicate II foram adquiridas em tempos diferentes, onde passaram em média 251 e 627 dias, respectivamente. Há ainda alguma sobreposição entre os subconjuntos Duplicate. O restante das categorias englobam variações na pose que variam de $+60^\circ$ a -60° (GROSS, 2005). Esse banco de dados pode ser encontrado publicamente em <http://face.nist.gov/colorferet/>. Algumas amostras são apresentadas na Figura 3.5.

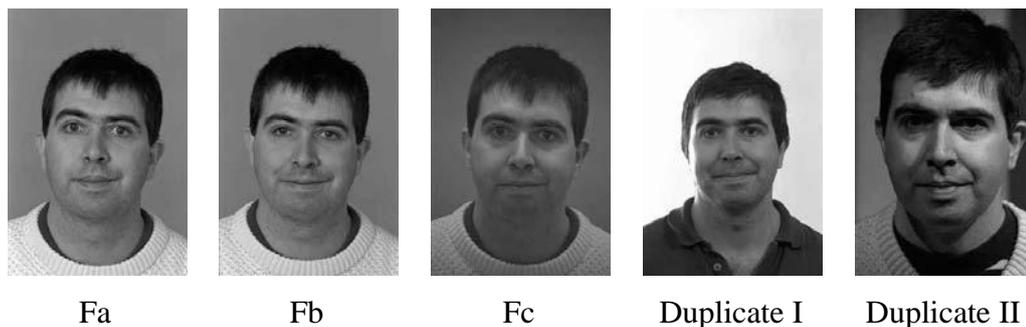


Figura 3.5. Algumas amostras de faces frontais do banco de dados FERET: Fa, Fb, Fc, Duplicate I e Duplicate II são nomes de algumas das principais categorias de divisão do FERET. Retirado de (GROSS, 2005).

3.2.2 AR

O banco de dados AR contém pouco mais de 3.000 imagens faciais coloridas (RGB) de tamanho 768 x 576 pixels de 126 indivíduos (70 homens e 56 mulheres), sendo 26 amostras por classe distribuídas em variações de expressão facial (neutro, sorrindo, zangado, gritando), iluminação, oclusão e idade com 4, 4, 2 e 2 condições, respectivamente (MARTÍNEZ e BENAVENTE, 1998; GROSS, 2005; ABATE, NAPPI *et al.*, 2007). Foram realizadas duas seções com intervalos de duas semanas. Algumas amostras do banco de dados são mostradas na Figura 3.6. Esse banco de dados pode ser encontrado publicamente em http://cobweb.ecn.purdue.edu/~aleix/aleix_face_DB.html.



Figura 3.6. Algumas amostras de um sujeito do banco de dados AR. As duas primeiras linhas representam a primeira seção e as restantes a segunda seção. Retirado de (MARTÍNEZ e BENAVENTE, 1998).

3.2.3 ORL

O banco de dados ORL (Laboratório de Pesquisa da Olivetti) foi coletado entre 1992 e 1994. Ele contém um total de 400 imagens faciais em tons de cinza de tamanho 92 x 112 pixels de 40 indivíduos (4 mulheres e 36 homens), sendo 10 amostras por classe variando suavemente em expressão facial (olhos abertos/fechados, sorrindo/não sorrindo), iluminação e detalhes faciais (óculos ou não) (GROSS, 2005). Algumas amostras do banco de dados são mostradas na Figura 3.7. Esse banco de dados pode ser encontrado publicamente em <http://www.cl.cam.ac.uk/Research/DTG/attarchive/facedatabase.html>.



Figura 3.7. Algumas amostras de um sujeito do banco de dados ORL. Retirado de (ZUO, WANG *et al.*, 2007).

3.2.4 CMU PIE

O banco de dados CMU PIE (*Pose, Illumination and Expression*) contém 41.368 imagens faciais coloridas (RGB) de tamanho 640 x 486 pixels de 68 indivíduos, sendo aproximadamente 608 amostras por classe distribuídas em variações de expressão facial, iluminação e pose com 4, 43, 13 condições, respectivamente (GROSS, 2005; ABATE, NAPPI *et al.*, 2007). Algumas amostras do banco de dados são mostradas na Figura 3.8. Esse banco de dados pode ser encontrado publicamente em http://www.ri.cmu.edu/projects/project_418.html.



Figura 3.8. Algumas amostras de um sujeito do banco de dados CMU PIE. Essas amostras foram recortadas a fim de eliminar o fundo e conter somente a face do indivíduo. Retirado de (HUANG, YUEN *et al.*, 2006).

3.2.5 XM2VTS

O banco de dados XM2VTS é um banco de dados multimodal muito utilizado para sistemas de verificação/autenticação de identidade. Ele contém imagens faciais coloridas (RGB) de tamanho 720 x 576 pixels de 295 indivíduos variando somente em pose e idade (foram tomadas em 4 sessões) (GROSS, 2005; ABATE, NAPPI *et al.*, 2007). Algumas amostras do banco de dados são ilustradas na Figura 3.9. Esse banco de dados não é gratuito. Informações podem ser encontradas em <http://www.ee.surrey.ac.uk/CVSSP/xm2vtsdb/>.

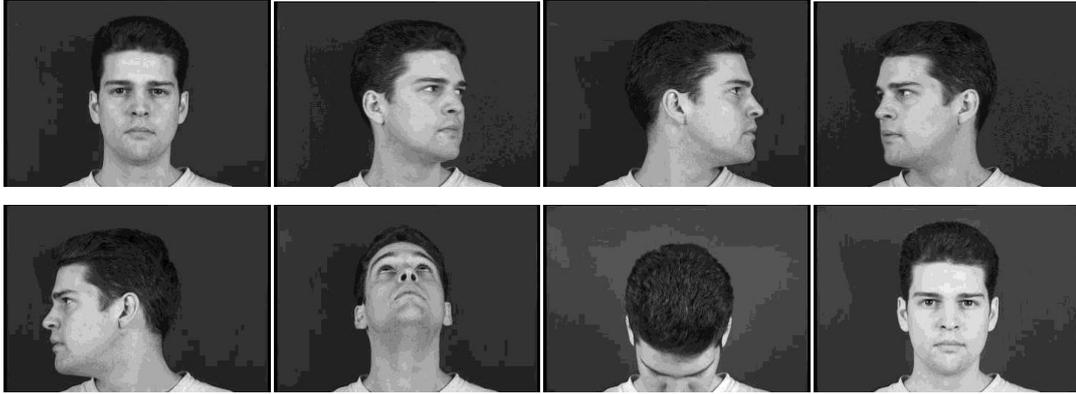


Figura 3.9. Algumas amostras de um sujeito do banco de dados XM2VTS. Retirado de (MESSER, MATAS *et al.*, 1999).

3.2.6 Yale

O banco de dados Yale contém um total de 165 imagens faciais em tons de cinza de tamanho 320 x 243 pixels de 15 indivíduos, sendo 11 amostras por classe distribuídas em variações de e sem óculos, iluminação e expressão facial com 3 e 6 amostras para as últimas duas variações. A iluminação varia em luz central, direita e esquerda. As expressões encontradas são normal, feliz, triste, sonolento, surpreso e piscando. Algumas amostras do banco de dados são mostradas na Figura 3.10. Esse banco de dados pode ser encontrado publicamente em <http://cvc.yale.edu/projects/yalefaces/yalefaces.html>.



Figura 3.10. Algumas amostras de um sujeito do banco de dados Yale. Retirado de <http://cvc.yale.edu/projects/yalefaces/yalefaces.html>.

3.2.7 YaleB

O banco de dados YaleB contém um total de 5.760 imagens faciais em tons de cinza de tamanho 640 x 480 pixels de 10 indivíduos, sendo 576 amostras por classe distribuídas em variações de expressão facial e iluminação com 9 poses \times 64 condições de iluminação. As poses e fontes de iluminação variam em até 24° e 77°, respectivamente (GROSS, 2005). Algumas amostras do banco de dados são mostradas na Figura 3.11. Esse banco de dados pode ser encontrado publicamente em <http://cvc.yale.edu/projects/yalefacesB/yalefacesB.html>.

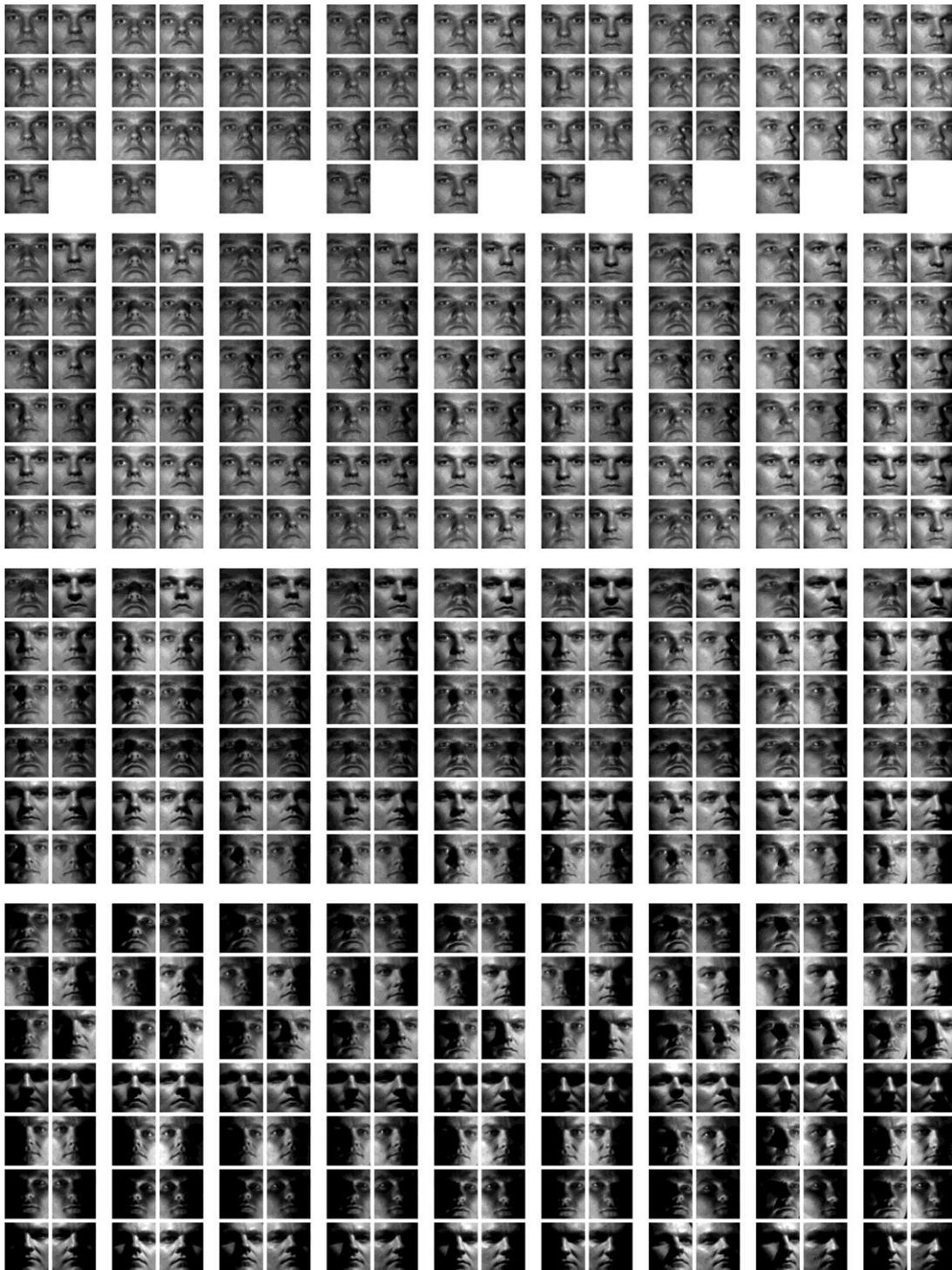


Figura 3.11. Algumas amostras de um indivíduo do banco de dados YaleB. Retirado de (GEORGIADES, BELHUMEUR *et al.*, 2001).

3.2.8 MIT

O banco de dados MIT contém 2.500 imagens faciais em tons de cinza de tamanho 512 x 512 pixels de 16 indivíduos (todos homens), variando entre pose, escala e iluminação com 3 condições para cada variação (TURK e PENTLAND, 1991; MARCIALIS e ROLI, 2004). Além disso, as imagens apresentam fundo complexo. Algumas amostras do banco de dados são mostradas na Figura 3.12.



Figura 3.12. Algumas amostras de um sujeito do banco de dados do MIT. Retirado de (MARCIALIS e ROLI, 2004).

3.3 Metodologias de Avaliação FERET e FRVT 2006

Metodologias de avaliação com seus protocolos de teste são extremamente importantes para permitir uma avaliação comparativa entre os diversos algoritmos desenvolvidos e assim, dar suporte à evolução da tecnologia e da área. Para reconhecimento de faces várias metodologias e testes comparativos foram propostos, sendo os patrocinados pelo Governo Americano os mais conhecidos, como é o caso do FERET (PHILLIPS, MOON *et al.*, 2000) e do FRVT (*Face Recognition Vendor Test*) (PHILLIPS, SCRUGGS *et al.*, 2007) em suas várias versões. Aliados aos métodos de avaliação, um banco de dados grande em número de indivíduos e que considera diferentes cenários é essencial. Alguns exemplos de cenários poderiam ser: 1) galerias e provas correspondendo a imagens tiradas no mesmo dia e 2) galerias e provas correspondendo a imagens tiradas em sessões diferentes, com intervalos de tempo sendo considerados. Dessa forma, uma metodologia de avaliação deve além de propor um protocolo de teste, definir também o banco de dados a ser utilizado.

Os principais objetivos com essa avaliação comparativa são avaliar o estado da arte, identificar áreas ou cenários futuros de pesquisa e medir o desempenho dos algoritmos (PHILLIPS, MOON *et al.*, 2000).

Para permitir essa comparação, protocolos de teste são definidos. Entre outras coisas, “um protocolo de testes estabelece o número de imagens de cada pessoa no teste, como as

saídas dos algoritmos são registradas e como os resultados de desempenho são relatados” (PHILLIPS, MOON *et al.*, 2000).

Para os casos específicos do FERET (1996) e FRVT 2006 (último grande teste realizado), algumas definições sobre nomes dados a alguns subconjuntos do banco de dados devem ser definidos que são galeria (gallery), prova (probe), alvo (target) e consulta (query). Galeria e prova correspondem ao conjunto de indivíduos conhecidos e os desconhecidos que descreverão um cenário e serão apresentados aos algoritmos e usados na computação do desempenho, respectivamente. Entretanto, os conjuntos de galeria e prova são subconjuntos de alvo e consulta, respectivamente. Alvo e consulta englobam imagens faciais de todos os cenários a serem analisados.

O FERET propôs que deveriam ser calculadas medidas de similaridade entre todas as imagens nos conjuntos alvo e consulta. Para relatar o desempenho, foi proposta a utilização de um gráfico de *Cumulative Match Score*, onde no eixo das abscissas temos os *ranks* e no eixo das ordenadas a probabilidade de identificação, ou seja, indica a probabilidade de a classe correta (ou imagem da galeria correspondente à classe correta) estar entre as *n* faces mais similares a uma dada amostra de prova.

Do FERET ao FRVT 2006 foi identificada uma grande evolução dos algoritmos de reconhecimento de face no que diz respeito à redução da taxa de rejeição falsa e robustez contra variações. No FERET havia sido relatado que os trabalhos futuros deveriam focar na solução ou compensação de variações nas condições, em especial a iluminação (PHILLIPS, MOON *et al.*, 2000). Enquanto, que finalmente, o FRVT 2006 relatou e identificou que isso fora feito (PHILLIPS, SCRUGGS *et al.*, 2007). Para se ter uma idéia da evolução dos algoritmos de reconhecimento de faces, um teste realizado durante o FRVT 2006 comparou o desempenho dos algoritmos de reconhecimento de faces com a capacidade dos humanos e foi relatado um desempenho superior (no que diz respeito à taxa de rejeição falsa) dos algoritmos sobre humanos para reconhecer faces focando no problema de mudança de condições de iluminação.

O FRVT 2006 também analisou imagens com mais alta resolução e um banco de dados crescente se analisado desde o FERET. Foi identificado que resoluções maiores contribuem significativamente para o aumento de desempenho. Assim, uma das questões-chaves e um dos desafios para a próxima geração de algoritmos de reconhecimento de faces será a aquisição e utilização de imagens de resolução mais alta (PHILLIPS, SCRUGGS *et al.*, 2007). Dessa forma, técnicas que utilizem super resolução ou *halucination* (como é conhecida a super

resolução no contexto de reconhecimento de faces) (WANG e TANG, 2003; LIU, SHUM *et al.*, 2007), que podem gerar imagens de mais alta resolução partindo de um conjunto de imagens de resoluções menores, podem ser promissoras para resolver esse problema da resolução e devem ganhar um destaque crescente nos algoritmos de reconhecimento de faces.

3.4 Principais Métodos

Na literatura, podem ser encontrados diversos algoritmos de reconhecimento de faces. Entretanto, alguns se destacam, especialmente pelo balanço entre simplicidade e altas taxas de reconhecimento correto. Basicamente, essas técnicas utilizam reconhecimento de padrões estatísticos e são classificadores individuais.

Nessa seção apresentaremos quatro algoritmos que podem ser considerados os principais e mais comuns, sendo muito utilizados na comparação com novos algoritmos sendo propostos. Para uma visão geral das técnicas de reconhecimento de faces sugerimos a leitura de alguns surveys (SAMAL e IYENGAR, 1992; CHELLAPPA, WILSON *et al.*, 1995; ZHAO, CHELLAPPA *et al.*, 2003; ABATE, NAPPI *et al.*, 2007).

3.4.1 Eigenfaces (PCA)

Eigenfaces correspondem aos autovetores ou vetores base obtidos aplicando a técnica de extração de atributos PCA. As imagens a serem classificadas e as da galeria são centralizadas e projetadas no espaço PCA. Assim, cada imagem é representada por um vetor de pesos que será utilizado para realizar a classificação, normalmente, por Vizinho Mais Próximo. Alguns exemplos de eigenfaces são mostradas na Figura 3.13. Essa técnica é a mais popular e foi proposta nos trabalhos (KIRBY e SIROVICH, 1990; TURK e PENTLAND, 1991). Foi identificado que os Eigenfaces são bons contra pequenas oclusões (ZHAO, CHELLAPPA *et al.*, 2003), mas retém muito da variação de iluminação (BELHUMEUR, HESPANHA *et al.*, 1997; MARTINEZ e KAK, 2001). Além disso, foi mostrado que a técnica de eigenfaces ultrapassa o desempenho da técnica de fisherfaces quando temos poucas amostras (MARTINEZ e KAK, 2001).

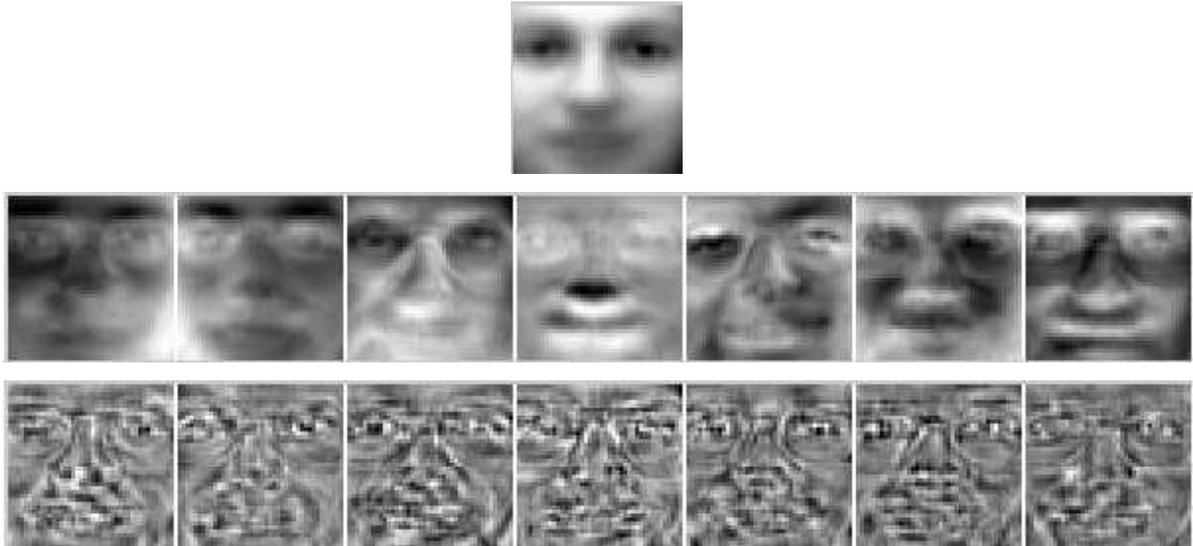


Figura 3.13. Exemplos de Eigenfaces sobre o banco de dados ORL. Na primeira linha temos a imagem média, na segunda os 7 autovetores correspondendo aos 7 maiores autovalores e na última linha os 7 autovetores correspondendo aos 7 menores autovalores. Retirado de (LU, 2003).

3.4.2 Fisherfaces (LDA)

A técnica de Fisherfaces consiste basicamente em selecionar alguns vetores base no espaço LDA. Esses vetores que formam a base LDA são os Fisherfaces. As imagens a serem classificadas e as da galeria são centralizadas e projetadas no espaço LDA. Assim, cada imagem é representada por um vetor de pesos que será utilizado para realizar a classificação, normalmente, por Vizinho Mais Próximo. Alguns exemplos podem ser vistos na Figura 3.14. Vale ressaltar que essa técnica sofre com o problema do Small Sample Size. Para solucionar esse problema, muitas vezes o LDA é aplicado após a transformação das amostras para um espaço PCA. Entretanto, foi verificado que juntando essas duas técnicas, informação discriminante é perdida (ABATE, NAPPI *et al.*, 2007). Ainda, foi identificado que essa técnica lida bem com variação em condições de iluminação e expressão facial (BELHUMEUR, HESPANHA *et al.*, 1997).



Figura 3.14. Exemplos dos primeiros Fisherfaces sobre o banco de dados ORL. Retirado de (LU, 2003).

3.4.3 Análise de Componentes Independentes (ICA)

ICA é uma técnica de extração de atributos que pode ser considerada uma generalização do PCA, que procura encontrar vetores base que são estatisticamente independentes considerando uma distribuição não Gaussiana. Com isso ICA possui três vantagens sobre o PCA (BARTLETT, MOVELLAN *et al.*, 2002; ABATE, NAPPI *et al.*, 2007): 1) melhor representação dos dados no espaço n -dimensional; 2) As bases encontradas não são necessariamente ortogonais, o que implica em erro menor de reconstrução na presença de ruído e 3) como consideram estatísticas de ordem mais alta, além da segunda (que é dada pela matriz de covariância no caso do PCA), mais informações discriminantes são capturadas. Estatísticas de primeira e segunda ordem consideram somente a amplitude do espectro e descartam a informação de fase do espectro (BARTLETT, MOVELLAN *et al.*, 2002; ABATE, NAPPI *et al.*, 2007). Entretanto, alguns experimentos têm identificado que a capacidade humana para reconhecer faces e outros objetos considera, principalmente, a informação de fase do espectro (BARTLETT, MOVELLAN *et al.*, 2002; ABATE, NAPPI *et al.*, 2007). Ainda, por obter um menor erro de reconstrução na presença de ruído, ICA tem sido uma solução robusta sobre o problema de iluminação e outras variações como expressão facial. Para uma comparação entre PCA e LDA, a Figura 3.15 apresenta alguns vetores base extraídos por ICA.



Figura 3.15. Exemplo de alguns vetores base ICA gerados sobre o banco de dados ORL. Retirado de (LU, 2003).

3.4.4 Máquinas de Vetores de Suporte (SVM)

Máquinas de Vetores de Suporte ou *Support Vector Machines* foram propostas por (VAPNIK, 1995). SVM considera um problema de duas classes e procura num espaço de dimensão mais alta um hiperplano linear capaz de separar as classes. Entretanto, uma superfície de decisão não linear pode ser obtida utilizando um Kernel adequado que satisfaça a condição de Mercer (VAPNIK, 1995). Como o SVM é um problema binário e a aplicação de reconhecimento de faces consiste num problema de n classes, o problema deve ser modelado de for-

ma diferente. Em (PHILLIPS, 1999) foi proposto utilizar as imagens representadas num espaço de dissimilaridades. Assim, uma única SVM é treinada, sendo as duas classes: diferenças entre faces de um mesmo indivíduo e diferenças entre faces de indivíduos diferentes do conjunto de treinamento. Dessa forma, a saída do SVM será uma medida de dissimilaridade, sendo que quanto menor o valor melhor. Assim, para uma dada face a ser reconhecida, sua dissimilaridade é calculada para cada indivíduo na galeria e essa dissimilaridade será a entrada do SVM. A classe que minimizar a saída do SVM será a associada ao padrão de entrada. Além disso, PCA foi utilizado antes de realizar o treinamento dos Vetores de Suporte.

3.5 Combinando Classificadores para Reconhecimento de Faces Frontais Bidimensionais: Uma Visão Geral

Como pôde ser visto nas seções anteriores, os espaços de faces são altamente complexos devido aos fatores ambientais no momento da aquisição, principalmente. Dessa forma, utilizar um algoritmo simples, individual para manipular todas essas variações é impraticável e não robusto contra todos os tipos de variações. Os algoritmos diferentes de reconhecimento de face encontrados na literatura são normalmente baseados em domínios de representação e atributos diferentes sendo que cada algoritmo apresenta-se com capacidades diferentes de lidar com cada fator de variação. Isso fornece uma idéia de complementariedade, já que as amostras classificadas erroneamente podem não ser sobrepostas. Portanto, o uso de uma técnica de combinação de classificadores pode aumentar o desempenho geral do sistema aproveitando-se das vantagens de cada especialista individual (HUANG, YUEN *et al.*, 2006). Com base nisso, nos últimos 15 anos e com mais intensidade nos últimos 5 anos, diversas abordagens utilizando combinação de classificadores foram propostas na literatura.

Baseados na divisão realizada por Zhao e seus colegas (ZHAO, CHELLAPPA *et al.*, 2003) para as técnicas de Reconhecimento de Faces, podemos dividir as técnicas de combinação de classificadores para reconhecimento de face de maneira semelhante:

- **Holísticas:** que utilizam a face inteira como entrada para o sistema;
- **Baseadas em Componentes:** que combina os classificadores de diversos componentes faciais como olhos, nariz e boca ou áreas diferentes extraídas da imagem original;
- **Híbridas:** que junta as duas categorias anteriores, ou seja, a combinação é realizada unindo classificação de componentes faciais e face inteira.

Nas subseções seguintes, as diferentes abordagens holísticas (3.5.1), baseadas em componentes (3.5.2) e híbridas (3.5.3) são descritas considerando o reconhecimento de face frontal 2D.

3.5.1 Técnicas Holísticas

Em (ARBUCKLE, LANGE *et al.*, 1995) é apresentada uma técnica que combina imagens em diversas resoluções. Para isso, inicialmente, das imagens adquiridas com tamanho 160 x 120 pixels são geradas imagens novas com resoluções menores aplicando a média sobre blocos de 2 x 2, 3 x 3, 4 x 4 e 5 x 5 pixels da imagem, obtendo-se assim quatro imagens novas com 80 x 60, 53 x 40, 40 x 30 e 32 x 24 pixels, respectivamente. Em seguida, são gerados vetores de atributos descritivos das imagens pela concatenação de coeficientes de autocorrelação local de ordem zero, um e dois (Figura 3.16), aplicando a equação

$$C_K = \frac{1}{N(K)} \sum_{i=2}^{I-1} \sum_{j=2}^{J-1} \prod_{n=0}^{O(K)} f_{i+i_n, j+j_n},$$

onde f é uma imagem em 256 tons de cinza, I e J são as dimensões da imagem, K são os kernels de autocorrelação definidos como conjuntos de $O(K) + 1$ pares que representam um deslocamento do pixel central definidos por $(i_n, j_n) = \{-1, 0, 1\} * \{-1, 0, 1\}$, $O(K)$ é a ordem de autocorrelação e $N(K)$ é uma constante de normalização definida por

$$N(K) = \frac{1}{(I-2)(J-2)(2^8-1)^{O(K)+1}}.$$

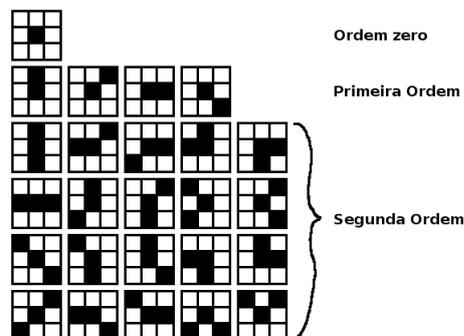


Figura 3.16. Kernels de autocorrelação de zero a segunda ordem. Adaptado de (ARBUCKLE, LANGE *et al.*, 1995).

A etapa de treinamento consiste de dois passos:

- 1) Para cada resolução separadamente, é aplicado o algoritmo *clustering* LDA sobre os dados do conjunto de treinamento a fim de obter matrizes de mapeamento para um espaço que maximiza a variância interclasses enquanto minimiza a variância intraclasses. A seguir, com os dados de treinamento mapeados nos seus respectivos espaços LDA, são encontrados o vetor média (centro) de cada classe;
- 2) Esta etapa pode ser sub-dividida em outras duas:
 - a) Utilizando o conjunto de teste mapeado no espaço LDA determinado na etapa anterior é realizada a classificação das faces calculando a distância Euclidiana entre cada vetor de teste e o centro de cada classe (obtidos no passo anterior). Para cada vetor de teste, as duas menores distâncias $d1$ e $d2$, respectivamente, são utilizadas para calcular a medida de contraste dada por

$$cont = \frac{d2 - d1}{d1 + d2}.$$

Essa medida será utilizada para verificar a aceitação ou rejeição da classificação de acordo com um limiar. Se a medida for maior que o valor do limiar, essa amostra de teste é classificada como pertencente à classe que gerou a menor distância. Caso contrário, essa amostra é rejeitada. Dessa forma, conseguimos calcular as taxas de reconhecimento correto (*rec*) e aceitação falsa (*far*) para cada resolução;

- b) Finalmente, utilizando as taxas obtidas no passo anterior podemos calcular as densidades de medida Fuzzy usando a seguinte equação

$$g_{ij} = 1 - \sqrt{\frac{(far_{ij}^2 + (1 - rec_{ij})^2)}{2}}.$$

Durante a etapa de teste, inicialmente, as densidades de medida Fuzzy são escaladas utilizando ou a soma de todos os elementos de g_{ij} ou a soma de todos os g_i e as distâncias d_i^k de cada vetor do conjunto de teste mapeado no espaço LDA para os centros de cada classe i e cada resolução k são normalizadas utilizando a equação

$$h_i^k = \frac{1/d_i^k}{\sum_{i=1}^p 1/d_i^k},$$

onde p é o número de distâncias indexadas pela classe i . Dessa forma, obtemos valores na faixa $[0, 1]$, condição necessária para a aplicação das integrais de Sugeno (SUGENO, 1974) e Choquet (CHOQUET, 1953), que são funções não-lineares Fuzzy responsáveis por realizar a fusão da decisão dos classificadores (um classificador para cada resolução). A integral de Sugeno de uma função h com relação à medida fuzzy g sobre um conjunto de fontes X (distâncias normalizadas, nesse caso) é definida por

$$S_g(h(x_1), h(x_2), \dots, h(x_m)) \stackrel{\text{def}}{=} \bigvee_{i=1}^m (h(x_{(i)}) \wedge g(A_{(i)})).$$

Enquanto a integral de Choquet é definida por

$$C_g(h(x_1), h(x_2), \dots, h(x_m)) \stackrel{\text{def}}{=} \sum_{i=1}^m (h(x_{(i)}) - h(x_{(i-1)})) \cdot g(A_{(i)}),$$

onde $A_{(i)}$ corresponde ao conjunto das últimas $m - i + 1$ fontes em X que pode ser denotado por $A_{(i)} = \{x_{(i)}, x_{(i+1)}, \dots, x_{(m)}\}$ e a medida Fuzzy $g(A_{(i)})$ é obtida utilizando a seguinte condição

$$g(A \cup B) = g(A) + g(B) + \lambda g(A)g(B), \quad \lambda > -1, \quad A, B \in 2^X,$$

Sendo que λ é determinado pela equação abaixo, selecionando o único valor que satisfaz $\lambda > -1$ e $\lambda \neq 0$

$$\prod_{i=1}^m (1 + \lambda g(x_{(i)})) = \lambda + 1.$$

Para essas integrais, as melhores medidas são as de maior valor. Assim, um valor de integral para cada classe conhecida é obtido e são selecionados o maior $d2$ e o segundo maior valor $d1$ para calcular a medida de contraste¹. Se o valor dado pela medida de contraste for maior que o valor de um limiar, esse vetor de teste é classificado como pertencente à classe que gerou o maior valor de integral ($d2$). Caso contrário, é rejeitado. O valor do limiar é dado pela maximização da seguinte equação

¹ A definição da medida de contraste contempla as menores distâncias – que são as melhores – mas agora é adaptada para aplicar com os valores das integrais.

$$score = 1 - \sqrt{\frac{(1 - rec^2) + far^2}{2}}$$

São verificados dois esquemas: o esquema 1 é o apresentado acima, que utiliza o conjunto de treinamento para obter o mapeamento para o espaço LDA e o de teste para obter tanto as densidades de medida Fuzzy quanto para testar o método e o esquema 2 consiste em utilizar o conjunto de teste para obter o mapeamento para o espaço LDA, obter as densidades de medida Fuzzy e testar o método. Entretanto, na fase de testes o mapeamento LDA utilizado é obtido do conjunto de treinamento, ou seja, antes da etapa de teste existe uma nova etapa de treinamento do mapeamento LDA. É importante notar que em ambos teremos algum *bias* adicionado, pois “o treinamento ou fornecerá conhecimento adicional das faces conhecidas ou da distribuição das faces desconhecidas” (ARBUCKLE, LANGE *et al.*, 1995).

Para avaliação do resultado foi utilizado um banco de dados de imagens de face próprio com 11.600 imagens de face frontais em tons de cinza de tamanho 160 x 120 pixels, sendo 100 imagens para cada um dos 116 indivíduos e com pequenas variações de pose (cerca de 15°). O conjunto foi dividido igualmente entre amostras de treinamento e teste, porém somente as amostras dos primeiros 60 indivíduos foram utilizadas para o treinamento. Alguns exemplos de indivíduos e variações de um mesmo indivíduo desse banco de dados são apresentados na Figura 3.17. E os melhores resultados obtidos nesse banco de dados foram 98,4% e 97,9 % de taxa de acerto, utilizando os esquemas 1 e 2, respectivamente e Integral de Sugeno. Essa técnica mostrou-se robusta a deslocamentos.

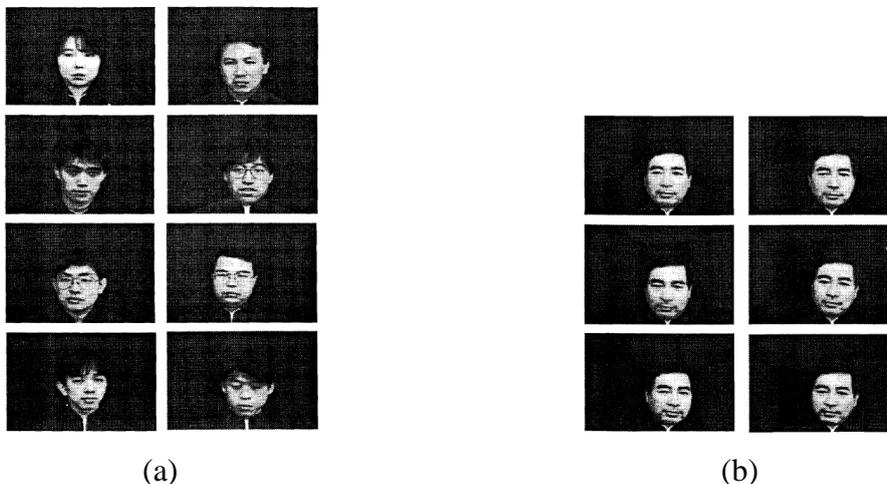


Figura 3.17. Exemplos de imagens no banco de dados: a) indivíduos diferentes; b) mesmo indivíduo. Extraído de (ARBUCKLE, LANGE *et al.*, 1995).

Em (ACHERMANN e BUNKE, 1996a) e (ACHERMANN e BUNKE, 1996b), diversas combinações de *eigenfaces* e Modelos Ocultos de Markov (HMM) são relatadas. Para permitir a aplicação do HMM, uma etapa de pré-processamento consistindo de equalização de histograma e normalização da largura é necessária. Além disso, como o HMM normalmente trabalha com sinais unidimensionais, a extração de atributos é feita por uma janela deslizante (sentido vertical) que cobre toda a largura da imagem e contém alguma sobreposição (para manter alguma informação contextual e evitar o corte de atributos de face significativos), sendo que o vetor de atributos corresponde aos valores de intensidade na janela deslizante. A idéia por trás disso é aproveitar a propriedade de que regiões faciais (testa, olhos, boca) são “identificáveis por um humano mesmo quando a imagem da face é dividida em janelas deslizantes como descrito acima”. Dessa forma, uma face humana pode ser representada por um modelo linear consistindo de cinco estados correspondendo a componentes faciais (1 = testa, 2 = olhos, 3 = nariz, 4 = boca e 5 = queixo) como mostrados na Figura 3.18. Finalmente, para a combinação foram analisados os seguintes combinadores: Votação por Maioria, Votação Consensual, Votação por Maioria Ponderada, Regra da Soma para *Score* e *Rank* e Regra do Produto. Antes de combinar os *scores* obtidos dos dois classificadores (distância Euclidiana e probabilidades para *eigenfaces* e HMM, respectivamente), eles precisam ser normalizados para um mesmo domínio. Para isso, foram analisadas quatro transformações: linear, logarítmica, exponencial e logística. Para a avaliação dos resultados um banco de dados próprio contendo 10 imagens em tons de cinza de cada um dos 30 indivíduos com variações de pose e dimensão de 512 x 342 pixels. Como exemplo, as amostras de um indivíduo do banco de dados são mostradas na Figura 3.19. O conjunto foi dividido ao meio em conjunto de teste e treinamento, ou seja, 5 imagens de cada indivíduo para treinamento e outras 5 para teste. O melhor resultado relatado foi 98,7% (considerando o primeiro *rank*) obtido pela utilização da Regra da Soma do *Score* e transformação logarítmica.

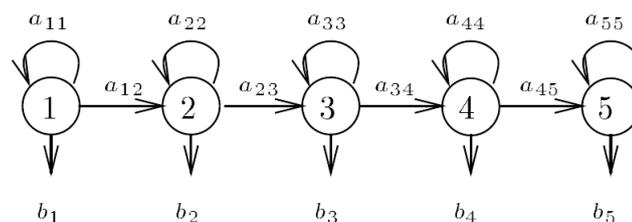


Figura 3.18. Modelo Esquerda-Direita de Face Humana Linear, onde a_{ij} representa a probabilidade de transição do estado i ao j e b_i representa a probabilidade de emissão de saída num estado i . Extraído de (ACHERMANN e BUNKE, 1996a).



Figura 3.19. Exemplos de um indivíduo no banco de dados. Adaptado de (ACHERMANN e BUNKE, 1996a).

Em (TOLBA e ABU-REZQ, 2000), um combinador baseado em regras é desenvolvido para combinar duas redes neurais: um classificador de Quantização Vetorial por Aprendizado (LVQ) e um classificador de Função de Base Radial (RBF). O conjunto de treinamento é dividido entre os classificadores neurais após treinar o LVQ (foi escolhido por apresentar maior taxa de classificação correta se comparado ao RBF) com todo o conjunto de treinamento e testá-lo no conjunto de teste, a fim de determinar quais classes são influenciadas por outras e quais não. Dessa forma, o conjunto de treinamento pode ser separado no grupo classificado corretamente (DC) e no grupo classificado erroneamente (FC), juntando-se a esse último as classes familiares e que interferiram na classificação. Ao grupo FC são incluídas um conjunto de amostras sintetizadas aplicando o LVQ (com 1 neurônio na camada escondida e 1 neurônio na camada de saída) entre as amostras das classes familiares a fim de aumentar a robustez do sistema. Alguns exemplos de faces sintetizadas pela técnica acima são mostradas na Figura 3.20. O grupo FC final é utilizado para treinar uma rede RBF e o DC é utilizado para treinar uma nova rede LVQ. Para combinar a saída dessas duas redes, são utilizadas as seguintes regras: 1) Rejeitar o padrão quando o LVQ rejeitá-lo; 2) Associar o padrão à classe determinada pelo LVQ quando a saída do RBF associá-lo à mesma classe ou o RBF rejeitá-lo; 3) Associar o padrão à classe determinada pelo RBF quando o LVQ associá-lo a uma classe diferente. Utilizando o banco de dados ORL dividido igualmente em conjunto de treinamento e teste, com as faces interpoladas utilizando o interpolador de vizinho mais próximo para um tamanho de 32 x 32 pixels aplicando-se antes um filtro passa-baixas (a idéia é “reduzir o efeito dos padrões de Moiré e os padrões *ripple* que resultam do *aliasing* durante a reamostragem”), e arquiteturas 1024-1200-40 e 1024-200-40, respectivamente para LVQ (com taxa de aprendizado de 0,5 e 8000 épocas de treinamento) e RBF (com constante *spread* igual 5) obteve uma taxa de classificação correta de 99,5% e 0,5% de rejeição. Foi relatado também, que utilizando de 6 a 8 amostras por classe no treinamento a taxa de classificação

correta foi de 100%, entretanto, o classificador individual LVQ já atingia essa taxa. Essa técnica apresentou-se bastante robusta a variações na pose e em expressões faciais.

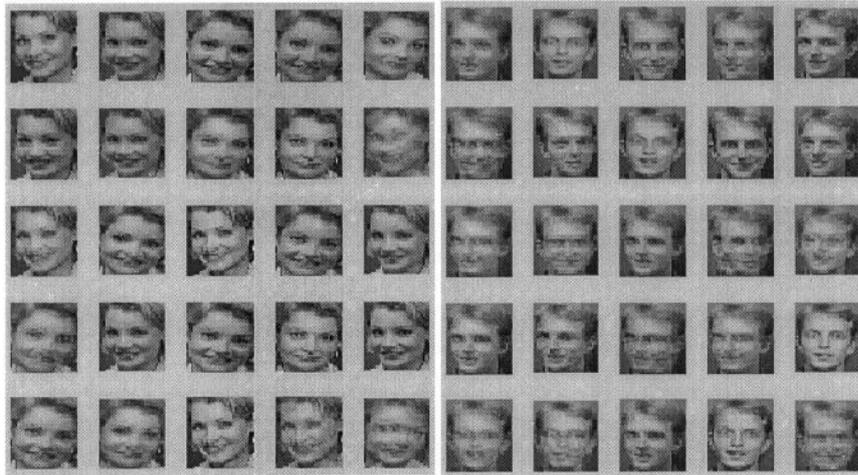


Figura 3.20. Exemplo de faces geradas artificialmente usando uma rede LVQ. Retirado de (TOLBA e ABUREZQ, 2000).

Marcialis e Roli propuseram combinar as técnicas estatísticas bem-conhecidas *eigenfaces* e *fisherfaces* (MARCIALIS e ROLI, 2002; MARCIALIS e ROLI, 2003). Como apresentado na Figura 3.21, as faces do banco de dados são projetadas nos espaços LDA e PCA. Para a combinação foram propostas duas abordagens KNN e Média Mais Próxima (NM). A geração dos vetores de distância d^{PCA} e d^{LDA} depende da regra de combinação a ser utilizada. Para o KNN, as distâncias são calculadas para cada amostra no conjunto de treinamento em seu respectivo espaço, ou seja, N (como mostrado na Figura 3.21) é igual ao número de amostras no conjunto de treinamento. Caso a combinação utilizada seja o NM, a distância é calculada com relação aos vetores média de cada classe no seu respectivo espaço, ou seja, N é igual a C , onde C é o número classes. Antes de realizar a combinação, todas as distâncias são normalizadas para $[0, 1]$. A combinação é dividida em duas etapas: 1) geração de um vetor de distância combinado d que deverá conter as informações de ambos os espaços e 2) aplicação do KNN ou NM. Para gerar o vetor de distância d são utilizados um dos dois modos seguintes:

a) Média:
$$d = \left\{ \frac{d_1^{PCA} + d_1^{LDA}}{2}, \dots, \frac{d_N^{PCA} + d_N^{LDA}}{2} \right\};$$

b) União (Append):
$$d = \{d_1^{PCA}, \dots, d_N^{PCA}, d_1^{LDA}, \dots, d_N^{LDA}\}.$$

Dessa forma, temos 4 combinações diferentes denominadas M-KNN, A-KNN, M-NM, A-NM. Essas combinações foram testadas com as bases de faces ORL e Yale. Para o ORL, utilizando metade do conjunto para treinamento e metade para teste foi obtida uma taxa de reconhecimento médio de 97,25% usando o esquema M-KNN com 119 componentes principais e 39 para o LDA. Já para a base Yale, foi relatada uma taxa de reconhecimento médio de 84,22% usando o esquema A-KNN com 33 componentes principais e 14 para o LDA, sendo que para treinamento e teste foram utilizados 5 e 6 imagens por classe, respectivamente. As taxas de reconhecimento relatadas correspondem à média de 10 repetições do experimento, cada uma com uma partição aleatória do conjunto de teste.

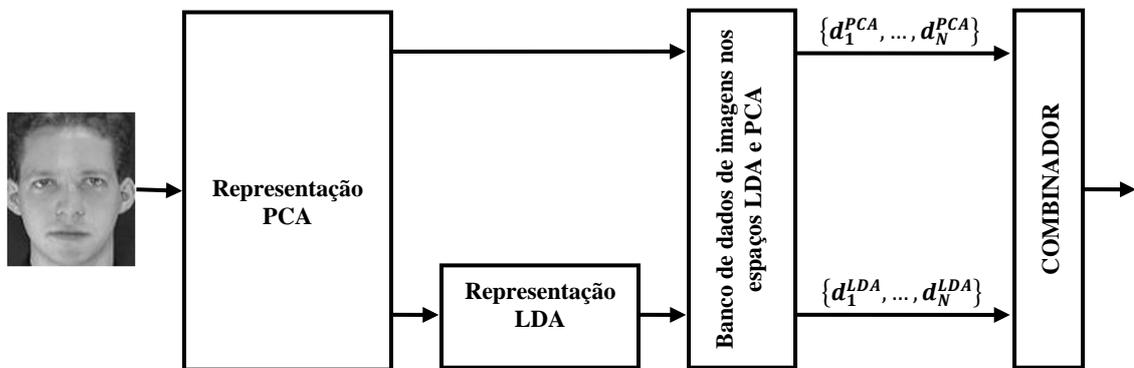


Figura 3.21. Metodologia de fusão proposta por Marcialis e Roli. Adaptado de (MARCIALIS e ROLI, 2002).

Em (MARCIALIS e ROLI, 2004), o algoritmo acima foi também aplicado à base de dados do MIT, obtendo uma taxa de reconhecimento médio de aproximadamente 79% com o esquema A-KNN. O conjunto de treinamento era constituído das imagens mais próximas e mais distantes da câmera e o de teste pelas intermediárias.

Ryu e Oh (RYU e OH, 2002) propuseram uma abordagem em regras para combinar um classificador de Ângulo de Atributos Mais Próximo (NFA) e um *ensemble* de redes neurais MLP treinadas pelo algoritmo *Resilient Backpropagation* (ERN), com ambos os classificadores aplicados no espaço PCA. A medida de similaridade utilizada em todo o sistema é uma medida de ângulo de atributos e é definida como

$$S(x_i, x_j, x_r) = \cos \theta = \frac{\bar{x}_i \cdot \bar{x}_j}{\|\bar{x}_i\| \|\bar{x}_j\|},$$

onde $\bar{x}_i = x_i - x_r$ e $\bar{x}_j = x_j - x_r$. A classificação por NFA é dada pela classe da amostra que maximiza a medida de similaridade acima e pode ser denotada como

$$S(x, x_{i^*}^{c_{NFA}}, x_r) = \max_{i \leq c \leq C} \max_{1 \leq i \leq N_c + K_c} S(x, x_i^c, x_r),$$

onde x é a face de teste, C é o número de classes, N_c é o número de amostras da classe c , K_c é o número de amostras virtuais da classe c , x_r é o vetor nulo, x_i^c são os protótipos i da classe c e $x_{i^*}^{c_{NFA}}$ é o protótipo mais similar a x . Essas amostras virtuais são geradas com a finalidade de contemplar mais variações da face e, em consequência, ampliar a robustez contra variações e aumentar o desempenho da classificação. Para gerar essas amostras virtuais, considere a Figura 3.22. São considerados dois tipos de regiões A e B, esparsa e densa, respectivamente. Para uma região esparsa A, onde $S(x_1, x_2, x_r) \leq \alpha$, uma amostra virtual v_1 pode ser gerada como

$$v_1 = x_1 + \mu(x_2 - x_1),$$

e para uma região densa B, onde $S(x_3, x_4, x_r) > \alpha$, uma amostra virtual v_2 pode ser gerada seguindo

$$v_2 = x_4 + \mu(x_M - x_4)$$

$$x_M = x_2 + \mu(x_3 - x_2),$$

onde x_1, x_2, x_3 e x_4 são amostras do conjunto de treinamento de uma determinada classe, μ é igual a 0,5. Uma outra forma de gerar amostras virtuais é simplesmente espelhando a amostra da classe.

O classificador ERN é composto por L redes MLP (FRN) (como pode ser visto na Figura 3.23). Cada rede é treinada individualmente, sendo que a face mais frontal é selecionada como a face desejada (ou alvo), ou seja,

$$\tilde{x}^c \cong \hat{x}_{i,l} = f_l(x_i^c),$$

onde $l = 1, 2, \dots, L$, L é o número de FRNs e $i = 1, 2, \dots, N_c + K_c$ e $f_l(\cdot)$ é a função de mapeamento da l -ésima FRN. As respostas das FRN podem ser fundidas calculando-se a média, isto é,

$$\tilde{x}^c \cong \hat{x}_i = \frac{1}{L} \sum_{l=1}^L \hat{x}_{i,l}.$$

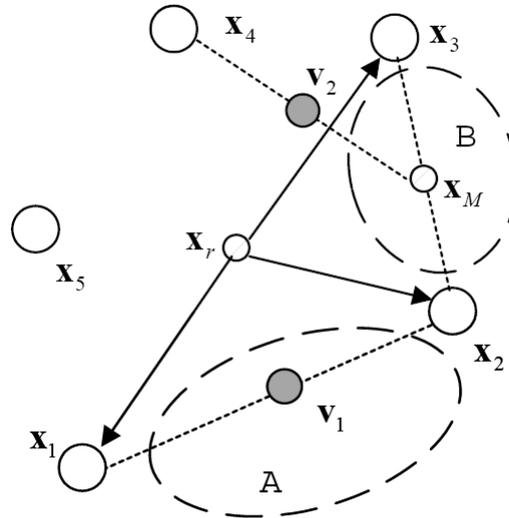


Figura 3.22. Exemplo de distribuição do conjunto de treinamento e geração de amostras virtuais. Retirado de (RYU e OH, 2002).

E finalmente, a classificação realizada pela ERN é dada por

$$S(\hat{x}, \tilde{x}^{c_{ERN}^*}, \tilde{x}_r) = \max_{1 \leq c \leq C} S(\hat{x}, \tilde{x}^{c_{ERN}}, \tilde{x}_r),$$

onde \hat{x} é a amostra de teste aplicada à ERN, \tilde{x}_r é a média das faces frontais frontais \tilde{x}^c , $\tilde{x}^{c_{ERN}^*}$ é a amostra frontal \tilde{x}^c da classe que maximiza a equação acima e fornecerá, portanto, o seu rótulo a \hat{x} . Enfim, para combinar a saída das duas fontes utiliza-se a seguinte regra de combinação

$$c_{final}^* = \begin{cases} c_{ERN}^*, & \text{se } c_{NFA}^* = c_{ERN}^* \\ c_{ERN}^*, & \text{se } S(\hat{x}, \tilde{x}^{c_{ERN}^*}, \tilde{x}_r) \geq S(x, x_i^{c_{NFA}^*}, x_r), & c_{NFA}^* \neq c_{ERN}^* \\ c_{NFA}^*, & \text{se } S(\hat{x}, \tilde{x}^{c_{ERN}^*}, \tilde{x}_r) < S(x, x_i^{c_{NFA}^*}, x_r), & c_{NFA}^* \neq c_{ERN}^* \end{cases}$$

Essa técnica foi aplicada ao banco de dados ORL, sendo dividido ao meio entre amostras de teste e treinamento (sendo que a amostra frontal e uma semi-frontal devem ser garantidas no conjunto de treinamento). Dessa forma, foi atingida uma taxa de reconhecimento de 96,33%, com $\alpha = 0,5$, sendo 5 FRNs ($L = 5$) com 2 camadas escondidas cada e configurações nas camadas escondidas 150-150, 80-80, 50-200, 150-200 e 80-100,

respectivamente. Além disso, a taxa de aprendizado das FRNs foi 0,2, 50 componentes principais foram utilizados e foi tentado restringir o tamanho máximo de cada classe a 40 amostras por classe (ou seja, foram geradas no máximo 30 amostras virtuais por classe).

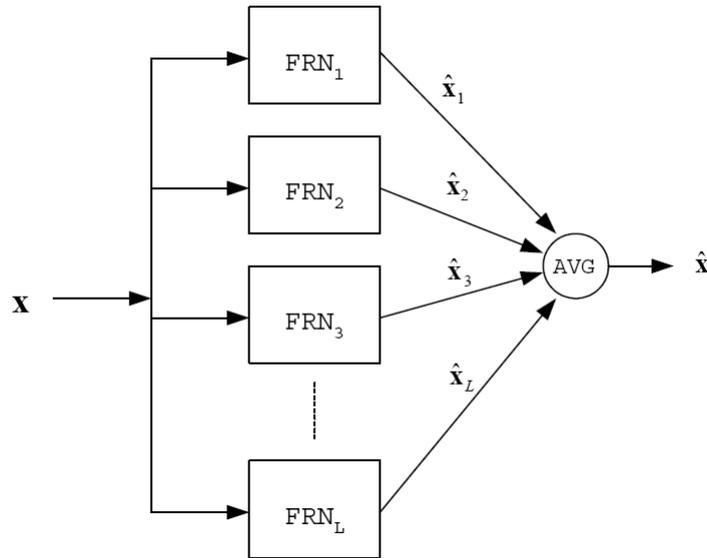


Figura 3.23. Arquitetura da ERN. Retirado de (RYU e OH, 2002).

Um novo esquema de combinação baseado em algoritmo genético foi apresentado em (JING e ZHANG, 2003). A fim de aumentar o desempenho da combinação são utilizadas diferentes fontes de informação. Para isso, às imagens é aplicada a Transformada Wavelet Ortogonal de Daubechies, o que resulta em 4 sub-imagens diferentes descorrelacionadas. Em seguida, são extraídas *Fisherfaces* separadamente considerando cada subimagem e, finalmente, um classificador NN é utilizado. O combinador proposto para fundir a saída desses 4 classificadores é denominado Critério Máximo de Complementaridade (MCC). Esse combinador utiliza uma medida de correlação $R_{a,b}$ entre dois classificadores a e b como

$$R_{a,b} = \sum_k^c w_k r_k$$

$$w_k = \frac{e^{-k}}{\sum_{s=1}^c e^{-s}}$$

$$r_k = 1 - \left| \frac{a_k}{\sum_{s=1}^c a_s} - \frac{b_k}{\sum_{s=1}^c b_s} \right| / \left(\frac{a_k}{\sum_{s=1}^c a_s} + \frac{b_k}{\sum_{s=1}^c b_s} \right),$$

onde a_k, a_s denotam o k -ésimo e o s -ésimo componente de a , b_k, b_s são o k -ésimo e o s -ésimo componente de b , c é o número de classes, sendo que a e b são ordenados em ordem decrescente de similaridade, ou seja, as mínimas distâncias primeiro. Para o cálculo da correlação, caso os primeiros componentes de a e b fossem de classes diferentes, esses componentes não eram utilizados no seu cálculo, pois é dada mais importância a eles (veja o cálculo do peso w_k) e o seu valor de r_k seria 0. O critério de combinação final J (MCC) será utilizado em um algoritmo genético para determinar os pesos para os classificadores e é calculado pelas seguintes equações

$$J = \max \left(\sum_{i=1}^n E_i \right) = \max \left(\sum_{i=1}^n (G_i \cdot U_i^\beta) \right)$$

$$U = \|R_0 - R\| / \|R_0\|$$

$$R = [R_{1,2}, R_{1,3}, \dots, R_{1,d}, R_{2,3}, R_{2,4}, \dots, R_{2,d}, \dots, R_{d-2,d-1}, R_{d-2,d}, R_{d-1,d}]$$

$$G = (x_1 - x_2) / (x_1 - \bar{x}),$$

onde n é o número de amostras de treinamento, i é a i -ésima amostra de treinamento, E_i é a complementaridade entre os classificadores, G_i é o grau de correção do resultado combinado, U_i é o grau total de descorrelacionamento dos classificadores, d é o número de classificadores, x_1, x_2, \bar{x} são os 2 componentes mais prováveis e a média de um resultado combinado correto X (vide referência para maiores detalhes) e β foi definido experimentalmente como 0,5. Um dos bancos de dados de imagens faciais NUST603 foi utilizado. Ele contém 18 indivíduos com 12 imagens faciais de 64x64 pixels por indivíduo (algumas amostras são exibidas na Figura 3.24). Foram realizadas seleções aleatórias de 3 amostras por classe para treinar os classificadores, outras 3 por classe para calcular o MMC e os pesos para os classificadores usando algoritmos genéticos e o restante para calcular a taxa de reconhecimento. Executando cada experimento 10 vezes, foi obtida uma taxa de reconhecimento de 96,4%. Foi relatada também a comparação com os combinadores de Votação por Maioria e a Regra de Taxa de Erro Mínimo, porém estas com taxas de reconhecimento menores.

Lu e seus colaboradores (LU, WANG *et al.*, 2003) utilizaram a Regra da Soma e a uma rede neural RBF para combinar 3 classificadores de face baseados nas técnicas de extração de atributos PCA, ICA e LDA. Para comparar duas faces, a medida de similaridade

utilizada é a distância do cosseno. Foi gerado um banco de dados próprio coletando imagens de face de outros 4 bancos de dados públicos: ORL, Yale, AR e NLPR+MSU. Essa base é composta por 206 indivíduos com 10 amostras cada e as imagens tem tamanho 42 x 42 pixels. Alguns exemplos de imagens faciais são apresentados na Figura 3.25. Foram utilizadas 9 amostras para treinamento e 1 para teste, sendo executados 10 experimentos trocando-se a partição de forma que toda amostra tenha sido utilizada para teste. Assim, a melhor taxa de reconhecimento média relatada foi 90,2% utilizando a RBF como combinador.



Figura 3.24. Algumas amostras de um indivíduo do banco de dados NUST603. Retirado de (JING e ZHANG, 2003).



Figura 3.25. Exemplos de imagens faciais na base de dados: a) ORL, b) Yale, c) AR e d) NLPR+MSU. Retirado de (LU, WANG *et al.*, 2003).

Kim e Kittler investigaram algumas técnicas a fim de fornecer um reconhecimento de face invariante à pose (KIM e KITTLER, 2004; 2006). Segundo eles, os “métodos de reconhecimento de face invariante à visão existentes podem ser geralmente decompostos em três passos sequenciais: representação, transformação de visão e função de discriminação”. A fim de satisfazer esses passos, foram propostos 4 classificadores base. O primeiro deles utilizava PCA para representação, uma técnica de Matriz Linear (LM) para aprender uma função de transferência entre faces rotacionadas e face frontal e que minimiza o MSE e, finalmente, o LDA é escolhido como função discriminante. Esse método é chamado PCA-LM-LDA. O segundo, denominado Análise de Discriminante Linear Localmente (LLDA), foi

proposto de forma a suprir o problema de suboptimalidade do primeiro, causado pela realização dos 3 passos individualmente. O LLDA consiste de um *framework* unificado para os 3 estágios. Ele “encontra concorrentemente o conjunto de transformações lineares localmente a fim de obter classes de face transformadas linearmente e localmente que maximizam a covariância entre classes enquanto minimiza a covariância dentro da classe”, como é exibido na Figura 3.26. O terceiro classificador proposto difere do primeiro por aplicar uma função de kernel que transforma o espaço de entrada num de mais alta dimensão de tal modo que nessa nova dimensão os dados possam ser separados linearmente. Essa técnica é denominada Análise de Discriminante Generalizado (GDA) e um kernel RBF foi utilizado como função de transformação. E por último, uma técnica baseada em *Look-Up Table* (LUT) é proposta para gerar imagens virtuais rotacionadas a partir da imagem original como mostrado na Figura 3.27. LDA é utilizado como função discriminante. Todos esses classificadores base possuem ao final vetores de atributos no espaço Euclidiano e devem ser normalizados para o intervalo $[0, 1]$. Esses classificadores foram combinados utilizando 6 combinadores: Regra do Produto, Regra da Soma, Regra da Soma Ponderada, Regra do Máximo, Regra do Mínimo e Regra da Mediana. Aplicando-os sobre a base de dados XM2VTS, foi obtida uma taxa média de reconhecimento correto (considerando as diferentes visões) de 73,2% para faces da mesma sessão e 54,4% (KIM e KITTLER, 2004) para sessões diferentes com as Regras da Soma e Soma Ponderada. As imagens foram redimensionadas para 46 x 56 pixels e a base de dados foi dividida em 125 pessoas, 45 pessoas e 125 pessoas para conjuntos de treinamento, avaliação e teste respectivamente. Para o conjunto de teste a face frontal F-S1 foi utilizada como galeria e as restantes como consulta. Em (KIM e KITTLER, 2006) foram relatados 70% e 42% para faces da mesma sessão e sessões diferentes sobre classes de face rotacionadas por mais de 30° da imagem frontal, respectivamente.

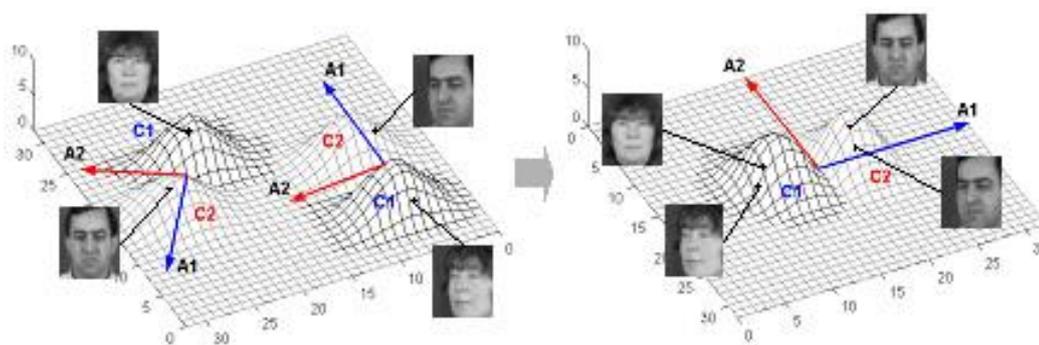


Figura 3.26. Classificação de face usando o LLDA. A distribuição original dos dados e a distribuição transformada são exibidas do lado esquerdo e direito, respectivamente. Retirado de (KIM e KITTLER, 2004).

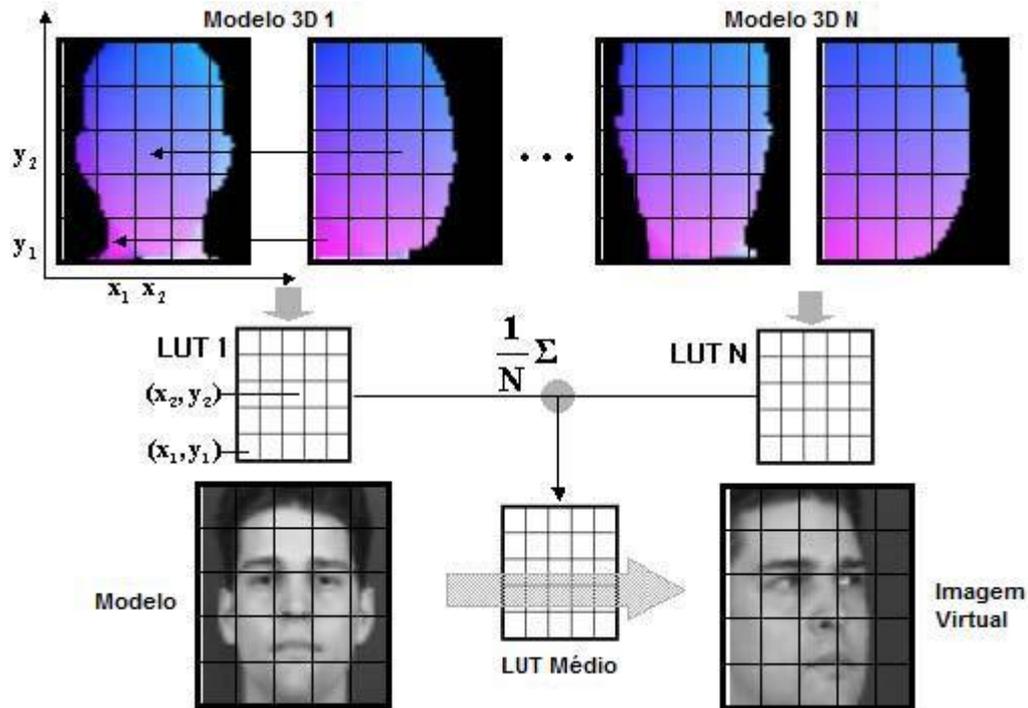


Figura 3.27. Geração de Visão Virtual Rotacionada usando LUT 3D. Adaptado de (KIM e KITTLER, 2004).

Em (ZHANG, SHAN *et al.*, 2004) são propostos dois esquemas de fusão de nível de atributos e uma fusão em nível de decisão. Para todas as fusões são utilizadas imagens de intensidade recortadas para 64 x 64 pixels com histogramas normalizados e máscara para remoção de cabelo aplicado, e 40 características de Gabor (5 escalas e 8 orientações) dessas imagens de intensidade são extraídas. A Figura 3.28 apresenta uma amostra do banco de dados JDL e o pré-processamento executado. O primeiro esquema de fusão de nível de atributos consiste em aplicar PCA às características de Gabor e à imagem de intensidade, unindo os seus atributos, aplicando o LDA e depois um algoritmo de *matching*. O outro esquema de fusão em nível de atributos consiste em aplicar LDA sobre componentes principais de ambas as características, unindo os componentes de Fisher e aplicado um algoritmo de *matching* para fazer a decisão final. Por fim, o esquema de fusão em nível de decisão utiliza as Regras da Soma e do Produto para combinar as saídas do módulo de *matching* no espaço LDA de componentes principais. Testado sobre o banco de dados FERET, o combinador em nível de decisão com a Regra do Produto como combinador obteve o melhor desempenho com 98,2% de taxa de reconhecimento correto para um EER (*Equal Error Rate*) de 0,8% com o conjunto de treinamento consistindo de 1.002 imagens que são

subconjuntos de F_a e F_b , a galeria contendo uma imagem por classe e o conjunto de prova sendo o F_b . Utilizando o banco de dados próprio denominado JDL, o melhor resultado foi 88,5% de taxa de reconhecimento correto e 3,95% de EER pelo esquema de fusão em nível de atributos PCA para 14 amostras de cada um de 300 indivíduos no conjunto de treinamento e 6577 imagens dos 500 indivíduos no conjunto de teste.

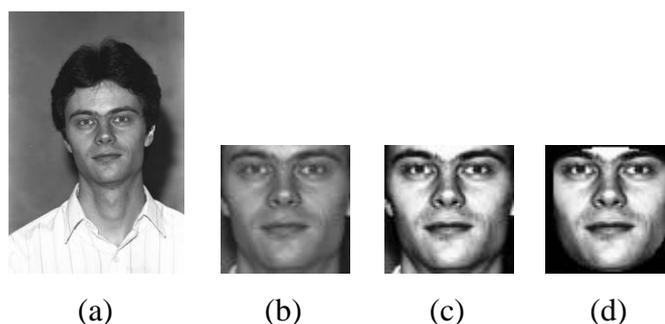


Figura 3.28. Uma amostra do banco de dados JDL e o pré-processamento executado: a) Imagem Original; b) Imagem Recortada; c) Histograma Equalizado e d) Máscara para Remoção de Cabelo aplicada. Retirado de (ZHANG, SHAN *et al.*, 2004).

Uma outra técnica utilizando máquinas de comitê com estratégia de Votação por Maioria de redes neurais RBF treinada por um algoritmo de Aprendizado Híbrido (HLA) que combina os paradigmas do Gradiente e Mínimos Quadrados Lineares (LLS) foi proposta em (ZHAO, HUANG *et al.*, 2004a; b). O comitê é composto por N redes neurais, sendo N dependente do número de blocos e domínios utilizados no sistema. As técnicas de extração de atributos que definem o domínio são aplicadas sobre blocos da imagem original. As técnicas de extração de atributos utilizadas são Operador de Interesse, PCA, LDA e Kernel LDA e as imagens originais foram divididas em $1 * 1$, $2 * 2$, $4 * 2$, $4 * 3$ ou $4 * 4$ blocos. Essa técnica foi testada com o banco de dados ORL e o melhor resultado relatado foi 97,7% de taxa média de reconhecimento correto para 10 execuções quando utilizados $4 * 4$ blocos para Operador de Interesse e PCA e $1 * 1$ para Kernel LDA com o banco de dados dividido igualmente entre conjunto de treinamento e teste. Foram extraídos 2185, 180, 30 e 40 componentes para Operador de Interesse, PCA, LDA e Kernel PCA, respectivamente. O número de blocos foi diferente para o Kernel LDA, pois foi identificado que divisões em blocos para Kernel LDA e LDA foram ineficientes nos experimentos. Portanto, havia $4 * 4 * 2 + 1 * 1 * 1 = 33$ RBFs.

Uma abordagem utilizando *Ensemble* de classificadores NN com a Distância do Cosseno no Espaço de Mahalanobis como medida de similaridade foi proposto em (CHAWLA e BOWYER, 2005). Para construir os *Ensemble-1*, *Ensemble-2* e *Ensemble-3* foram realizadas uma seleção aleatória de 1, 2 ou 3 imagens, respectivamente, de um total de 4 amostras por indivíduo para construir 50 espaços de face diferentes para cada *Ensemble*. Nesses espaços as imagens foram normalizadas e recortadas para 130x150 pixels para em seguida aplicar o PCA extraíndo 120, 241 e 362 vetores de base para o *Ensemble-1*, *Ensemble-2* e *Ensemble-3*, respectivamente. Finalmente, a fusão dos classificadores é realizada atribuindo ao padrão a classe associada à menor distância entre as distâncias selecionadas pelos 50 classificadores. Experimentando com o banco de dados Notre Dame (ND), o Ensemble-3 atingiu os melhores resultados com taxa de reconhecimento correto médio para 10 execuções aleatórias de 70,7%, 74,3%, 75,6% e 76,9% para os conjuntos de prova FA-LF, FA-LM, FB-LF e FB-LM, respectivamente. Onde FA e FB indicam expressões neutra e sorrindo, respectivamente e LF indicam que a fonte de iluminação é lateral (ambos os lados) e LM que é, além de lateral, também central. Algumas amostras desse banco de dados são apresentadas na Figura 3.29. Vale ressaltar que as imagens nos conjuntos de treinamento, teste e galeria não eram sobrepostas e 121 dos 484 sujeitos foram selecionados aleatoriamente 10 vezes.

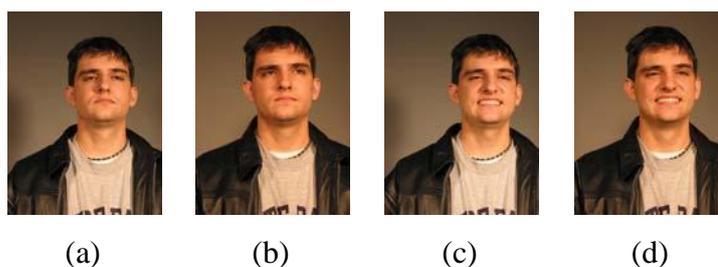


Figura 3.29. Algumas amostras do banco de dados da Universidade de Notre Dame (ND): a) FA-LF; b) FA-LM; c) FB-LF e d) FB-LM. Onde FA e FB indicam expressões neutra e sorrindo, respectivamente e LF indica que a fonte de iluminação é lateral (ambos os lados) e LM que é, além de lateral, também central. Retirado de (CHAWLA e BOWYER, 2005).

Uma nova abordagem dividir para conquistar foi proposta em (EBRAHIMPOUR, EHTEHAM *et al.*, 2005) que combina redes neurais MLP usando a Regra do Máximo. As redes neurais MLP são utilizadas para classificação binária e há uma MLP para cada classe. A arquitetura dessa rede é $M - K - 1$ com funções de ativação Tangente Hiperbólica Não-

Linear para os nós da camada escondida e uma Linear para os nós de saída, onde M é o número de vetores de base PCA utilizados, K é o número de nós na camada escondida e 1 é o número de nós de saída que representa a probabilidade de um dado padrão observado pertencer à classe i . O PCA é calculado sobre a face pura que consiste na região da face compreendida por uma elipse que melhor se ajusta à área da face que contém a informação pertinente ao reconhecimento de faces e que tenta excluir ao máximo regiões como o cabelo, pescoço e fundo. Essa abordagem foi testada sobre as bases de dados ORL e Yale e apresentou, para ambos, uma taxa de reconhecimento correto médio de 100% para 5 repetições do experimento. Para o ORL, as 40 MLPs apresentavam uma arquitetura 50-15-1 e foram treinadas por 70 épocas, sendo que a base foi dividida igualmente entre conjuntos de treinamento e teste. Já para o Yale, as 15 MLPs apresentavam uma arquitetura 50-7-1 e foram treinadas 55 épocas, sendo que 90 imagens foram usadas para treinamento e as restantes para teste.

Uma técnica extraindo atributos PCA de imagens direcionais geradas por Banco de Filtros Direcionais (DFB) de estágio 3 foi proposta por (KHAN, IBRAHIM *et al.*, 2005). Para combinar as saídas dos 8 especialistas DFB-PCA são comparadas a utilização de uma Medida de Correlação Cruzada e Votação por Maioria. Aplicando sobre a base de dados ORL, a combinação por medida de Correlação obteve um desempenho superior com taxa de reconhecimento correto de 97%, sendo que houve uma etapa de pré-processamento onde todas as imagens foram recortadas para conter somente as regiões faciais e normalizadas para média zero e variância unitária. E ainda, o conjunto de treinamento foi composto por 9 imagens por pessoa e as imagens restantes formaram o conjunto de teste.

A Regra da Média foi proposta para combinar dois esquemas de classificadores em (LUMINI e NANNI, 2005). O primeiro esquema proposto usava 6 classificadores Parzen Window com domínios de atributos PCA, DCT, Filtro de Gabor e LDA aplicado sobre esses 3 espaços anteriores. O segundo esquema proposto era composto por classificadores *Parzen Window* e método *Subspace* sobre o domínio DCT e DCT-LDA e Vizinho Mais Próximo sobre DCT-LDA, totalizando 5 classificadores. O primeiro esquema obteve o melhor resultado para o banco de dados YALE-B com taxa de acerto médio de 99,7% e ERR de 0,1%, enquanto o segundo esquema obteve o melhor resultado para o ORL com taxa de acerto médio de 98% e ERR de 2,2%. Essas taxas são referentes a 5 repetições dos experimentos. Ainda, foram considerados 50 componentes para PCA, DCT. Já para o Filtro de Gabor foi utilizada Wavelet de Gabor em somente uma escala e oito direções e aplicado PCA em

seguida para obter um espaço com menos de 50 dimensões. Para o LDA foram considerados 39 e 9 componentes para ORL e Yale-B, respectivamente. Vale ressaltar que o ORL foi dividido igualmente entre conjunto de treinamento e teste e para o YALE-B foram consideradas somente as imagens com faces frontais, totalizando 108 amostras por classe. E finalmente, as imagens do ORL passaram por um pré-processamento que consistia de aumentar a dimensão da imagem suavizando as linhas e colunas extremas até o tamanho final desejado para a imagem (142x122 pixels) como mostrado na Figura 3.30.

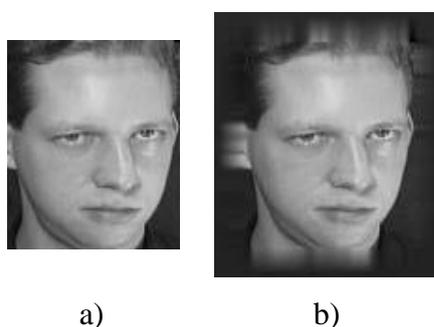


Figura 3.30. Imagem do banco de dados ORL a) antes e b) depois da etapa de pré-processamento. Retirado de (LUMINI e NANNI, 2005).

Em (GAO e WANG, 2006) a Regra da Soma foi utilizada para combinar classificadores gerados por *Boosting* em Subespaços Aleatórios. O algoritmo AdaBoost é o responsável por gerar os classificadores em subespaços de atributos de Gabor (gerados por Filtros de Gabor com 5 escalas e 8 orientações) e Padrão Binário Local (LBP) que utiliza distâncias Chi quadrado entre os histogramas LBP de subregiões como característica discriminativa para classificação binária (intrapessoal e extrapessoal). Essas técnicas foram testadas sobre o banco de dados FERET e o melhor resultado foi obtido com múltiplos subespaços aleatórios em atributos LBP com 99% de taxa de reconhecimento médio para 10 execuções (excluindo a pior e a melhor precisão). O conjunto de treinamento era do conjunto padrão FERET e incluía 736 imagens de 314 indivíduos, o que gerou 592 e 269.888 pares intrapessoal e extrapessoal, respectivamente. Entretanto, 4.000 faces extrapessoais foram selecionadas aleatoriamente para cada repetição do experimento a fim de manter um balanço entre amostras positivas e negativas. O conjunto de teste era composto pelo Fb e havia 1.196 imagens (1 por pessoa) na galeria. Todas as imagens foram normalizadas para 60x50 e

142x120 pixels para Gabor e LBP, respectivamente. Os subespaços originais de Gabor e LBP continuaram 120.000 e 11.700 dimensões, respectivamente.

Uma técnica combinando 4 variantes lineares e não lineares do LDA foi proposta por (HUANG, YUEN *et al.*, 2006). Essas variantes são Direct-LDA (DLDA), Subspace-LDA (SLDA), Kernel Direct-LDA (KDDA), Kernel Subspace-LDA (KSLDA), sendo as duas primeiras métodos de extração de atributos lineares e as outras não-lineares. O DLDA “descarta o espaço nulo da matriz de covariância entre classes aplicando antes o PCA e mantendo o espaço nulo da matriz de covariância intraclasses”. O SLDA “encontra o subespaço nulo da matriz de covariância intraclasses e descarta o espaço nulo da matriz de covariância entre classes”. KDDA e KSLDA são as versões não lineares de DLDA e SLDA, respectivamente. Descartar o espaço nulo da matriz de covariância entre classes é uma tentativa de resolver o problema de poucas amostras. O combinador utilizado para agregar esses domínios é a Regra da Soma Ponderada, sendo os pesos $\frac{r_i}{\gamma}$ determinados por

$$r_i = \left(\frac{\overline{BeDis}_i}{\overline{InDis}_i} \right)^2$$

$$\gamma = \sum_{i=1}^R r_i,$$

onde R é o número de classificadores, \overline{BeDis}_i e \overline{InDis}_i são as distâncias médias entre classes e intraclasses, respectivamente. Entretanto, antes de realizar a combinação, as medidas de distância dos diferentes domínios precisam ser normalizadas para um mesmo domínio. Isso é feito calculando a medida de similaridade S dada por

$$S_{ij} = \exp\left(-\frac{d_{ij}^2}{2\sigma^2}\right)$$

$$\sigma = \sqrt{\frac{\sum_{i,j} d_{ij}^2}{N}},$$

onde N é o número total de distâncias e d_{ij} é a distância entre os padrões Z_i e Z_j . Essa técnica foi aplicada nas bases de dados FERET, YaleB e CMU PIE e todas as imagens passaram por uma etapa de pré-processamento que consistiu em reamostrar para um tamanho de 92x112 pixels, aplicar a transformada Wavelet de Daubechies-4 e selecionar a banda LL com

resolução 30 x 25 pixels como as imagens de entrada final. Para uma média de 10 repetições foi relatada uma taxa de reconhecimento correto médio de 83,21%, 89,92% e 79,42% para o FERET, YaleB e CMU PIE, respectivamente. Foram selecionados 240 pessoas do FERET com 4 imagens por pessoa, sendo que foram escolhidas 120 pessoas e suas imagens foram divididas aleatoriamente com 2 amostras por indivíduo para treinamento e outras 2 para teste. Para o YaleB foram selecionadas aleatoriamente 72 e 333 amostras por indivíduo para treinamento e teste, respectivamente para um total de 10 indivíduos. Finalmente, para o CMU PIE foram selecionadas aleatoriamente 14 e 41 imagens por indivíduo para treinamento e teste, respectivamente para um total de 68 indivíduos. Todas as imagens nas três bases foram alinhadas antes de serem reamostradas, sendo que no FERET e YaleB pelo centro dos olhos e boca, no CMU PIE pelo centro dos olhos e centro do nariz, e ainda para o YaleB as imagens não frontais foram alinhadas pelo centro da face.

A transformação do espaço de atributos para um espaço de dissimilaridades foi proposta por (KIM, 2006) a fim de ultrapassar o problema de poucas amostras. A vantagem de utilizar medidas de dissimilaridade para a representação está em “não operar nas distribuições condicionais das classes, o que pode fazer com que a precisão exceda o limite de erro de Bayes”. Para gerar a matriz de dissimilaridades D que é definida como um espaço de dissimilaridades e contém as distâncias de uma imagem de treinamento para todas as outras, são utilizadas, além da distância Euclidiana, a Distância Regional (RD), que consiste na “média da diferença mínima entre o valor de cinza de um pixel e o valor de cinza de cada pixel na vizinhança 5x5 do pixel correspondente”. Sobre o espaço de dissimilares são definidos 2 classificadores: um que utiliza o KNN sobre D e o outro que considera o NN para a média de cada classe. Ainda um terceiro classificador é gerado sobre o espaço de atributos que consiste no KNN referente aos vetores médias de um conjunto de *clusters* gerados por K-médias. Votação por Maioria é usada para combinar esses 3 classificadores. Essa técnica foi testada sobre os bancos de dados Yale e ORL e validados por *leave-one-out*. Distância Euclidiana e RD apresentaram os melhores resultados para Yale e ORL, respectivamente, com taxa de reconhecimento correspondente a 93,03 e 99,25% com 9 *clusters* por classe. Para Yale, as imagens faciais foram recortadas manualmente para 236x178 pixels e a intensidade foi normalizada.

Um *ensemble* de LDA foi proposto em (KONG, LI *et al.*, 2006). Para gerar os classificadores fracos LDA, Boosting (realizado por AdaBoost) é aplicado sobre Subespaços Aleatórios projetados antes no espaço PCA. Como classificador foi utilizado o NN referente

ao centro de cada classe no seu respectivo subespaço. Para combiná-los é utilizada Votação por Maioria. Para validar o método, ele foi aplicado sobre as bases ORL e Yale B, tendo como conjunto de treinamento 2 a 9 e 2 a 12 amostras por classe, respectivamente e as restantes para teste. Na base de dados ORL, foi obtida uma taxa média de reconhecimento para 50 execuções de aproximadamente 88% e 100% para 2 e 9 amostras por classe no conjunto de treinamento, respectivamente e com imagens reduzidas pela metade em tamanho e 30 e 230 principais componentes correspondentemente. Finalmente para o Yale B, foi atingida uma taxa de reconhecimento médio para 100 execuções de cerca de 60% e 96% para 2 e 12 amostras por classe no conjunto de treinamento, respectivamente e com imagens redimensionadas para 50x60 pixels e 10 e 90 componentes principais extraídos. Porém, para os valores de principais componentes citados acima foram selecionados N_0 entre os maiores autovalores e N_1 selecionados aleatoriamente dos autovetores restantes.

Duas novas variantes de LDA, Fisherfaces pós-processados (pFisherfaces) e BDPCA+LDA são combinadas por Regra da Soma de distâncias que seleciona a menor soma de distâncias (ZUO, WANG *et al.*, 2007). A variante pFisherface consiste em aplicar um filtro Gaussiano para suavizar os Fisherfaces. BDPCA (PCA Bidimensional) é aplicado sobre o espaço original e posteriormente LDA para gerar BDPCA+LDA. Sobre esses domínios LDA, as menores distâncias para cada classe são entradas que serão normalizadas para permitir, posteriormente, combiná-las. Aplicando sobre o banco de dados ORL e FERET para 10 repetições foi atingida uma taxa de reconhecimento médio de 97,65% e 90,55%, respectivamente. O ORL estava dividido igualmente entre treinamento e teste, enquanto para o FERET foram utilizadas 600 imagens para treinamento e 800 para teste para um total de 200 indivíduos. Para o FERET todas as imagens foram recortadas em 80x80 pixels e tiveram o histograma equalizado.

Para verificar o efeito de transformações na imagem sobre reconhecimento de face biométrico, Jarillo e seus colegas propuseram em (JARILLO, PEDRYCZ *et al.*, 2008) um sistema com 3 classificadores NN aplicados sobre espaços PCA, LDA, Kernel-PCA (com Kernel Gaussiano, Kernel Polinomial, Kernel Sigmóide), Eigen-Isomap ou Fisher-Isomap. Essas técnicas de extração de atributos foram aplicadas sobre a imagem original, imagem com realce de contraste e imagem com realce de contraste e histograma equalizado. Para combinar esses classificadores foram verificadas a Regra do Produto e Votação por Maioria. Utilizando *3-Fold Cross Validation*, o melhor resultado para o FERET foi 93,44% de taxa de acerto usando Regra do Produto e Fisher-Isomap enquanto que usando 10 repetições aleatórias para

o Yale foi 96,44% de taxa de acerto com a Regra do produto e Kernel-PCA com Kernel Polinomial de Segunda Ordem. O conjunto FERET foi dividido igualmente entre conjuntos de treinamento, validação e teste com 1 amostra por indivíduo para cada conjunto de um total de 600 indivíduos. O conjunto Yale foi dividido em 4, 3 e 4 imagens por pessoa para conjuntos de treinamento, validação e teste, respectivamente.

3.5.2 Técnicas Baseadas em Componentes

Em (MIRHOSSEINI, YAN *et al.*, 1998) também foi proposta a utilização da integral fuzzy de Choquet para realizar a fusão dos classificadores, porém agora combinando componentes faciais de olhos, nariz e boca. Antes de aplicar o processo de classificação, uma etapa de detecção dos componentes faciais ocorre. Essa etapa consiste, para uma dada imagem de face, em: 1) detecção da região da cabeça utilizando *graph templates* e filtros de Gabor e extração dos limites da face usando *snakes*; 2) detecção e extração das regiões dos olhos e da boca aplicando *graph templates* e uma rede neural artificial retro-propagação para os olhos e um modelo deformável para a boca; 3) detecção da linha base do nariz aplicando uma projeção integral horizontal na região compreendida entre os olhos e a boca localizadas anteriormente. Com foco no problema da rotação no eixo y, utilizando as regiões faciais localizadas e um modelo 3D de uma cabeça é possível estimar a rotação facial como uma tentativa de compensar esse problema. Quanto à fusão de classificador fuzzy, essa abordagem difere da proposta em (ARBUCKLE, LANGE *et al.*, 1995) no cálculo da densidade de medida fuzzy na etapa de treinamento e na medida de similaridade utilizada. Aqui, a medida de similaridade consiste em uma medida de correlação baseada na norma L_1 que é menos sensível ao ruído do que a distância euclidiana, sendo essa medida definida como

$$C_{L_1}(x, y) = 1 - \frac{\sum_i |x_i - y_i|}{\sum_i (|x_i| + |y_i|)}.$$

Já a densidade de medida fuzzy é baseada no cálculo de uma matriz de confusão $P^k = [p_{ij}^k]$ para cada classificador r_k e definida, inicialmente, por

$$g_{i/k} = \frac{p_{ii}^k}{\sum_{j=1}^I p_{ij}^k}.$$

Em seguida, a densidade de medida fuzzy passa por uma fase de atualização que pretende “considerar o efeito da relação entre as frequências de classificação correta e desclassificação” já que “é sempre observado que quando um classificador é mais robusto do que outros classificadores em classificar uma certa característica, ele é também mais propenso a erros na classificação de outras características do que outros classificadores menos robustos”. Para fazer essa atualização, são calculados dois fatores $\delta^{k/m}$ e $\gamma^{k/m}$, definidos por

$$\delta^{k/m} = \begin{cases} 1, & i/k = j/m \\ \frac{p_{i/k, i/k}^k - p_{i/k, j/m}^k}{p_{i/k, i/k}^k}, & i/k \neq j/m \end{cases}$$

$$\gamma^{k/m} = \begin{cases} 1, & p_{i/k, j/m}^k \leq p_{i/k, j/m}^m \text{ ou } i/k = j/m \\ \frac{p_{i/k, j/m}^m}{p_{i/k, j/m}^k}, & p_{i/k, j/m}^k \geq p_{i/k, j/m}^m \text{ e } i/k \neq j/m \\ \xi, & p_{i/k, j/m}^k = 0 \end{cases}$$

onde $\xi = 0.0001$ e a densidade fuzzy final é

$$g_{i/k}^* = g_{i/k} \times (\delta^{k/1} \times \dots \times \delta^{k/M})^{w_1} \times (\gamma^{k/1} \times \dots \times \gamma^{k/M})^{w_2},$$

sendo w_1 e w_2 pesos exponenciais. Para avaliar o desempenho dessa técnica o banco de dados ORL foi utilizado. Diversos experimentos foram feitos, testando todas as combinações possíveis de componentes faciais, sendo que foi obtida uma taxa de reconhecimento geral de 85% utilizando todas as componentes e 200 amostras para treinamento (5 amostras por indivíduo) e 160 para teste (4 amostras por indivíduo).

Uma abordagem combinando SVMs com Kernel Polinomial de grau 2 treinados individualmente sobre 10 componentes faciais (mostrados na Figura 3.31) extraídos automaticamente de imagens faciais de um sistema de vigilância é apresentado em (IVANOV, HEISELE *et al.*, 2004). As imagens faciais, após serem detectadas, passam por uma equalização de histograma e depois são reamostradas para 70 x 70 pixels. Exemplos dessa base de dados são apresentados na Figura 3.32. Para realizar a combinação, as saídas dos classificadores individuais são normalizadas aplicando a função softmax. Foi proposto um novo *framework* Bayesiano para combinar as fontes usando a “distribuição empírica do erro derivado da matriz de confusão”. Esse combinador é dado pela seguinte equação

$$P(w|x) = \sum_{i=1}^C \sum_{k=1}^K P(w|\tilde{w}_k, x, \lambda_i) P(\tilde{w}_k|x, \lambda_i) P(\lambda_i|x)$$

$$\approx \sum_{i=1}^C \left[\sum_{k=1}^K P(w|\tilde{w}_k, \lambda_i) P(\tilde{w}_k|x, \lambda_i) \right] P(\lambda_i|x),$$

onde C é o número de classificadores, \tilde{w} é uma variável aleatória, w é o rótulo da classe verdadeira, x é uma dada observação, K é o número de classes, λ_i “são ponderados por uma medida de confiança imposta sobre o desempenho do classificador”. Além do combinador proposto, foram verificadas a Regra da Soma, a Regra do Produto e Votação por Maioria. Essas 3 regras foram aplicadas diretamente após a normalização dos dados por softmax e também após aplicar a equação acima, o que gerou os combinadores denominados Regra da Soma Ponderada, Regra do Produto Ponderado e Votação por Maioria Ponderada. Foi identificado que todas as 6 combinações apresentaram um resultado próximo, com cerca de 90% de taxa de aceitação para um $\log P(\text{erro})$ igual a -7 . Essa semelhança no resultado entre os combinadores ponderados e não ponderados pode ser explicada “pelo fato que os classificadores individuais são muito fortes sobre esse conjunto de dados e suas matrizes de confusão são muito próximas à identidade, e conseqüentemente, tem pouco efeito nos *scores* combinados”. Vale ressaltar que foi verificada também a utilização dos vetores componentes empilhados em um único vetor e classificado com o SVM, porém o mesmo apresentou resultado inferior à combinação no nível de decisão. E ainda, o classificador de face inteira e das narinas foram, respectivamente, o melhor e o os piores classificadores de componentes.

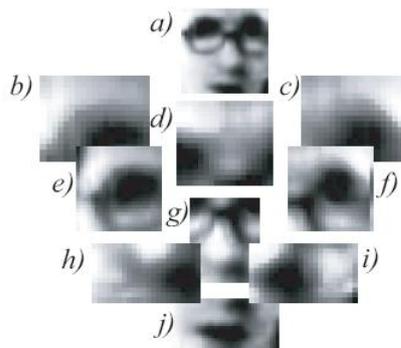


Figura 3.31. Exemplo dos componentes faciais extraídos: a) *bounding box*; b, c) sobrancelhas; d) ponte do nariz; e, f) olhos; g) nariz; h, i) narinas e j) boca. Retirado de (IVANOV, HEISELE *et al.*, 2004).



Figura 3.32. Exemplos do conjunto de dados usado por Ivanov e seus colegas. Os exemplos das duas primeiras linhas são do conjunto de treinamento e as restantes, do conjunto de teste. Retirado de (IVANOV, HEISELE *et al.*, 2004).

Uma abordagem de dividir para conquistar foi proposta em (TOYGAR e ACAN, 2004a; b). A idéia consiste em dividir uma imagem facial em 5 segmentos horizontais igualmente distribuídos como mostrado na Figura 3.33 e em atribuir importâncias diferentes para áreas faciais, já que a utilização de métodos estatísticos baseados em aparência e aplicados à faces inteiras consideram todas as áreas faciais com pesos iguais. A cada segmento é aplicado PCA, LDA ou ICA (que são métodos estatísticos baseados em aparência para extração de características) e em seguida classificados por NN. Diversos combinadores foram testados sobre a base de dados FERET: *Borda Count*, Votação por Maioria, Maior Ranking, Regra da Soma, Regra da Mediana e Regra do Máximo. A Regra da Soma foi quem apresentou o melhor desempenho e aplicado sobre o espaço PCA obteve cerca de 89% com 50 imagens frontais na galeria (2 amostras por indivíduo) e 100 imagens frontais no conjunto de prova. Outro resultado relatado foi de cerca de 42% com 50 imagens frontais na galeria e 300 imagens com alta variabilidade em iluminação, escala e pose no conjunto de prova. Todas as imagens foram recortadas para 45 x 35 pixels.

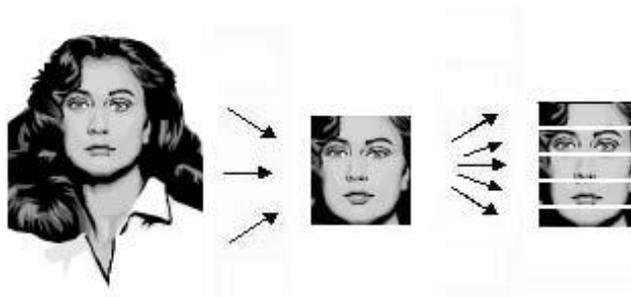


Figura 3.33. O processo de divisão da imagem recortada em 5 segmentos: imagem original, imagem recortada e imagem dividida. Retirado de (TOYGAR e ACAN, 2004a).

Em (MU, WATTA *et al.*, 2005) foi proposto combinar medidas de similaridade local dadas por *Template Matching* comparando as Regras da Soma, Votação por Maioria e Votação por Maioria Ponderada. Para gerar classificadores locais, as imagens são particionadas em partes não sobrepostas. O combinador de Votação por Maioria Ponderada proposto considera a quantidade de partições que são classificadas em um determinado rank e são ponderadas por $\alpha_r = P(w_t | rank = r)$, onde t é a classe considerada. Esse algoritmo foi testado nos bancos de dados CNNL (*Computation and Neural Networks Laboratory* da *Wayne State University*) e FERET. O banco de dados CNNL contém 1300 indivíduos com 10 amostras por pessoa com variações em expressão facial (veja a Figura 3.34 como exemplo). Para ambos os bancos de dados, a Votação por Maioria Ponderada obteve o melhor desempenho. Para o CNNL, a galeria continha uma amostra por classe com expressão normal e dois conjuntos de teste foram utilizados, sendo que um continha uma outra imagem com expressão normal por classe e o segundo uma amostra aleatória por classe. Para o primeiro conjunto de testes foi atingida uma taxa de classificação correta de 100% e cerca de 99% para 100 e 1000 indivíduos na base de dados, respectivamente. Já para o segundo conjunto de testes foi obtida uma taxa de classificação correta de cerca de 99% e aproximadamente 87% para 100 e 1000 indivíduos na base de dados, respectivamente. Aplicando o algoritmo sobre o banco de dados FERET foi obtida uma taxa de classificação correta de 96,1% usando todo o conjunto Fa como treinamento e Fb como teste. Vale ressaltar ainda que as taxas obtidas pelo CNNL continham uma modificação no algoritmo a fim de incorporar robustez a pequenos deslocamentos e rotações. Essa modificação consistia em computar também a distância em posições próximas da janela, selecionando a menor distância para aquele local.



Figura 3.34. Algumas amostras de um indivíduo do banco de dados CNNL. Retirado de (MU, WATTA *et al.*, 2005).

3.5.3 Técnicas Híbridas

Lucas desenvolveu um novo classificador para reconhecimento de padrões denominado “Classificador de n -tuplas contínuas” e o aplicou ao reconhecimento de faces

(LUCAS, 1997b; a; 1998). As n -tuplas são geradas a partir da amostragem aleatória do espaço de entrada d -dimensional selecionando m tuplas de n dimensões (a Figura 3.35 ilustra esse procedimento). Essa etapa de geração das tuplas pode ser entendida como a etapa de treinamento. A classificação é efetuada associando a imagem de teste z à classe que minimiza a distância r_c , dada por

$$r_c = \sum_{j=1}^m \min_k D(y_{jk}^c, z_j),$$

onde m é o número de n -tuplas, $z_j = \{a_{j1}, a_{j2}, \dots, a_{jn} \mid 1 \leq a_{ji} \leq d\}$ é a j -ésima tupla da imagem de teste z , y_{jk}^c é a j -ésima tupla da k -ésima amostra da classe C e D é a distância de Manhattan (*city-block*), definida por

$$D(x, y) = \sum_{i=1}^n |x_i - y_i|.$$

Além da versão apresentada, existe uma outra versão, denominada de versão compilada. A diferença entre elas é que a versão compilada utiliza-se da quantização do espaço de entrada para σ níveis e uma *look-up-table* com endereços b_j e memória n_{cj} , que armazenarão as distâncias mínimas entre cada classe c e n -tupla j e todas as σ^n combinações de tuplas possíveis que podem ser acessadas por $n_{cj}[b]$ e será gerada durante a etapa de treinamento. O mapeamento de um padrão de entrada x com d dimensões para um endereço $b_j(x)$ pode ser calculado para cada n -tupla pela equação

$$b_j(x) = \sum_{k=1}^n x(a_{jk}) \times \sigma^{k-1}.$$

Dessa forma, o tempo de treinamento da versão compilada é maior que a da versão não compilada, entretanto o tempo de teste será menor. Em analogia com a Combinação de Classificadores, o classificador de n -tuplas contínuas pode ser entendido como um classificador Nearest Neighbour com a Regra do Mínimo para combinar as saídas desses classificadores. Essa técnica pode ser classificada como Híbrida, pois apesar de não utilizar explicitamente componentes faciais (o que a classificaria como sendo uma técnica baseada em componentes), ela o faz implicitamente ao utilizar diferentes pontos da imagem sem repeti-los e cada classificador analisa alguns pontos da face. Além disso, quando “ $n = d$ e $m = 1$,”

essa técnica implementa exatamente o classificador Nearest Neighbour”. Portanto, n-tuplas contínuas tanto pode ser baseada em componentes quanto holística. Na avaliação dessa técnica, a base de dados ORL foi utilizada, sendo que os melhores resultados relatados foram taxas de acerto de 97,3% (média de 5 experimentos, com $n = 4$ e $m = 500$) (LUCAS, 1997b) e 96,41% em versão compilada (média de 100 experimentos, com $n = 2$, $m = 500$ e $\sigma = 16$) (LUCAS, 1997a; 1998). Em ambos os casos, cada partição era composta por 5 amostras de treinamento e 5 amostras de teste para cada classe. Como foi ressaltado pelo autor, essa técnica, por ser muito rápida para realizar as etapas de treinamento e teste, pode ser utilizada para aplicações de tempo real.

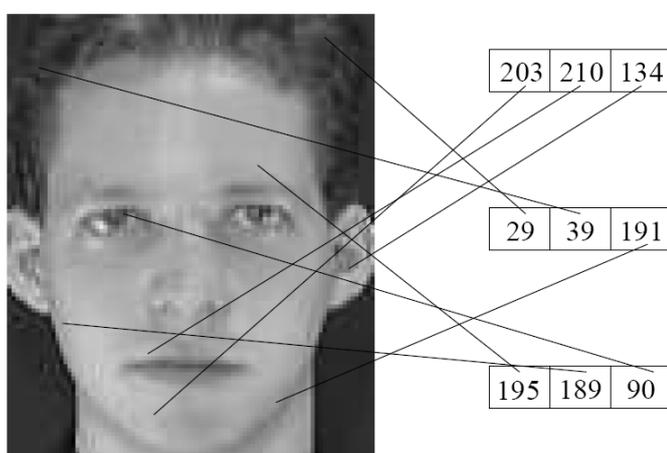


Figura 3.35. Ilustração do processo de amostragem de n-tuplas contínuas. Retirado de (LUCAS, 1997b).

Uma combinação ou comitê de redes neurais MLP (Multilayer Perceptron) foi proposto por Caeanu (CALEANU, 2000). Para isso, foram treinadas 3 redes neurais MLP individualmente, pelo algoritmo do gradiente conjugado escalado e com características extraídas por um operador de interesse que calcula 5 variâncias direcionais nas diversas direções para blocos não sobrepostos 5x5. As áreas da imagem original utilizadas para extrair as características poderiam ser sobrepostas ou não. Para realizar a combinação das saídas das redes neurais foram verificadas a Regra da Média e uma nova MLP. Utilizando a base de dados ORL e somente 10 dos 40 indivíduos dessa base e ainda com a imagem reduzida pela metade, foram executados experimentos para cada arquitetura de sistema utilizado, considerando 1, 2 ou 3 imagens por pessoa para o conjunto de treinamento e as restantes para teste. O melhor resultado médio de 10 experimentos foi obtido pela Regra da Média com uma

taxa de acerto de 97,143%, utilizando 3 amostras por pessoa no conjunto de treinamento, a imagem inteira (ou sobreposição completa) e a arquitetura de cada MLP era 600-25-10 (ou seja, 600 neurônios na camada de entrada, 25 na camada escondida e 10 na saída, respectivamente), sendo que o Erro Médio Quadrático (MSE) considerado no treinamento das redes foi de 0,01%.

Uma abordagem combinando três técnicas de reconhecimento de face de características globais e uma local foi proposta em (HUANG, YUEN *et al.*, 2002). A técnica local utilizada é o Wavelet de Gabor Local que utiliza 15 pontos importantes da face como mostrado na Figura 3.36a. Já as técnicas globais são PCA, ICA e o *Spectroface*, que é um método “que combina as transformadas Wavelet e de Fourier para a extração das características”. Foram verificados 3 métodos para normalizar as saídas dos classificadores e tornar adequada a combinação: normalização linear, exponencial e exponencial-linear. E finalmente, para combinar os classificadores foram testadas 4 regras: Produto, Soma, Mínimo e Máximo. Utilizando o banco de dados Yale, com imagens redimensionadas para 128x128 pixels, os melhores resultados relatados foram obtidos utilizando a Regra da Soma e métodos de normalização exponencial e exponencial-linear com 94,16% de taxa de classificação correta considerando somente a amostra com expressão normal para treinamento e todas as outras para teste, exceto as com expressões “sem óculos” e “feliz”. Foi citado também, que utilizando a base de dados ORL, foram obtidos resultados semelhantes.

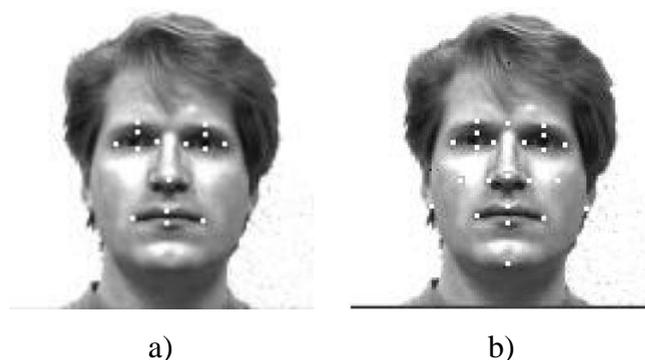


Figura 3.36. Pontos importantes na face para o reconhecimento pelo método Wavelet de Gabor Local: a) 15 marcos e b) 23 marcos. Retirado de (HUANG, YUEN *et al.*, 2002; 2004), respectivamente.

Uma extensão do trabalho (HUANG, YUEN *et al.*, 2002) foi proposto em (HUANG, YUEN *et al.*, 2004). Nesse novo trabalho são utilizados os mesmos classificadores

individuais, entretanto, para o Wavelet de Gabor Local são utilizados 23 pontos importantes como mostrado na Figura 3.36b. Para a normalização das saídas dos classificadores o método exponencial-linear é utilizado e um novo método de normalização baseado na distribuição normal denominado Método de Normalização de Distribuição Ponderada Gaussiana (DGWNM) foi proposto. Como combinador, além das regras utilizadas no artigo anterior, a Regra da Soma Ponderada também foi utilizada. Além disso, um algoritmo de seleção dos classificadores individuais foi proposto baseado na idéia básica que “se um classificador é redundante, a precisão aumentará se aquele classificador for removido da combinação”. Esse sistema foi avaliado nos bancos de dados Yale, ORL e FERET. Respectivamente, para as bases dados Yale e ORL, o melhor resultado obtido utilizando os 4 classificadores foi 95% e 85% com a normalização DGWNM e a Regra da Soma Ponderada. Já para o FERET, foi verificado o método de seleção de classificadores e o melhor resultado relatado foi 92,5% utilizando os classificadores *Spectroface*, ICA e Wavelet Local de Gabor, a normalização DWGNM e a Regra da Soma Ponderada para combinação. Esses resultados foram obtidos considerando a seguinte divisão das bases de dados: a) Yale: uma única amostra por classe para treinamento e as restantes para teste; b) ORL: uma imagem de face por classe para treinamento (a de visão mais frontal) e 7 imagens por indivíduo para teste e c) FERET: 30 dos 70 indivíduos usados para treinamento e algoritmo de seleção de classificadores e o restante para testar o esquema final.

Em (KHUWAJA, 2002) foi proposta uma combinação de redes LVQ treinadas sobre 6 componentes faciais extraídos manualmente (veja a Figura 3.37) mais a face completa utilizando Votação por Maioria e Regra da Soma para Ranking. As redes LVQ são otimizadas por uma técnica de podas de neurônios da camada escondida, já que uma das dificuldades no projeto de redes neurais é justamente a determinação do número de neurônios da camada escondida que maximiza o desempenho da rede. Inicialmente, foi definido que a camada escondida teria $C \times N_i$ neurônios (onde C é o número de classes e N_i é o número de amostras por classe no conjunto de treinamento). E a poda focaliza na remoção de neurônios cegos que possuem informação redundante e não contribuem para uma melhor generalização, aliás prejudicam. Para determinar se um neurônio é cego, é calculado o desvio padrão dos vetores pesos para cada neurônio da camada escondida e verificado se o valor é menor que um determinado limiar que varia de acordo com a região de atributos considerada. Para avaliar o desempenho dessa abordagem foi utilizado o banco de dados de faces AR, sendo que as imagens foram redimensionadas aplicando a mesma técnica usada em (TOLBA e ABU-

REZQ, 2000). Vários experimentos foram realizados para determinar os valores dos parâmetros e a configuração da rede mais adequados e eficientes. Foi relatada uma taxa de classificação correta de 98,03% utilizando 5 imagens por pessoa para treinamento e outras 5 imagens por pessoa para teste de um total de 100 indivíduos, sendo que as redes LVQ foram treinadas por 10.000 épocas com uma taxa de aprendizado de 0,7 e as imagens de face inteira utilizadas haviam sido redimensionadas para 64x64 pixels.

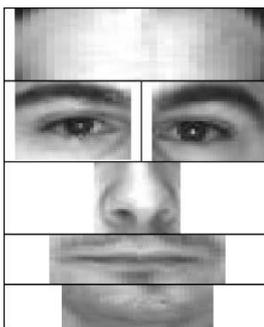


Figura 3.37. Componentes Faciais extraídos manualmente. Retirado de (KHUWAJA, 2002).

Em (ZHOU e ZHANG, 2002) foi apresentado um *framework* para fusão baseado em Máquinas de Vetores de Suporte (SVM), verificando-se a utilização de Kernel Polinomial e Kernel Gaussiano. São combinados 3 classificadores: 1) *Eigenfaces* com 16 componentes principais extraídos, 2) uma técnica baseada em atributos geométricos que extrai um vetor com 24 atributos e utiliza o quadrado da distância de Mahalanobis como medida de similaridade e 3) uma técnica de *template matching* que utiliza áreas do nariz, olhos e boca, sendo que, para comparar duas faces, as distâncias entre cada componente facial correspondente são somadas e a medida de similaridade entre uma dada imagem de face e uma pessoa registrada é dada pela menor soma entre uma dada face e uma dada amostra daquela classe. Um banco de dados de faces próprio, composto de 60 classes e 10 amostras por classe com alguma variação de pose, iluminação e expressão foi dividido de forma a ser utilizado para estimar os parâmetros das SVMs e verificar o desempenho do método. O conjunto de treinamento era composto por 4 amostras por classe e as restantes formavam o conjunto de teste. Para estimar os parâmetros das SVMs foram utilizados apenas 10 indivíduos e os 50 restantes foram tomados para teste, obtendo-se uma taxa de acerto de 95,7% utilizando Kernel Polinomial.

O combinador Borda Count foi utilizado para combinar classificadores de *Eigenfaces*, Transformada Discreta do Cosseno (DCT), Modelos de Markov Escondidos (HMM) e

Eigenobjects em (LEMIEUX e PARIZEAU, 2003). O HMM utiliza, como entrada para o sistema, coeficientes DCT computados nas regiões apresentadas pela Figura 3.38. A técnica *Eigenobjects* é semelhante à *Eigenfaces*, porém na primeira são extraídos atributos de imagens de nariz e olhos (exemplos são mostrados na Figura 3.39). A justificativa de se utilizar *Eigenobjects* consiste de ganhar robustez contra oclusões. Já para o DCT, o motivo está na menor influência sofrida por variações de iluminação. Vale ressaltar que foi utilizado o NN com norma L1 (*city-block*) para classificar os atributos extraídos por DCT, *Eigenfaces* e *Eigenobjects*. Para combinar as respostas dessas técnicas foi utilizado o combinador Borda Count com uma pequena modificação: as listas com a classificação das classes de cada classificador individual foram limitadas ao rank 10. Dessa forma, no cálculo da média, ranks maiores que 10 foram definidos como 10 a fim de evitar que *outliers* prejudicassem o desempenho da combinação. O banco de dados FERET foi utilizado para avaliar o desempenho dessa abordagem. As imagens dessa base passaram pelas seguintes etapas de pré-processamento: 1) Os centros dos olhos foram extraídos e as imagens rotacionadas para que a linha dos olhos se tornasse alinhada horizontalmente; 2) As imagens foram escaladas de forma que a distância entre os olhos fosse de 70 pixels; 3) As imagens foram recortadas para 130 x 150 pixels considerando a face centralizada e os olhos na linha 45; 4) Cabelo e fundo restantes foram removidos aplicando uma máscara oval e 5) Foi aplicada uma equalização de histograma. Vários experimentos foram realizados sobre essa base considerando 200 *Eigenfaces*, HMM (5 superestados e 3-6-6-6-3 estados embutidos), 75 *Eigenobjects* (25 por componente facial utilizado) e 192 coeficientes DCT. O melhor resultado obtido foi cerca de 87%, combinando *Eigenfaces*, *Eigenobjects* e HMM.

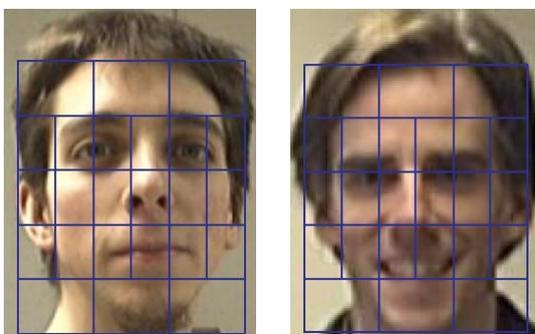


Figura 3.38. Alguns exemplos da segmentação inicial para o HMM. Retirado de (LEMIEUX e PARIZEAU, 2003).

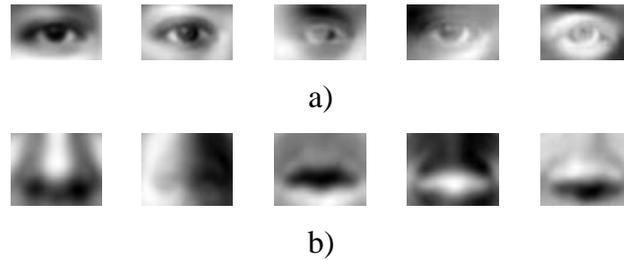


Figura 3.39. Imagem média e os 4 primeiros *Eigenobjects* para a) Olho esquerdo e b) o nariz. Retirado de (LEMIEUX e PARIZEAU, 2003).

Em (RAO, GHADI *et al.*, 2004) um novo esquema de fusão em nível de decisão foi proposto para combinar classificadores de componentes de Fisher extraídos de *eigencomponents* de 3 características faciais: a face inteira, a região dos olhos e um mapa de bordas da imagem inteira gerada calculando a soma dos componentes horizontal e vertical de um operador diferencial de primeira ordem derivado da função Gaussiana 1-D, após aplicar um filtro Gaussiano 1-D para suavização. Além disso, uma técnica de equalização de histograma em blocos foi utilizada para efetuar a normalização da intensidade numa etapa de pré-processamento. O classificador individual utilizado foi a Distância no Espaço de Atributos (DIFS), ou seja, a classe que fornece a melhor descrição da imagem de teste é aquela que minimiza a norma L2 da diferença entre a imagem de teste e o vetor média da classe no espaço de Fisher. O combinador proposto consiste da probabilidade conjunta de variáveis independentes entre as probabilidades de cada espaço de características, sendo que cada probabilidade de cada característica $P(i)$ é dada por

$$P_{caraterística} = [P_i(\Delta'_i) P_{ordem}(i)]_{caraterística}, \quad 1 \leq i \leq I,$$

onde I é o número de classes, $P_{ordem}(i)$ é definida por uma distribuição Gama e é responsável por “associar um peso adequado a uma classe dependendo do *ranking* posicional como também do menor valor DIFS” e “ $P_i(\Delta'_i)$ descreve o quanto próximo o valor DIFS do i -ésimo *ranking* está do *top rank*”. O classificador associa o padrão à classe que maximiza a probabilidade conjunta, desde que ela seja maior que um limiar. Caso contrário, é considerado um impostor. Esse esquema foi testado sobre o banco de dados FERET, sendo que o conjunto de treinamento era composto de 482 indivíduos com 3 amostras por classe e a galeria utilizada foi o conjunto Fa. Foram relatadas as seguintes taxas de reconhecimento correto 98,3%, 68%, 59% e 54% para os conjuntos de teste FaFb, *duplicate I*, FaFc e *duplicate II*, respectivamente com valores de ERR 31%, 30%, 34% e 12% para face, mapa de bordas, olhos e fusão, respectivamente.

Em (KWAK e PEDRYCZ, 2005) duas novas abordagens são propostas utilizando fusão de informação fuzzy através da Integral de Choquet. A primeira abordagem considera o domínio dos componentes faciais nariz, boca e região dos olhos e a face inteira enquanto a outra considera o domínio Wavelet com as 4 subimagens geradas pela decomposição Wavelet de Daubechies. Para ambos os esquemas, LDA de componentes principais é aplicado para extrair atributos e a distância Euclidiana é utilizada como medida de similaridade, obtendo-se assim 4 classificadores para cada esquema. Para gerar a integral de Choquet são calculados a função h e a densidade de medida Fuzzy g^i que será utilizada no cálculo da medida fuzzy g_λ . A função h é dada por

$$h(y_{ik}) = \frac{1}{N_k} \sum_{\mu_{ij} \in C_k} (\mu_{ij})$$

$$\mu_{ij} = \frac{1}{1 + \left(\frac{d_{ij}}{\bar{d}_i}\right)},$$

onde $i = 1, \dots, 4$ e denota o classificador, N é o número de imagens no conjunto de treinamento, d_{ij} é a distância Euclidiana entre o vetor de atributos da imagem de treinamento e a de teste, \bar{d}_i é a distância média entre todos os valores de distância no classificador i , N_k é o número de amostras na k -ésima classe C_k . Já a densidade de medida Fuzzy g^i é executada da seguinte maneira

$$\begin{cases} g^i = \beta p_i, & i = 1 \\ g^i = (1 - \beta)p_i, & i = 2, 3, 4 \end{cases}$$

onde $i = 1$ é o classificador da característica imagem inteira ou da subimagem LL da decomposição Wavelet, p_i é a taxa de classificação do especialista obtido usando *leave-one-out* e $\beta \in [0, 1]$ é um fator que tenta balancear entre os resultados de classificação produzidos por classificadores baseados em componentes e por aqueles baseados em imagens globais. Essas abordagens foram validadas usando *3-fold cross-validation* sobre o banco de dados FERET e além da fusão de informação Fuzzy, foram também utilizadas as Regras da Média, Média Ponderada e a Regra do Máximo para combinar os classificadores. Foram selecionadas 600 imagens frontais de 200 indivíduos, sendo que havia 3 imagens por pessoa incluindo alguma variação na expressão e iluminação. As imagens foram recortadas e redimensionadas para 94 x 82 pixels. Os componentes faciais continham 20x50, 22x50 e 20x30 pixels para

olhos, nariz e boca, respectivamente e 47x41 pixels para cada subimagem da decomposição Wavelet. Para ambas as abordagens, o método usando integral fuzzy obteve o melhor desempenho com taxa de reconhecimento de 95% e 93,83% para a primeira ($\beta = 0,6$) e segunda abordagem ($\beta = 0,7$), respectivamente. Esse desempenho foi atingido usando 79 componentes principais e 33 vetores discriminantes, além do conjunto de dados ter sido dividido em 400 imagens para treinamento e 200 para teste.

Uma abordagem baseada em classificadores SVM foi proposta por (HEISELE, SERRE *et al.*, 2007). Sobre 14 componentes faciais (ver a Figura 3.40) extraídos por um detector de componentes também baseado em SVM, mais a face completa são aplicados classificadores SVM com Kernel Polinomial de segundo grau. Para combinar esses classificadores, que são 15 por classe, já que o SVM é binário, SVM com Kernel Linear é usado, sempre na estratégia “um contra todos”. Cada componente utilizado tem dimensão 40 x 40 pixels e tem o seu histograma equalizado. Esse método foi aplicado ao CMU-PIE e comparado com outras técnicas de fusão de classificadores tal como Votação por Maioria, Regra do Produto, Regra da Soma, atributos empilhados e outros ainda, sendo que o esquema proposto obteve o melhor desempenho com 89,25% de taxa de reconhecimento correto. Vale ressaltar ainda, que para aumentar o número de amostras de treinamento, foram utilizados modelos de cabeça 3D para sintetizar novas amostras em diversas poses. Para o conjunto de treinamento foram utilizadas 7.040 imagens sintéticas e 200 imagens por indivíduo para teste. E ainda, essa técnica foi aplicada em imagens com oclusões para verificar a interferência das mesmas no momento da detecção dos componentes. Foi identificado que isso realmente prejudicou o desempenho.

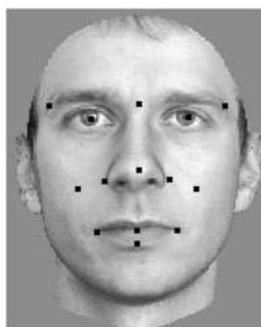


Figura 3.40. Pontos de referência usados para o aprendizado dos componentes. Retirado de (HEISELE, SERRE *et al.*, 2007).

Regra da Soma e Votação por Maioria foram utilizadas para combinar Classificadores *Moving Window* (MWC) que são baseados em n-tuplas realçadas em (SIRLANTZIS, HOQUE *et al.*, 2008). Os extratores de características utilizados são o Bit-Plane e um Bit-Plane aleatório (motivado pelo método de Subspaço Aleatório) que consiste em definir aleatoriamente um template com o bit a ser considerado de cada nível de cinza em cada ponto como mostrado na Figura 3.41. Para entender melhor as imagens selecionadas veja a Figura 3.42. Aplicando esse método sobre o ORL com 5-cross validation e coeficiente Kappa foi obtida com a Regra da Soma, uma taxa de reconhecimento correto de 97,1% utilizando o bit-plane padrão, selecionando as camadas de 4 a 8. Além disso, o ORL foi dividido igualmente entre conjuntos de teste e treinamento.

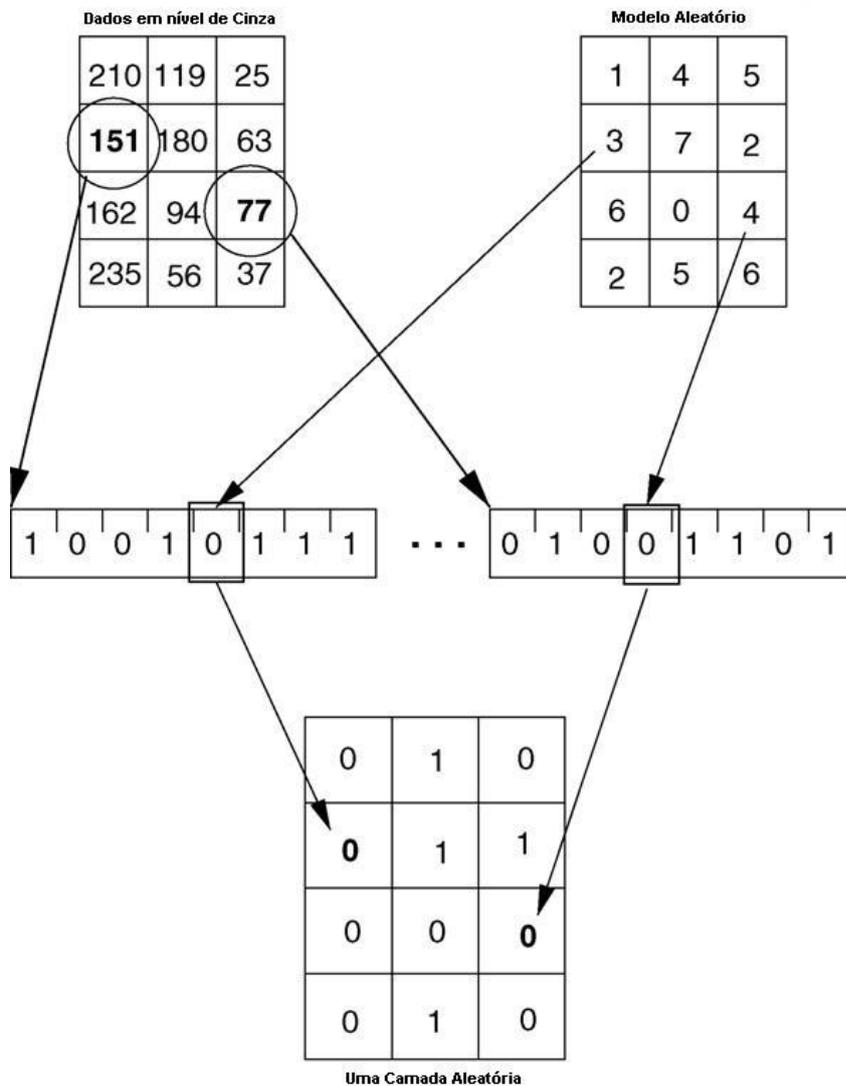


Figura 3.41. Esquema de Quantização de Características Aleatória (Bit-plane aleatório). Adaptado de (SIRLANTZIS, HOQUE *et al.*, 2008).



(a)



Camada 8 (Bit Mais Significativo)



Camada 7



Camada 6



Camada 5



Camada 4



Camada 3



Camada 2



Camada 1

(b)



Camada Aleatória 1



Camada Aleatória 3



Camada Aleatória 7



Camada Aleatória

10

(c)

Figura 3.42. Efeito da quantização em escala de nível de cinza: a) Imagem Original; b) Quantização Ordenada (Bit-Plane) e c) Quantização de Camada Aleatória (Bit-Plane Aleatório). Adaptado de (SIRLANTZIS, HOQUE *et al.*, 2008).

3.5.4 Discussões

Nesta seção, foram apresentadas diversas abordagens de sistemas de reconhecimento de faces usando combinação de classificadores. Na Tabela 3.1 encontra-se um resumo das técnicas de reconhecimento de faces humanas que utilizam combinação de classificadores que foram abordadas na Seção 3.5. Devido à utilização de diferentes bases de dados e metodologias para a estimação de erros, não é possível fazer comparações entre as diversas abordagens. De qualquer forma, até o momento podemos afirmar que não existe uma abordagem fortemente superior.

Entretanto, em geral, os sistemas aqui apresentados mostram, na prática, que a combinação de classificadores melhora o desempenho da classificação com relação a um classificador individual, o que mostra maior robustez em relação às diversas variações, principalmente, pose e iluminação. Além disso, muitos dos sistemas consideram classificadores simples e fáceis de implementar para atingir um desempenho superior e ainda, reduzir a variância do erro estimado (mais estabilidade) (MARCIALIS e ROLI, 2004; CHAWLA e BOWYER, 2005; HUANG, YUEN *et al.*, 2006).

Tudo isso indica o potencial de se utilizar combinação de classificadores em aplicações de reconhecimento de faces. Baseado nisso, um sistema de combinação de múltiplos classificadores variando, principalmente, as técnicas de extração de atributos e os classificadores é proposto no próximo capítulo.

Tabela 3.1. Resumo dos métodos de reconhecimento de faces humanas usando combinação de classificadores.

Método	Extração de Atributos	Classificação	Combinação	Banco de Dados	Taxa de Acerto
Técnicas Holísticas					
(ARBUCKLE, LANGE <i>et al.</i> , 1995)	Concatenação de Coeficientes de Autocorrelação e depois LDA para diversas resoluções	1-NN	Integral de Sugeno	Próprio (116 classes)	98,4%
(ACHERMANN e	PCA	1-NN	Normalização	Próprio (30 classes)	98,7%

BUNKE, 1996b; a)	Janela Deslizante (Sentido Vertical)	HMM	por Transformação Logística e Combinação por Regra da Soma de Scores		
(TOLBA e ABU- REZQ, 2000)	-	LVQ e RBF	Baseado em Regras	ORL (40 classes)	99,5%
(MARCIALIS e ROLI, 2002; MARCIALIS e ROLI, 2003)	PCA	1-NN	Normalização por Média de Componentes e Combinação por KNN	ORL (40 classes)	97,25%
			Normalização por União de Componentes e Combinação por KNN	Yale (15 classes)	84,22%
(MARCIALIS e ROLI, 2004)	PCA-LDA			MIT (16 classes)	79%
(RYU e OH, 2002)	PCA	NFA e Ensemble de redes neurais MLP	Baseado em Regras	ORL (40 classes)	96,33%
(JING e ZHANG, 2003)	Transformada Wavelet Ortogonal de Daubechies obtendo 4 sub- bandas e, em seguida, LDA	1-NN	MCC	NUST603 (18 classes)	96,4%

(LU, WANG <i>et al.</i> , 2003)	PCA	Distância do Cosseno	Rede Neural RBF	Próprio coletado de ORL, Yale, AR e NLPR+MSU (206 classes)	90,2%
	ICA				
	LDA				
(KIM e KITTLER, 2004)	PCA-LM-LDA	1-NN	Regra da Soma	XM2VTS (295 classes)	73,2% na mesma sessão e 54,4% em para sessões diferentes
	LLDA				
	GDA				
	3D LUT-LDA				
(KIM e KITTLER, 2006)	PCA-LM-LDA	1-NN	Regra da Soma	XM2VTS (295 classes)	70% na mesma sessão e 42% em para sessões diferentes
	LLDA				
	GDA				
	3D LUT-LDA				
(ZHANG, SHAN <i>et al.</i> , 2004)	PCA-LDA	Não Especificado	Regra do Produto	FERET (1.196 classes)	98,2%
	Transformada de Gabor-PCA-LDA				
(ZHAO, HUANG <i>et al.</i> , 2004a; b)	Operador de Interesse	Redes Neurais RBF	Votação por Maioria	ORL (40 classes)	97,7%
	PCA				
	Kernel LDA				
(CHAWLA e BOWYER, 2005)	PCA	Ensemble de 1-NN com Distância do Cosseno no Espaço de Mahalanobis como medida	Regra do Mínimo (mínimas distâncias)	Notre Dame (121 classes)	76,9%

		de Similaridade			
(EBRAHIMPOUR, EHTERAM <i>et al.</i> , 2005)	PCA	Redes Neurais MLP (uma por classe)	Regra do Máximo	ORL (40 classes)	100%
				Yale (15 classes)	100%
(KHAN, IBRAHIM <i>et al.</i> , 2005)	Banco de Filtros Direcionais (DFB)-PCA	Não Especificado	Medida de Correlação Cruzada	ORL (40 classes)	97%
(LUMINI e NANNI, 2005)	PCA	Parzen Window	Regra da Média	YaleB (10 classes)	99,7%
	DCT				
	Filtro de Gabor				
	PCA-LDA				
	DCT-LDA				
	Filtro de Gabor- LDA				
	DCT-LDA	1-NN	Regra da Média	ORL (40 classes)	98%
	DCT	Parzen Window e Subspace			
DCT-LDA					
(GAO e WANG, 2006)	Padrão Binário Local (LBP)	Boosting em Subspaços Aleatórios	Regra da Soma	FERET (1.196 clas- ses)	99%
(HUANG, YUEN <i>et al.</i> , 2006)	Direct-LDA	Classificador Baseado em distância euclidiana e Kernel Gaussiano	Regra da Soma Ponderada	FERET (120 classes)	83,21%
	Subspace-LDA			YaleB (10 classes)	89,92%
	Kernel Direct- LDA			CMU PIE	79,42%

	Kernel Subspace-LDA			(68 classes)	
(KIM, 2006)	-	K-NN sobre médias geradas por K-médias	Votação por Maioria	Yale (15 classes)	93,03%
	Espaço de Dissimilaridades usando Distância Euclidiana	KNN			
		1-NN para a média de cada classe			
	-	K-NN sobre médias geradas por K-médias		ORL (40 classes)	99,25%
Espaço de Dissimilaridades usando Distância Regional (RD)	KNN				
	1-NN para a média de cada classe				
(KONG, LI <i>et al.</i> , 2006)	Ensemble LDA sobre Subspaços Aleatórios PCA	1-NN sobre centro de cada classe	Votação por Maioria	ORL (40 classes)	100%
				YaleB (10 classes)	96%
(ZUO, WANG <i>et al.</i> , 2007)	pFisherfaces	1-NN	Regra da Soma para Distâncias	ORL (40 classes)	97,65%
	BDPCA+LDA			FERET (200 classes)	90,55%

(JARILLO, PEDRYCZ <i>et al.</i> , 2008)	Fisher-Isomap sobre 3 imagens: 1) imagem original, 2) imagem com realce de contraste e 3) imagem com realce de contraste e histograma equalizado	1-NN	Regra do Produto	FERET (600 classes)	93,44%
	Kernel-PCA com Kernel Polinomial sobre imagem original, imagem com realce de contraste e imagem com realce de contraste e histograma equalizado			Yale (15 classes)	96,44%
Técnicas Baseadas em Componentes					
(MIRHOSSEINI, YAN <i>et al.</i> , 1998)	Área dos Olhos, Nariz e Boca	Medida de Correlação baseada na norma L_1	Integral de Choquet	ORL (40 classes)	85%

(IVANOV, HEISELE <i>et al.</i> , 2004)	10 componentes faciais	SVM com Kernel Polinomial de grau 2	Regra da Soma, Regra do Produto, Votação por Maioria ou estas 3 aplicadas sobre um Framework Bayesiano	Próprio (6 classes)	90%
(TOYGAR e ACAN, 2004a; b)	PCA de 5 segmentos horizontais da face	1-NN	Regra da Soma	FERET (25 classes)	89%
(MU, WATTA <i>et al.</i> , 2005)	Partições não sobrepostas da Imagem	Template Matching	Votação por Maioria Ponderada	CNNL (1.000 classes)	99%
				FERET (1.196 classes)	96,1%
Técnicas Híbridas					
(LUCAS, 1997b; a; 1998)	n-tuplas contínuas	1-NN	Regra do Mínimo para Distâncias	ORL (40 classes)	97,3%
(CALEANU, 2000)	Operador de Interesse	Redes Neurais MLP	Regra da Média	ORL (10 classes)	97,143%
(HUANG, YUEN <i>et al.</i> , 2002)	PCA	Medida de Distância	Regra da Soma com normalização exponencial-linear	Yale (15 classes)	94,16%
	ICA				
	Spectroface				
	Wavelet de Gabor Local	Medida de Similaridade			
(HUANG, YUEN <i>et al.</i> , 2004)	PCA	Medida de Distância	Regra da Soma Ponderada com normalização DGWNM	Yale (15 classes)	95%
	ICA				
	Spectroface				
	Wavelet de Gabor Local	Medida de Similaridade		ORL (40 classes)	85%
	Spectroface	Medida de		FERET	92,5%

	ICA	Distância		(70 classes)	
	Wavelet de Gabor Local	Medida de Similaridade			
(KHUWAJA, 2002)	6 componentes facias, além da face completa	Redes LVQ	Votação por maioria ou Regra da Soma para Ranking	AR (100 classes)	98,03%
(ZHOU e ZHANG, 2002)	PCA	1-NN	SVM com Kernel Polinomial	Próprio (60 classes)	95,7%
	24 Atributos Geométricos	Distância de mahalanobis			
	Áreas do nariz, Olhos e Boca	Template Matching			
(LEMIEUX e PARIZEAU, 2003)	PCA	1-NN com norma L_1	Borda Count	FERET (1.196 classes)	87%
	PCA de Nariz e Olhos (<i>Eigenobjects</i>)				
	DCT sobre partição da face	HMM			
(RAO, GHADI <i>et al.</i> , 2004)	PCA-LDA sobre face inteira, região dos olhos e mapa de bordas da imagem inteira	1-NN sobre a média das classes	Máxima Probabilidade Conjunta	FERET (1.199 classes)	98,3%
(KWAK e PEDRYCZ, 2005)	PCA-LDA de componentes faciais nariz, boca e região dos olhos além da face inteira	1-NN	Integral de Choquet	FERET (200 classes)	95%
	PCA-LDA sobre 4 subimagens geradas pela decomposição Wavelet de Daubechies			FERET (200 classes)	93,83%

(HEISELE, SERRE <i>et al.</i> , 2007)	14 componentes faciais extraídos por SVM, além da face completa	SVM com Kernel Polinomial do segundo grau	SVM com Kernel Linear com estratégia “um contra todos”	CMU-PIE (10 classes)	89,25%
(SIRLANTZIS, HOQUE <i>et al.</i> , 2008)	Bit-Plane	MWC	Regra da Soma	ORL (40 classes)	97,1%

4 DESENVOLVIMENTO DO PROJETO

O presente trabalho propõe a combinação de classificadores de Máxima Verossimilhança e K Vizinhos Mais Próximos treinados sobre diferentes espaços de atributos como o PCA e algumas variantes, além de LDA, para aplicações em Reconhecimento de Faces.

Neste capítulo, a metodologia utilizada será apresentada e explorada em maiores detalhes na Seção 4.1. Na Seção 4.2, os bancos de dados de faces para teste, assim como os softwares utilizados na implementação computacional do sistema proposto serão identificados. Finalmente, nas Seções 4.3 e 4.4, os métodos de avaliação dos resultados e os resultados obtidos bem como os experimentos realizados são expostos.

4.1 Metodologia

A metodologia proposta consiste em, partindo de um banco de dados de faces humanas, comparar múltiplos classificadores e regras de combinação a fim de encontrar um esquema que atinja o melhor desempenho sobre uma classificação supervisionada. A Figura 4.1 apresenta um diagrama de blocos com o esquema inicial proposto. Basicamente, procurou-se investigar neste trabalho aquele que obtém o melhor desempenho entre o esquema inicial proposto e subesquemas do mesmo.

O esquema inicial considera uma entre nove regras de combinação de classificadores (Regra do Produto, Regra da Soma, Regra do Máximo, Regra do Mínimo, Regra da Mediana, Votação por Maioria, Votação por Maioria Ponderada, Decision Templates e Dempster-Shafer). Elas combinam a saída de dois classificadores: Máxima Verossimilhança e K Vizinhos Mais Próximos (KNN). O KNN é treinado sobre cinco subespaços diferentes: PCA e 4 variações denominadas KL2, KL3, KL4 e KL5. Já o classificador de Máxima Verossimilhança é treinado sobre dois subespaços, sendo denominados aqui como RPCA+RLDA e RBPCA. Os subespaços do classificador de Máxima Verossimilhança são diferentes daqueles definidos para o KNN, pois as matrizes de covariância das classes geradas sobre as variantes do PCA não são possíveis de serem regularizadas, já que autovetores do espaço nulo são selecionados para formar a base do subespaço, o que implica na projeção de amostras de classes diferentes no mesmo ponto. O RPCA+RLDA é extraído utilizando os seguintes passos: 1) PCA; 2) Ge-

ração das Matrizes de Covariância das Classes sobre o espaço PCA e Regularização das mesmas; 3) LDA gerado com as matrizes regularizadas da etapa anterior e 4) Geração das Matrizes de Covariância das Classes sobre o espaço LDA e regularização das mesmas. Por sua vez, o RBPCA é extraído em duas etapas: 1) Block-Based PCA e 2) Geração das Matrizes de Covariância das Classes sobre o espaço do Block-Based PCA e Regularização das Mesmas. Como podemos notar, para o classificador de Máxima Verossimilhança, as Matrizes de Covariância das Classes e suas médias já estarão definidas na Extração dos Atributos. Todas essas técnicas de Extração de Atributos, Regularização de Matrizes de Covariância, Classificadores e Combinação foram explicados no Capítulo 2.

4.2 Materiais e Bancos de Dados Utilizados

O desenvolvimento do trabalho contemplou a implementação dos classificadores e das regras de combinação (com seus respectivos algoritmos de treinamento e teste), além dos métodos de extração de atributos e técnica de Regularização de Matriz de Covariância da solução proposta.

Para isso, o software Matlab que permite o desenvolvimento rápido de protótipos e um pacote de Reconhecimento de Padrões denominado PRTOOLS (DUIN, JUSZCZAK *et al.*, 2004), que foi desenvolvido pela Universidade de Delft e contém diversos algoritmos de reconhecimento de padrões já codificados, foram utilizados para a implementação da solução proposta. Com isso, ao final do processo, os resultados práticos foram obtidos e analisados.

Para executar os experimentos foi utilizado um computador pessoal com processador Intel Core 2 Quad Q6600 2,4 GHz, 8 GB de memória RAM, 2 TB de disco rígido e Sistema Operacional Ubuntu Linux 9.04 versão de 64 bits.

Para a realização dos treinamentos e testes, foram utilizados os bancos de dados de imagens de face humana ORL e AR. Desta forma, permitiu-se a comparação com outros métodos já propostos anteriormente na literatura.

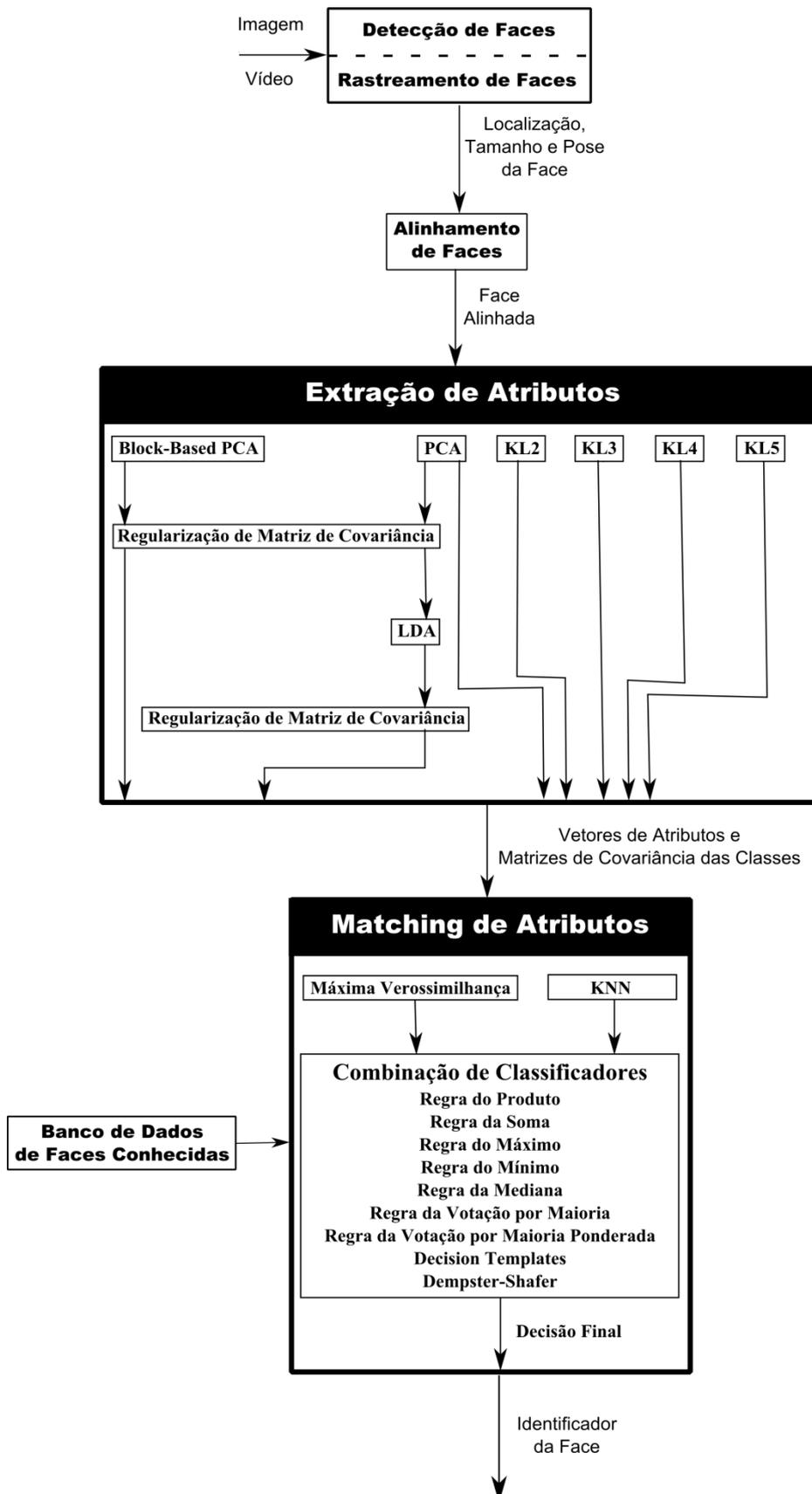


Figura 4.1. Diagrama de blocos do sistema inicial proposto.

4.3 Avaliação dos Resultados

A fim de se avaliar os resultados obtidos na classificação supervisionada e realizar uma comparação objetiva entre os diversos esquemas foram calculadas taxas de classificação correta e coeficientes Kappa. Além disso, para particionar os dados, os métodos Resubstitution e Holdout foram utilizados. Para este último, no caso do banco de dados ORL, diferentes partições do banco de dados foram considerados e não somente metade para treinamento e metade para teste.

Por fim, testes de significância Z e T foram realizados para medir se a superioridade de um esquema sobre outro ou sobre algum outro método é significativa.

4.4 Experimentos e Resultados

Nesta seção, os experimentos realizados são relatados, assim como os seus resultados obtidos. Estes experimentos têm a finalidade de demonstrar a viabilidade do método, avaliar o seu desempenho utilizando os mais diversos esquemas e subesquemas do proposto inicialmente e compará-los a alguns dos métodos mais conhecidos da literatura de reconhecimento de faces (*Eigenfaces* e *Fisherfaces*). Nas próximas subseções, os experimentos executados sobre as bases de dados de faces ORL e AR são descritos e analisados. Finalmente, na subseção 4.4.3, o custo computacional dos diversos extratores de atributos é apresentado.

4.4.1 ORL

As estimativas de erro do banco de dados ORL foram realizadas por *Resubstitution* e *Holdout*. Para o primeiro, todas as amostras foram utilizadas para treinamento e teste. Enquanto que para o segundo, o conjunto de dados foi dividido em subconjuntos disjuntos de treinamento e teste. Como havia 10 amostras por classe no conjunto todo, foram considerados sete experimentos cujos conjuntos de treinamento variaram de 3 a 9 amostras por classe e os de teste de 7 a 1 amostras por classe. Vale ressaltar que foram considerados os de melhor desempenho aqueles que tiveram a maior média de Taxa de Acerto e Kappa.

Para este banco de dados foram utilizadas as imagens originais com 112 x 92 pixels, sendo que, exclusivamente, para o *Block-Based PCA* (BPCA) as imagens foram redimensio-

nadas para 112 x 88 pixels. E ainda para este método de extração de atributos, foram utilizados blocos de tamanho 8 x 8 pixels, obtendo uma imagem reduzida de 14 x 11 pixels, ou seja, foi reduzido de 10304 para 154 atributos no BPCA. Para PCA e suas variantes e LDA, 39 atributos foram extraídos, isto é, $(C - 1)$ atributos, onde C é o número de classes do ORL.

Nas Tabelas 4.1 a 4.4, coeficientes Kappa utilizando *Holdout*, respectivamente, para os classificadores individuais, para o esquema inicial proposto alternando o combinador utilizado referente à Figura 4.1, para os melhores classificadores combinados para cada combinador e uma comparação entre o melhor classificador individual (RBPCA/MaxVer), as técnicas de *Eigenfaces* (PCA/NN) e *Fisherfaces* (PCA-LDA/NN) e os dois melhores subesquemas de combinação obtidos para esse banco de dados que foram RBPCA/MaxVer-PCA/NN-KL5/NN combinados por Votação por Maioria e RBPCA/MaxVer-PCA/NN-KL4/NN-KL5/NN combinados por Soma Ponderada. Considerando esta mesma divisão dos resultados, foram também calculadas as variâncias do Kappa mostradas nas Tabelas 4.5 a 4.8. Nota-se que todas as variâncias dos coeficientes Kappa são praticamente nulas. Além dos coeficientes Kappa e suas variâncias, as taxas de acerto também foram calculadas. Entretanto, por apresentarem informação redundante ao Kappa, elas não foram adicionadas neste documento.

Vale ressaltar que as células realçadas nas tabelas (fundo preto), representam os melhores resultados considerando cada divisão no conjunto de dados e posteriormente, os melhores classificadores ou combinações indicados pela média das linhas da tabela. As médias das colunas de coeficientes Kappa indicam que geralmente, ao aumentar o número de amostras no conjunto de treinamento, obtém-se maior desempenho. Entretanto, geralmente, também ocorre o aumento nas variâncias dos coeficientes Kappa, como apresentado pelas médias das colunas de variâncias de Kappa. Isto é devido ao número de amostras de testes que é reduzido na medida em que o número de amostras no conjunto de treinamento aumenta, fazendo com que haja menos amostras para a estimação do erro.

Pode-se notar que o esquema que obteve o melhor desempenho usava três classificadores: RBPCA/MaxVer, PCA/NN e KL5/NN e Votação por Maioria para combiná-los, obtendo Kappa médio igual 0,9563 e variância 0,0003, que é considerado um excelente desempenho segundo o critério de (LANDIS e KOCH, 1977). O segundo melhor, combinava além dos três classificadores do melhor, também KL4/NN por Regra da Soma Ponderada. Como o banco de dados ORL possui, principalmente, variação de pose, podemos considerar que o método proposto é invariante às mudanças nesta condição.

Para avaliar se este resultado é significativo, a Estatística T considerando 95% de confiança foi utilizada, comparando o melhor esquema de combinação com as técnicas de Eigenfaces, Fisherfaces, o melhor classificador individual e o segundo melhor esquema de combinação. A Tabela 4.9 apresenta estes valores de T que devem ser menor que -1,943 para que a diferença de desempenho seja considerada significativa. Observa-se que o desempenho foi significativamente superior à *Eigenfaces* e ao melhor classificador individual.

Ainda, para uma melhor avaliação dos métodos propostos, Resubstitution foi utilizado. Para as mesmas técnicas relatadas nas Tabelas 4.1 a 4.4 referentes ao *Holdout*, foi obtida uma Taxa de Acerto de 100%, Kappa igual a 1 e variância dos Coeficientes Kappa igual a 0. Isso ocorre, pois o *Resubstitution* com número pequeno de amostras é viesado de maneira otimista como apresentado na Seção 2.4.1.

Tabela 4.1. Coeficientes Kappa dos classificadores individuais sobre o banco de dados ORL usando Holdout com diferentes conjuntos de treinamento e teste.

Métodos	Número de Amostras por indivíduo no conjunto de Treinamento							Média
	3	4	5	6	7	8	9	
RBPCA/MaxVer	0,934	0,906	0,944	0,962	0,949	0,949	0,974	0,945
RPCA+RLDA/MaxVer	0,876	0,855	0,903	0,897	0,932	0,949	0,974	0,912
PCA/NN (<i>Eigenfaces</i>)	0,850	0,880	0,918	0,942	0,966	0,974	0,974	0,929
KL2/NN	0,696	0,739	0,800	0,821	0,915	0,936	0,923	0,833
KL3/NN	0,784	0,671	0,718	0,712	0,658	0,615	0,718	0,697
KL4/NN	0,711	0,714	0,708	0,660	0,624	0,744	0,744	0,700
KL5/NN	0,832	0,872	0,933	0,942	0,940	0,962	0,949	0,918
PCA-LDA/NN (<i>Fisherfaces</i>)	0,832	0,880	0,918	0,923	0,957	0,987	1,000	0,928
Média	0,814	0,815	0,855	0,857	0,868	0,889	0,907	0,858

Tabela 4.2. Coeficientes Kappa do esquema proposto inicialmente com diferentes combinadores sobre o banco de dados ORL usando Holdout com diferentes conjuntos de treinamento e teste.

Combinadores	Número de Amostras por indivíduo no conjunto de							Média
	Treinamento							
	3	4	5	6	7	8	9	
Regra da Soma	0,894	0,902	0,928	0,942	0,957	0,962	0,974	0,937
Regra da Soma Ponderada	0,908	0,910	0,933	0,949	0,966	0,962	0,974	0,943
Regra do Produto	0,897	0,893	0,918	0,936	0,957	0,962	0,949	0,930
Regra do Máximo	0,886	0,919	0,928	0,949	0,966	0,949	0,974	0,939
Regra do Mínimo	0,857	0,786	0,831	0,756	0,727	0,705	0,769	0,776
Regra da Mediana	0,865	0,855	0,903	0,923	0,957	0,974	0,949	0,918
Votação por Maioria	0,919	0,910	0,939	0,949	0,966	0,974	0,974	0,947
Decision Templates	0,663	0,761	0,826	0,763	0,932	0,859	0,872	0,811
Dempster-Shafer	0,681	0,769	0,836	0,776	0,940	0,885	0,897	0,826
Média	0,841	0,856	0,893	0,882	0,930	0,915	0,926	0,892

Tabela 4.3. Coeficientes Kappa dos melhores subesquemas para cada combinador sobre o banco de dados ORL usando Holdout com diferentes conjuntos de treinamento e teste. Para facilitar a visualização consideramos 1 = RBPCA/MaxVer, 2 = RPCA+RLDA/MaxVer, 3 = PCA/NN, 4 = KL2/NN, 5 = KL3/NN, 6 = KL4/NN e 7 = KL5/NN.

Métodos	Número de Amostras por indivíduo no conjunto de							Média
	Treinamento							
	3	4	5	6	7	8	9	
1 e 3 → Soma	0,919	0,923	0,949	0,962	0,966	0,974	0,974	0,952
1, 3 e 7 → Produto	0,919	0,927	0,954	0,962	0,966	0,974	0,974	0,954
1, 3 e 7 → Votação por Maioria	0,927	0,936	0,964	0,974	0,957	0,962	0,974	0,956
1, 3, 6 e 7 → Soma Ponderada	0,927	0,923	0,954	0,955	0,957	0,974	1,000	0,956
1 e 7 → Mediana	0,938	0,927	0,939	0,955	0,957	0,974	0,974	0,952
1, 3, 4, 6 e 7 → Decision Templates	0,758	0,774	0,821	0,763	0,915	0,923	0,923	0,839
1, 3, 4, 6 e 7 → Dempster-Shafer	0,769	0,774	0,826	0,795	0,923	0,936	0,923	0,849
1 e 4 → Máximo	0,941	0,906	0,944	0,962	0,949	0,949	0,974	0,946
1 e 3 → Mínimo	0,857	0,880	0,918	0,942	0,966	0,974	0,974	0,930
Média	0,884	0,886	0,919	0,919	0,951	0,960	0,966	0,926

Tabela 4.4. Coeficientes Kappa do melhor método individual, das Eigenfaces e Fisherfaces e dos dois melhores esquemas de combinação sobre o banco de dados ORL usando Holdout com diferentes conjuntos de treinamento e teste.

Métodos	Número de Amostras por indivíduo no conjunto de Treinamento							Média
	3	4	5	6	7	8	9	
RBPCA/MaxVer	0,934	0,906	0,944	0,962	0,949	0,949	0,974	0,945
PCA/NN (Eigenfaces)	0,850	0,880	0,918	0,942	0,966	0,974	0,974	0,929
PCA-LDA/NN (Fisherfaces)	0,832	0,880	0,918	0,923	0,957	0,987	1,000	0,928
RBPCA/MaxVer-PCA/NN-KL5/NN → Votação por Maioria	0,927	0,936	0,964	0,974	0,957	0,962	0,974	0,956
RBPCA/MaxVer-PCA/NN-KL4/NN-KL5/NN → Soma Ponderada	0,927	0,923	0,954	0,955	0,957	0,974	1,000	0,956
Média	0,894	0,905	0,939	0,951	0,957	0,969	0,985	0,943

Tabela 4.5. Variâncias dos coeficientes Kappa dos classificadores individuais sobre o banco de dados ORL usando Holdout com diferentes conjuntos de treinamento e teste.

Métodos	Número de Amostras por indivíduo no conjunto de Treinamento							Média
	3	4	5	6	7	8	9	
RBPCA/MaxVer	0,000	0,000	0,000	0,000	0,000	0,001	0,001	0,000
RPCA+RLDA/MaxVer	0,000	0,001	0,001	0,001	0,001	0,001	0,001	0,001
PCA/NN (Eigenfaces)	0,001	0,001	0,000	0,000	0,000	0,000	0,001	0,000
KL2/NN	0,001	0,001	0,001	0,001	0,001	0,001	0,002	0,001
KL3/NN	0,001	0,001	0,001	0,001	0,002	0,003	0,005	0,002
KL4/NN	0,001	0,001	0,001	0,001	0,002	0,003	0,005	0,002
KL5/NN	0,001	0,001	0,000	0,000	0,001	0,001	0,001	0,001
PCA-LDA/NN (Fisherfaces)	0,001	0,001	0,000	0,001	0,000	0,000	0,000	0,000
Média	0,001	0,001	0,001	0,001	0,001	0,001	0,002	0,001

Tabela 4.6. Variâncias dos coeficientes Kappa do esquema proposto inicialmente com diferentes combinadores sobre o banco de dados ORL usando Holdout com diferentes conjuntos de treinamento e teste.

Combinadores	Número de Amostras por indivíduo no conjunto de							Média
	Treinamento							
	3	4	5	6	7	8	9	
Regra da Soma	0,000	0,000	0,000	0,000	0,000	0,001	0,001	0,000
Regra da Soma Ponderada	0,000	0,000	0,000	0,000	0,000	0,001	0,001	0,000
Regra do Produto	0,000	0,000	0,000	0,000	0,000	0,001	0,001	0,001
Regra do Máximo	0,000	0,000	0,000	0,000	0,000	0,001	0,001	0,000
Regra do Mínimo	0,000	0,001	0,001	0,001	0,002	0,003	0,005	0,002
Regra da Mediana	0,000	0,001	0,001	0,001	0,000	0,000	0,001	0,001
Votação por Maioria	0,000	0,000	0,000	0,000	0,000	0,000	0,001	0,000
Decision Templates	0,001	0,001	0,001	0,001	0,001	0,002	0,003	0,001
Dempster-Shafer	0,001	0,001	0,001	0,001	0,001	0,001	0,002	0,001
Média	0,000	0,001	0,000	0,001	0,001	0,001	0,002	0,001

Tabela 4.7. Variâncias dos coeficientes Kappa dos melhores subesquemas para cada combinador sobre o banco de dados ORL usando Holdout com diferentes conjuntos de treinamento e teste. Para facilitar a visualização consideramos 1 = RBPCA/MaxVer, 2 = RPCA+RLDA/MaxVer, 3 = PCA/NN, 4 = KL2/NN, 5 = KL3/NN, 6 = KL4/NN e 7 = KL5/NN.

Métodos	Número de Amostras por indivíduo no conjunto de							Média
	Treinamento							
	3	4	5	6	7	8	9	
1 e 3 → Soma	0,000	0,000	0,000	0,000	0,000	0,000	0,001	0,000
1, 3 e 7 → Produto	0,000	0,000	0,000	0,000	0,000	0,000	0,001	0,000
1, 3 e 7 → Votação por Maioria	0,000	0,000	0,000	0,000	0,000	0,001	0,001	0,000
1, 3, 6 e 7 → Soma Ponderada	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000
1 e 7 → Mediana	0,000	0,000	0,000	0,000	0,000	0,000	0,001	0,000
1, 3, 4, 6 e 7 → Decision Templates	0,001	0,001	0,001	0,001	0,001	0,001	0,002	0,001
1, 3, 4, 6 e 7 → Dempster-Shafer	0,001	0,001	0,001	0,001	0,001	0,001	0,002	0,001
1 e 4 → Máximo	0,000	0,000	0,000	0,000	0,000	0,001	0,001	0,000
1 e 3 → Mínimo	0,000	0,001	0,000	0,000	0,000	0,000	0,001	0,000
Média	0,000	0,000	0,000	0,000	0,000	0,000	0,001	0,000

Tabela 4.8. Variâncias dos coeficientes Kappa do melhor método individual, das Eigenfaces e Fisherfaces e dos dois melhores esquemas de combinação sobre o banco de dados ORL usando Holdout com diferentes conjuntos de treinamento e teste.

Métodos	Número de Amostras por indivíduo no conjunto de Treinamento							Média
	3	4	5	6	7	8	9	
RBPCA/MaxVer	0,000	0,000	0,000	0,000	0,000	0,001	0,001	0,000
PCA/NN (Eigenfaces)	0,000	0,001	0,001	0,001	0,001	0,001	0,001	0,001
PCA-LDA/NN (Fisherfaces)	0,001	0,001	0,000	0,000	0,000	0,000	0,001	0,000
RBPCA/MaxVer-PCA/NN-KL5/NN → Votação por Maioria	0,001	0,001	0,001	0,001	0,001	0,001	0,002	0,001
RBPCA/MaxVer-PCA/NN-KL4/NN-KL5/NN → Soma Ponderada	0,001	0,001	0,001	0,001	0,002	0,003	0,005	0,002
Média	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000

Tabela 4.9. Teste de Significância usando Estatística T sobre o banco de dados ORL.

Métodos	RBPCA/MaxVer-PCA/NN-KL5/NN → Votação por Maioria (Estatística T)	Diferença Significativa?
RBPCA/MaxVer	-2,3599	Sim
PCA/NN (Eigenfaces)	-2,0546	Sim
PCA-LDA/NN (Fisherfaces)	-1,6189	Não
RBPCA/MaxVer-PCA/NN-KL4/NN-KL5/NN → Soma Ponderada	-0,0944	Não

4.4.2 AR

Para uma melhor avaliação do método proposto neste trabalho, foi utilizada uma outra base de dados de faces denominada AR. Esta base é maior que a ORL por conter mais classes e mais amostras por classe. Além disso, sua organização permite uma avaliação mais adequada de robustez a iluminação, oclusão e expressão, por exemplo. Entretanto, um conjunto reduzido de número de classes foi utilizado neste trabalho. Este conjunto contém 100 classes (50 homens e 50 mulheres) com 26 amostras por classe (13 por sessão). E suas imagens foram recortadas para conter somente a face e redimensionadas para 165 x 120 pixels (considerada a

imagem original a partir de agora). Para maiores detalhes sobre o processo de aquisição destas imagens, consultar (MARTINEZ e KAK, 2001).

Devido ao alto custo computacional, as imagens foram redimensionadas para 82 x 60 pixels. Para PCA e suas variantes e LDA também foram extraídos (C-1) atributos, onde C é o número de classes. Ou seja, foi reduzido de 4920 para 99 atributos. Já para o *Block-based PCA* foram utilizados blocos de tamanho 5 x 5 sobre imagens redimensionadas para 80 x 60 pixels, obtendo uma imagem reduzida para 16 x 12 pixels. Isto é, de 4920 para 192 atributos.

Basicamente, foi utilizado o mesmo procedimento para avaliação de desempenho usado para o ORL, ou seja, foi verificada a fusão de várias combinações de classificadores e para estimar a probabilidade de erro, *Resubstitution* e *Holdout* também foram utilizados. Todavia, para o *Holdout*, o conjunto de dados foi dividido ao meio com 13 amostras de treinamento por classe (primeira sessão) e 13 amostras de teste por classe (segunda sessão). Para o *Resubstitution*, as 26 amostras de cada classe foram utilizadas tanto para teste quanto para treinamento.

Nas duas próximas subseções, resultados gerais sobre o banco de dados AR e avaliação de robustez à iluminação, oclusão e expressão são relatados.

4.4.2.1 Experimento 1 – Resultados Gerais

Neste experimento foram determinados os coeficientes Kappa e suas variâncias dos classificadores individuais, do uso dos diversos combinadores para o esquema proposto inicialmente, os melhores subesquemas para cada combinador e os melhores resultados comparados a *Eigenfaces* e *Fisherfaces*. Os resultados usando Holdout e Resubstitution são exibidos nas Tabelas 4.10 a 4.13. E ainda, os coeficientes Kappa e Taxa de Acerto apresentaram informação redundante. Desta forma, os gráficos com as taxas de acerto não foram plotados neste documento.

Pode-se notar que o esquema que obteve o melhor desempenho usava três classificadores: RLDA_RPCA/MaxVer, KL4/NN e KL5/NN e Regra da Soma para combiná-los, obtendo Kappa igual 0,8392 e variância 0,0001, que é considerado um excelente desempenho segundo o critério de (LANDIS e KOCH, 1977). O segundo melhor combinava os mesmos classificadores por Regra da Soma Ponderada. Para avaliar se este resultado é significativo, Estatística Z considerando 95% de confiança foi utilizada comparando o melhor esquema de

combinação com as técnicas de Eigenfaces, Fisherfaces, o melhor classificador individual e o segundo melhor esquema de combinação. A Tabela 4.14 apresenta estes valores de Z que devem ser maior que 1,96 para que a diferença de desempenho seja considerada significativa. Observa-se que o desempenho só não foi significativamente superior ao segundo melhor esquema de combinação.

Além disso, como os resultados apresentados utilizando os combinadores Decision Templates e Dempster-Shafer para Resubstitution não atingiram Kappa igual 1, foi verificado se algum outro subconjunto de classificadores atinge tal valor de Kappa. Foi encontrado que diversos alcançaram este desempenho, sendo que consideramos aqui os que utilizaram menor número de classificadores e que serviram para ambas as técnicas. Estes classificadores são RBPCA/MaxVer-KL3/NN ou KL3/NN-KL5/NN. Outro ponto importante com relação a Decision Templates e a Dempster-Shafer é que estes apresentaram resultado muito abaixo dos demais combinadores quando foi utilizado o Holdout. Isto pode ser justificado pela proximidade dos templates de decisão de diversas classes, já que o Decision Templates pode ser visto como o KNN aplicado a um espaço intermediário.

Tabela 4.10. Resultados gerais dos classificadores individuais usando Holdout e Resubstitution sobre o banco de dados AR, onde η é a Taxa de Acerto e σ^2 é a variância do Kappa.

Métodos	Holdout			Resubstitution		
	η	KHAT	σ^2	η	KHAT	σ^2
RBPCA/MaxVer	0,6515	0,6480	0,0002	0,9900	0,9899	0,0000
RPCA+RLDA/MaxVer	0,7915	0,7894	0,0001	1,0000	1,0000	0,0000
PCA/NN (Eigenfaces)	0,5831	0,5789	0,0002	1,0000	1,0000	0,0000
KL2/NN	0,5600	0,5556	0,0002	1,0000	1,0000	0,0000
KL3/NN	0,4415	0,4359	0,0002	1,0000	1,0000	0,0000
KL4/NN	0,4454	0,4398	0,0002	1,0000	1,0000	0,0000
KL5/NN	0,8100	0,8081	0,0001	1,0000	1,0000	0,0000
PCA-LDA/NN (Fisherfaces)	0,6946	0,6915	0,0002	1,0000	1,0000	0,0000
Média	0,6222	0,6184	0,0002	0,9988	0,9987	0,0000

Tabela 4.11. Resultados gerais do esquema proposto inicialmente com diferentes combinadores usando Holdout e Resubstitution sobre o banco de dados AR, onde η é a Taxa de Acerto e σ^2 é a variância do Kappa.

Combinadores	Holdout			Resubstitution		
	η	KHAT	σ^2	η	KHAT	σ^2
Regra da Soma	0,7677	0,7653	0,0001	1,0000	1,0000	0,0000
Regra da Soma Ponderada	0,7877	0,7855	0,0001	1,0000	1,0000	0,0000
Regra do Produto	0,7685	0,7661	0,0001	1,0000	1,0000	0,0000
Regra do Máximo	0,6515	0,6480	0,0002	1,0000	1,0000	0,0000
Regra do Mínimo	0,6938	0,6908	0,0002	1,0000	1,0000	0,0000
Regra da Mediana	0,5777	0,5734	0,0002	1,0000	1,0000	0,0000
Votação por Maioria	0,7862	0,7840	0,0001	1,0000	1,0000	0,0000
Decision Templates	0,1292	0,1204	0,0001	0,9773	0,9771	0,0000
Dempster-Shafer	0,2246	0,2168	0,0001	0,9969	0,9969	0,0000
Média	0,5985	0,5945	0,0001	0,9971	0,9971	0,0000

Tabela 4.12. Resultados gerais dos melhores subesquemas para cada combinador usando Holdout e Resubstitution sobre o banco de dados AR, onde η é a Taxa de Acerto e σ^2 é a variância do Kappa. Para facilitar a visualização consideramos 1 = RBPCA/MaxVer, 2 = RPCA+RLDA/MaxVer, 3 = PCA/NN, 4 = KL2/NN, 5 = KL3/NN, 6 = KL4/NN e 7 = KL5/NN.

Métodos	Holdout			Resubstitution		
	η	KHAT	σ^2	η	KHAT	σ^2
2, 6 e 7 → Soma	0,8408	0,8392	0,0001	1,0000	1,0000	0,0000
2, 6 e 7 → Soma Ponderada	0,8385	0,8368	0,0001	1,0000	1,0000	0,0000
2 e 7 → Produto	0,8338	0,8322	0,0001	1,0000	1,0000	0,0000
2 e 7 → Máximo	0,7915	0,7894	0,0001	1,0000	1,0000	0,0000
2 e 7 → Mínimo	0,8100	0,8081	0,0001	1,0000	1,0000	0,0000
2 e 7 → Mediana	0,8331	0,8314	0,0001	1,0000	1,0000	0,0000
2, 4, 5, 6 e 7 → Votação por Maioria	0,8092	0,8073	0,0001	1,0000	1,0000	0,0000
1 e 7 → Mínimo	0,8100	0,8081	0,0001	1,0000	1,0000	0,0000
1 e 7 → Decision Templates	0,4062	0,4002	0,0002	0,9992	0,9992	0,0000
1 e 7 → Dempster-Shafer	0,4054	0,3994	0,0002	0,9992	0,9992	0,0000
2 e 3 → Máximo	0,7915	0,7894	0,0001	1,0000	1,0000	0,0000
2 e 4 → Máximo	0,7915	0,7894	0,0001	1,0000	1,0000	0,0000
2 e 5 → Máximo	0,7915	0,7894	0,0001	1,0000	1,0000	0,0000
2 e 6 → Máximo	0,7915	0,7894	0,0001	1,0000	1,0000	0,0000
Média	0,7532	0,7507	0,0001	0,9999	0,9999	0,0000

Tabela 4.13. Resultados gerais do melhor método individual, das Eigenfaces e Fisherfaces e dos dois melhores esquemas de combinação usando Holdout e Resubstitution sobre o banco de dados AR, onde η é a Taxa de Acerto e σ^2 é a variância do Kappa.

Métodos	Holdout			Resubstitution		
	η	KHAT	σ^2	η	KHAT	σ^2
PCA/NN (<i>Eigenfaces</i>)	0,5831	0,5789	0,0002	1,0000	1,0000	0,0000
KL5/NN	0,8100	0,8081	0,0001	1,0000	1,0000	0,0000
PCA-LDA/NN (<i>Fisherfaces</i>)	0,6946	0,6915	0,0002	1,0000	1,0000	0,0000
RLDA+RPCA/MaxVer-KL4/NN-KL5/NN → Soma	0,8408	0,8392	0,0001	1,0000	1,0000	0,0000
RLDA+RPCA/MaxVer-KL4/NN-KL5/NN → Soma Ponderada	0,8385	0,8368	0,0001	1,0000	1,0000	0,0000
Média	0,7534	0,7509	0,0001	1,0000	1,0000	1,0000

Tabela 4.14. Teste de Significância usando Estatística Z sobre o banco de dados AR.

Métodos	RLDA+RPCA/MaxVer-KL4/NN-KL5/NN → Soma (Estatística Z)	Diferença Significativa?
KL5/NN	2,1991	Sim
PCA/NN (<i>Eigenfaces</i>)	15,0284	Sim
PCA-LDA/NN (<i>Fisherfaces</i>)	8,5275	Sim
RLDA+RPCA/MaxVer-KL4/NN-KL5/NN → Soma Ponderada	0,1697	Não

4.4.2.2 Experimento 2 – Avaliação de Robustez à Iluminação, Oclusão e Expressão

Como o banco de dados AR possui uma divisão clara das variações de face, ela permite uma análise adequada de robustez a alguns dos desafios de sistemas de reconhecimento de faces. Nas Figuras 4.2 e 4.3 têm-se os resultados de Taxa de Acerto usando Holdout para cada variação na Face. Para uma maior clareza o eixo das abscissas representa as variações das faces sendo, 1 = Expressão Neutra, 2 = Sorrindo, 3 = Zangado, 4 = Gritando, 5 = Iluminação do Lado Esquerdo, 6 = Iluminação do Lado Direito, 7 = Iluminação de Ambos os Lados, 8 = Usando Óculos de Sol, 9 = Usando Óculos de Sol e Iluminação do Lado Esquerdo, 10 = Usando Óculos de Sol e Iluminação do Lado Direito, 11 = Usando Lenço, 12 = Usando Lenço e Iluminação do Lado Esquerdo, 13 = Usando Lenço e Iluminação do Lado Direito.

Podemos observar que, tomando a condição de face sem variações (1 = Expressão Neutra) como a taxa mínima para considerar um método robusto contra alguma variação, o método proposto apresenta robustez à iluminação e a pequenas mudanças de expressão. Além disso, analisando a Figura 4.2, é possível perceber alguma semelhança no comportamento dos diversos classificadores individuais mesmo com taxas de acerto diferentes, o que pode significar uma forte correlação entre estes classificadores.

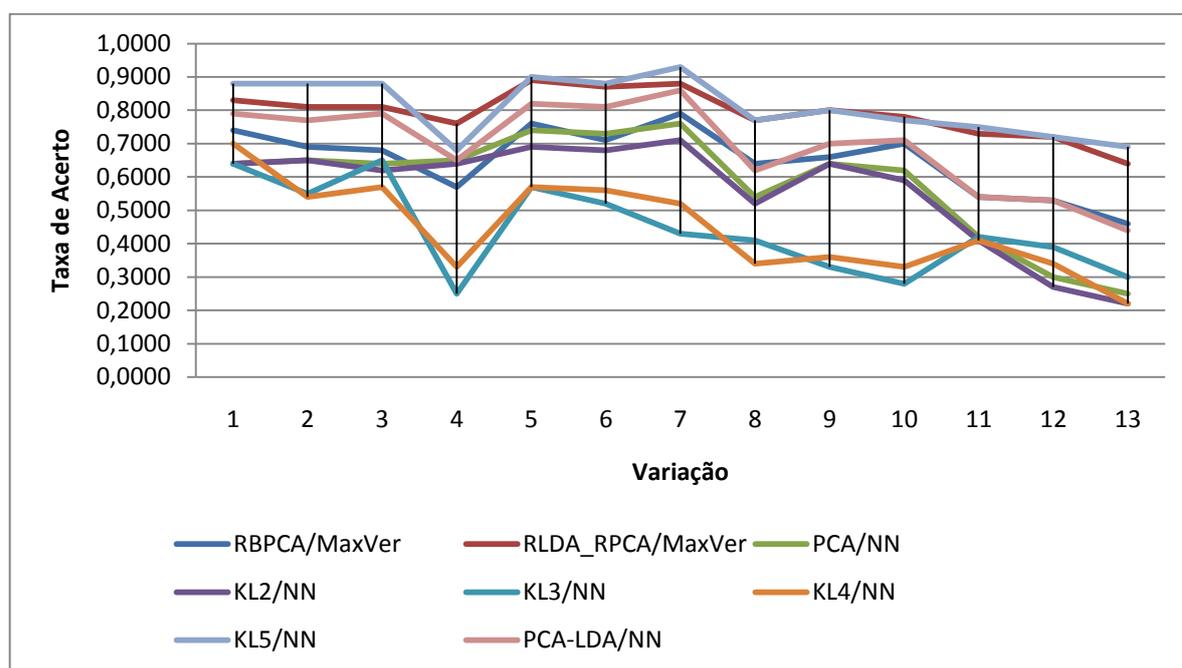


Figura 4.2. Taxas de Acerto por variação na face dos classificadores individuais usando Holdout sobre o banco de dados AR.

4.4.3 Custo Computacional

A fim de demonstrar a viabilidade, os tempos aproximados de cada extrator de atributos na etapa de treinamento para os bancos de dados AR e ORL são apresentados na Tabela 4.15 abaixo. Os tempos de teste e treinamento de combinadores treináveis não foram considerados aqui devido ao custo principal estar na etapa de treinamento.

Pode-se verificar da Tabela 4.15 que o KL4 é o método mais demorado. Nota-se que o custo do RBPCA e PCA se alternou para cada banco de dados como sendo os mais rápidos. Basicamente, temos que o RBPCA depende especialmente do número de classes e dimensão final obtida devido à etapa de regularização das matrizes de covariância das classes. Enquanto

que o PCA, assim como as suas variantes, dependem da dimensionalidade dos dados de entrada já que o passo mais demorado e que consome praticamente todo o tempo de treinamento destes extratores de atributos é a obtenção dos autovalores e autovetores da matriz de covariância utilizada. O RLDA_RPCA, por sua vez, depende de ambos os fatores do PCA e do RBPCA.

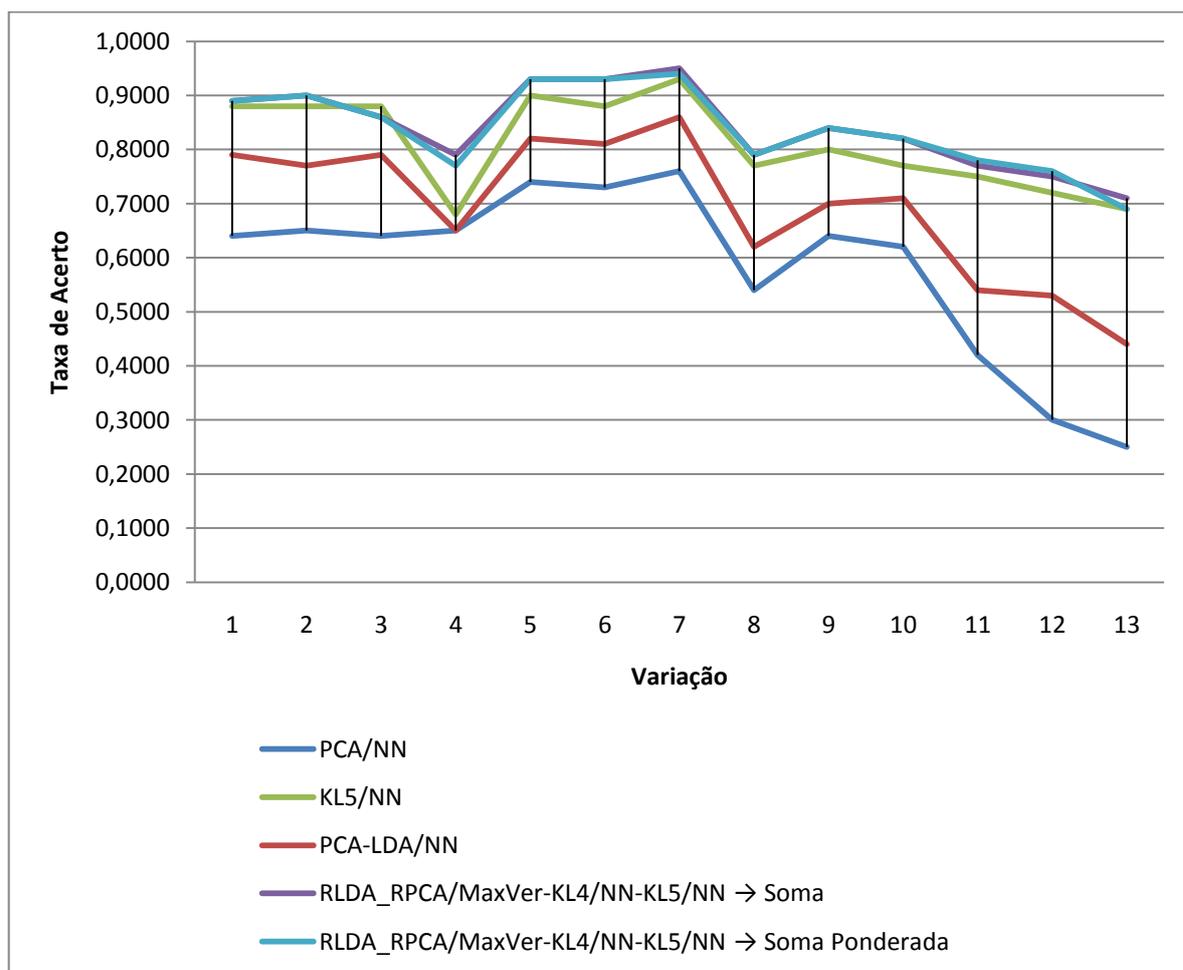


Figura 4.3. Taxas de Acerto por variação na face do melhor método individual, das Eigenfaces e Fisherfaces e dos dois melhores esquemas de combinação usando Holdout sobre o banco de dados AR

Tabela 4.15. Tempos aproximados de cada extrator de atributos na etapa de treinamento sobre os bancos de dados de faces AR e ORL.

Métodos	ORL	AR
RBPCA	4 minutos	30 minutos
RLDA_RPCA	2 horas e 35 minutos	30 minutos
PCA	2 horas e 34 minutos	8 minutos
KL2	2 horas e 57 minutos	33 minutos
KL3	3 horas e 23 minutos	36 minutos
KL4	4 horas e 11 minutos	1 hora e 50 minutos
KL5	3 horas e 38 minutos	43 minutos

4.5 Considerações finais

Neste capítulo o sistema proposto para reconhecimento de faces foi apresentado, os experimentos realizados foram descritos e os diversos resultados obtidos utilizando os bancos de dados de faces ORL e AR foram relatados e discutidos.

No próximo capítulo, algumas conclusões obtidas neste trabalho serão expostas, bem como sugestões para possíveis extensões e continuidade deste trabalho.

5 CONCLUSÕES E TRABALHOS FUTUROS

Neste capítulo, as conclusões deste trabalho são fornecidas na Seção 5.1 e encerrando, trabalhos futuros para a continuidade deste projeto são esboçadas na Seção 5.2.

5.1 Conclusões

Neste trabalho foi analisada a combinação de múltiplos classificadores para reconhecimento de faces. Para isso foram utilizados classificadores de Máxima Verossimilhança e KNN treinados sobre espaços definidos por PCA e algumas variantes, Block-Based PCA e LDA. Devido ao problema da Singularidade que decorre do mal condicionamento do problema foi também utilizada uma técnica de regularização de matrizes de covariância das classes. Um esquema inicial foi proposto utilizando 7 classificadores combinados por 1 de 9 combinadores (7 fixos e 2 treináveis). A idéia inicial era verificar a existência de um subesquema preponderante considerando dois bancos de dados de faces, ORL e AR. Não houve um esquema que atingisse tal desempenho para ambos os bancos de dados. Entretanto, foram obtidas altas taxas de acerto, altos coeficientes Kappa e pequenas variâncias se comparado a alguns classificadores individuais e a dois métodos bem conhecidos em reconhecimento de faces (*Eigenfaces* e *Fisherfaces*) usando Holdout e Resubstitution como métodos de partição dos conjuntos de dados e estimativas de erro. Além disso, em geral, os melhores resultados de combinação apresentaram diferença significativa nesta comparação como pôde ser determinado calculando as Estatísticas Z e T. Ainda, pelos experimentos, verificou-se que os melhores subesquemas são robustos à pose (no caso do ORL) e à iluminação e a pequenas variações em expressão facial (no caso do AR), mas foi identificado que as técnicas sofreram influência de oclusões. Vale ressaltar que o sistema de reconhecimento de faces é um problema extremamente mal condicionado. Enfim, podemos concluir que combinar múltiplos classificadores variando também as técnicas de extração de atributos, tem um grande potencial de melhorar o desempenho de sistemas de reconhecimento de faces.

5.2 Trabalhos Futuros

Durante o desenvolvimento deste trabalho foram identificados alguns pontos que necessitam de uma investigação mais aprofundada e servem como continuação deste trabalho.

Estes pontos são:

- Utilizar o Block-based PCA (BPCA) como etapa inicial de extração, ou seja, antes de outras técnicas de extração de atributos tal como PCA e suas variantes e LDA;
- Considerar linhas e colunas das imagens como blocos para o BPCA e possivelmente, treinar classificadores sobre BPCA extraído de cada tipo (bloco quadrado, linhas e colunas) ou ainda fazer uma combinação em nível de atributos;
- Comparar o desempenho obtido por Block-based PCA com técnicas de interpolação;
- Aplicar algum pré-processamento nas imagens a fim de torná-lo mais invariante à iluminação;
- Estudar os espaços gerados por Decision Templates e Dempster-Shafer com maiores detalhes;
- Avaliar o desempenho sobre outras bases de dados como CMU-PIE e FRAV2D (disponível em <http://www.fray.es/databases/FRAV2d/>), que são bancos de dados com maior número de classes e mais variações a serem consideradas.

REFERÊNCIAS

- ABATE, A. F. et al. 2D and 3D face recognition: A survey. **Pattern Recognition Letters**, v. 28, n. 14, p. 1885-1906, Oct 15 2007.
- ACHERMANN, B.; BUNKE, H. **Combination of Classifiers on the Decision Level for Face Recognition**. Universität Bern. Bern. 1996a. (IAM-96-002)
- _____. **Combination of face classifiers for person identification**. In: Proceedings of the 13th International Conference on Pattern Recognition, 1996b. p. 416-420.
- ARBUCKLE, T. D. et al. Fuzzy Information Fusion in a Face Recognition System. **International Journal of Uncertainty Fuzziness and Knowledge-Based Systems**, v. 3, n. 3, p. 217-246, Sep 1995.
- BARTLETT, M. S.; MOVELLAN, J. R.; SEJNOWSKI, T. J. Face recognition by independent component analysis. **IEEE Transactions on Neural Networks**, v. 13, n. 6, p. 1450-1464, Nov 2002.
- BELHUMEUR, P. N.; HESPANHA, J. P.; KRIEGMAN, D. J. Eigenfaces vs. Fisherfaces: Recognition using class specific linear projection. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, v. 19, n. 7, p. 711-720, Jul 1997.
- CALEANU, C. D. **Facial recognition using committee of neural networks**. In: Proceedings of the 5th Seminar on Neural Network Applications in Electrical Engineering (NEUREL'2000), 2000. p. 97-100.
- CHAWLA, N. V.; BOWYER, K. W. Designing multiple classifier systems for face recognition. **Multiple Classifier Systems**, v. 3541, p. 407-416, 2005.
- CHELLAPPA, R.; WILSON, C. L.; SIROHEY, S. Human and Machine Recognition of Faces: A Survey. **Proceedings of the IEEE**, v. 83, n. 5, p. 705-740, May 1995.
- CHIEN, Y.; KING-SUN, F. On the generalized Karhunen-Loève expansion. **IEEE Transactions on Information Theory**, v. 13, n. 3, p. 518-520, 1967.
- CHOQUET, G. Theory of Capacities. **Annales de l'Institut Fourier**, v. 5, p. 131-295, 1953.
- COHEN, J. A Coefficient of Agreement for Nominal Scales. **Educational and Psychological Measurement**, v. 20, n. 1, p. 37-46, 1960.
- CONGALTON, R. G., GREEN, KASS. **Assessing the Accuracy of Remotely Sensed data: Principles and Practices**. Lewis Publishers, 1999.
- DEVIJVER, P. A.; KITTLER, J. **Pattern Recognition: A Statistical Approach**. Englewood Cliffs, New Jersey: Prentice-Hall, 1982.
- DUDA, R.; HART, P.; STORK, D. **Pattern Classification**. 2. ed. Wiley-Interscience, 2000.

DUIN, R. P. W. et al. **PRTools 4: A Matlab Toolbox for Pattern Recognition**. Delft University of Technology: Delft, 2004. 39 p.

EBRAHIMPOUR, R.; EHTERAM, S. R.; KABIR, E. Face Recognition by Multiple Classifiers, a Divide-and-Conquer Approach. In: KHOSLA, R.; HOWLETT, R. J., *et al* (Eds.). **Knowledge-Based Intelligent Information and Engineering Systems**: Springer, 2005. p. 225-232. (Lecture Notes in Computer Science).

FUKUNAGA, K. **Introduction to statistical pattern recognition**. 2. ed. Academic Press Professional, Inc., 1990. 592 p.

GAO, Y.; WANG, Y. **Boosting in Random Subspaces for Face Recognition**. In: 18th International Conference on Pattern Recognition (ICPR'2006), 2006. p. 519-522.

GEORGHIADES, A. S.; BELHUMEUR, P. N.; KRIEGMAN, D. J. From Few to Many: Illumination Cone Models for Face Recognition under Variable Lighting and Pose. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, v. 23, n. 6, p. 643-660, 2001.

GROSS, R. Face Databases. In: LI, S. Z. e JAIN, A. K. (Eds.). **Handbook of Face Recognition**. New York: Springer-Verlag, 2005. Cap. 13, p. 301-328.

HEISELE, B.; SERRE, T.; POGGIO, T. A component-based framework for face detection and identification. **International Journal of Computer Vision**, v. 74, n. 2, p. 167-181, Aug 2007.

HUANG, J. et al. Linear and nonlinear feature-based fusion algorithms for face recognition. **International Journal of Wavelets Multiresolution and Information Processing**, v. 4, n. 4, p. 659-676, Dec 2006.

_____. **Combining local and global features for face recognition**. In: Proceedings of the Fifth Asian Conference on Computer Vision (ACCV'02), 2002. Melbourne. p. 568-573.

_____. Face recognition using local and global features. **EURASIP Journal on Applied Signal Processing**, v. 2004, n. 4, p. 530-541, Apr 1 2004.

IVANOV, Y.; HEISELE, B.; SERRE, T. **Using component features for face recognition**. In: Proceedings of the Sixth IEEE International Conference on Automatic Face and Gesture Recognition (FG'2004), 2004. p. 421-426.

JAIN, A. K.; DUIN, R. P. W.; MAO, J. C. Statistical pattern recognition: A review. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, v. 22, n. 1, p. 4-37, Jan 2000.

JAIN, A. K.; LI, S. Z. **Handbook of Face Recognition**. Springer-Verlag New York, Inc., 2005.

JARILLO, G.; PEDRYCZ, W.; REFORMAT, M. Aggregation of classifiers based on image transformations in biometric face recognition. **Machine Vision and Applications**, v. 19, n. 2, p. 125-140, Mar 2008.

JENG, S. H. et al. Facial feature detection using geometrical face model: An efficient approach. **Pattern Recognition**, v. 31, n. 3, p. 273-282, Mar 1998.

JING, X. Y.; ZHANG, D. Face recognition based on linear classifiers combination. **Neuro-computing**, v. 50, p. 485-488, Jan 2003.

KHAN, M. A. U. et al. **Cross correlation measure for decision fusion among multiple face classifiers**. In: Proceedings of the IEEE Symposium on Emerging Technologies, 2005. p. 126-131.

KHUWAJA, G. A. An adaptive combined classifier system for invariant face recognition. **Digital Signal Processing**, v. 12, n. 1, p. 21-46, Jan 2002.

KIM, S. W. On using a dissimilarity representation method to solve the small sample size problem for face recognition. **Proceedings of the Advanced Concepts for Intelligent Vision Systems**, v. 4179, p. 1174-1185, 2006.

KIM, T. K.; KITTLER, J. Combining Classifier for Face Identification at Unknown Views with a Single Model Image. In: FRED, A. L. N.; CAELLI, T., *et al* (Eds.). **Structural, Syntactic, and Statistical Pattern Recognition**: Springer, v. 3138, 2004. p. 565-573. (Lecture Notes in Computer Science).

_____. Design and fusion of pose-invariant face-identification experts. **IEEE Transactions on Circuits and Systems for Video Technology**, v. 16, n. 9, p. 1096-1106, Sep 2006.

KIRBY, M.; SIROVICH, L. Application of the Karhunen-Loeve Procedure for the Characterization of Human Faces. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, v. 12, n. 1, p. 103-108, Jan 1990.

KITTLER, J. et al. On combining classifiers. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, v. 20, n. 3, p. 226-239, Mar 1998.

KITTLER, J.; YOUNG, P. C. A new approach to feature selection based on the Karhunen-Loeve expansion. **Pattern Recognition**, v. 5, n. 4, p. 335-352, 1973.

KONG, H. et al. Ensemble LDA for face recognition. **Proceedings of the Advances in Biometrics**, v. 3832, p. 166-172, 2006.

KUNCHEVA, L. **Combining Pattern Classifiers: Methods and Algorithms**. Wiley-Interscience, 2004.

KUO, B. C.; LANDGREBE, D. A. A covariance estimator for small sample size classification problems and its application to feature extraction. **IEEE Transactions on Geoscience and Remote Sensing**, v. 40, n. 4, p. 814-819, Apr 2002.

KWAK, K.-C.; PEDRYCZ, W. Face recognition: A study in information fusion using fuzzy integral. **Pattern Recognition Letters**, v. 26, n. 6, p. 719-733, 2005.

LANDIS, J. R.; KOCH, G. G. The measurement of observer agreement for categorical data. **Biometrics**, v. 33, n. 1, p. 159-174, 1977.

LEMIEUX, A.; PARIZEAU, M. Flexible multi-classifier architecture for face recognition systems. **Vision Interface**, 2003.

LI, S. Z.; JAIN, A. K. Introduction. In: LI, S. Z. e JAIN, A. K. (Eds.). **Handbook of Face Recognition**. New York: Springer-Verlag, 2005. Cap. 1, p. 1-12.

LI, X.; MORI, G.; ZHANG, H. **Expression-Invariant Face Recognition with Expression Classification**. In: The 3rd Canadian Conference on Computer and Robot Vision, 2006. p. 77-77.

LIU, C.; SHUM, H. Y.; FREEMAN, W. T. Face hallucination: Theory and practice. **International Journal of Computer Vision**, v. 75, n. 1, p. 115-134, Oct 2007.

LU, X. **Image Analysis for Face Recognition**. NOTES, P. 2003.

LU, X.; WANG, Y.; JAIN, A. K. **Combining classifiers for face recognition**. In: Proceedings of the International Conference on Multimedia and Expo (ICME '03), 2003. p. 13-16.

LUCAS, S. M. Continuous n-tuple classifier and its application to face recognition. **Electronics Letters**, v. 33, n. 20, p. 1676-1678, Sep 25 1997a.

_____. **Face recognition with the continuous n-tuple classifier**. In: Proceedings of the British Machine Vision Conference, 1997b. p. 222-231.

_____. Continuous n-tuple classifier and its application to real-time face recognition. **IEEE Proceedings-Vision Image and Signal Processing**, v. 145, n. 5, p. 343-348, Oct 1998.

LUMINI, A.; NANNI, L. Combining classifiers to obtain a reliable method for face recognition. **Multimedia Cyberscape Journal**, v. 3, n. 3, p. 47-53, June 2005.

MARCIALIS, G. L.; ROLI, F. **Fusion of LDA and PCA for Face Recognition**. In: Proceedings of the Workshop on Machine Vision and Perception, 8 th Workshop of the Italian Association for Artificial Intelligence (AIIA'02), 2002. p. 1-5.

_____. Fusion of face recognition algorithms for video-based surveillance systems. In: FOREST, G. L.; REGAZZONI, C., *et al* (Eds.). **Multisensor Surveillance Systems: The Fusion Perspective**. Boston: Kluwer Academic Publishers, 2003. p. 235-250.

_____. Fusion of appearance-based face recognition algorithms. **Pattern Analysis and Applications**, v. 7, n. 2, p. 151-163, Jul 2004.

MARTÍNEZ, A.; BENAVENTE, R. **The AR Face Database**. CVC/Universitat Autònoma de Barcelona. Barcelona. 1998. (24)

MARTINEZ, A. M.; KAK, A. C. PCA versus LDA. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, v. 23, n. 2, p. 228-233, Feb 2001.

MESSER, K. et al. **XM2VTSDB: The Extended M2VTS Database**. In: Proceedings of the Second International Conference on Audio- and Video-based Biometric Person Authentication (AVBPA'99), 1999. 1999. p.

MIRHOSSEINI, A. R. et al. Human face image recognition: An evidence aggregation approach. **Computer Vision and Image Understanding**, v. 71, n. 2, p. 213-230, Aug 1998.

MONTGOMERY, D. C.; RUNGER, G. C. **Applied Statistics and Probability for Engineers**. 3. ed. John Wiley & Sons, 2002.

MU, X.; WATTA, P.; HASSOUN, M. H. **Combining local similarity measures: summing, voting, and weighted voting**. In: IEEE International Conference on Systems, Man and Cybernetics (SMC'2005), 2005. p. 2661-2666.

PHILLIPS, P. J. **Support vector machines applied to face recognition**. In: Advances in Neural Information Processing Systems 11, 1999. MIT Press. p. 803-809.

PHILLIPS, P. J. et al. The FERET evaluation methodology for face-recognition algorithms. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, v. 22, n. 10, p. 1090-1104, Oct 2000.

_____. **FRVT 2006 and ICE 2006 Large-Scale Results**. NIST. Gaithersburg, p. 55. 2007

RAO, K. S. et al. **Robust Face Recognition by Fusion**. Fourth Indian Conference on Computer Vision, Graphics and Image Processing (ICVGIP'2004). CHANDA, B.; CHANDRAN, S., et al. Kolkata: Allied Publishers Private Limited: p. 669-674. 2004.

RYU, Y. S.; OH, S. Y. Hybrid classification for face recognition with virtual samples and ensemble neural networks. **Intelligent Automation and Soft Computing**, v. 8, n. 4, p. 303-313, 2002.

SAMAL, A.; IYENGAR, P. A. Automatic Recognition and Analysis of Human Faces and Facial Expressions - a Survey. **Pattern Recognition**, v. 25, n. 1, p. 65-77, Jan 1992.

SIRLANTZIS, K.; HOQUE, S.; FAIRHURST, M. C. Diversity in multiple classifier ensembles based on binary feature quantisation with application to face recognition. **Applied Soft Computing**, v. 8, n. 1, p. 437-445, Jan 2008.

SUGENO, M. **Theory of fuzzy integrals and its applications**. 1974. Thesis (Ph. D.). Tokyo Institute of Technology, Tokyo.

THEODORIDIS, S.; KOUTROUMBAS, K. **Pattern Recognition**. 2. ed. Academic Press, 2003.

TOLBA, A. S.; ABU-REZQ, A. N. Combined Classifiers for Invariant Face Recognition. **Pattern Analysis & Applications**, v. 3, n. 4, p. 289-302, 2000.

TOYGAR, O.; ACAN, A. Boosting face recognition speed with a novel divide-and-conquer approach. **Proceedings of the Computer and Information Sciences (ISCIS'2004)**, v. 3280, p. 430-439, 2004a.

_____. Multiple classifier implementation of a divide-and-conquer approach using appearance-based statistical methods for face recognition. **Pattern Recognition Letters**, v. 25, n. 12, p. 1421-1430, Sep 2004b.

TURK, M.; PENTLAND, A. Eigenfaces for Recognition. **Journal of Cognitive Neuroscience**, v. 3, n. 1, p. 71-86, Win 1991.

VAPNIK, V. N. **The nature of statistical learning theory**. Springer-Verlag New York, Inc., 1995. 188 p.

WANG, X. G.; TANG, X. Face hallucination and recognition. **Proceedings of the Audio- and Video-Based Biometric Person Authentication (AVBPA'2003)**, v. 2688, p. 486-494, 2003.

WEBB, A. **Statistical Pattern Recognition**. 2. ed. John Wiley & Sons, 2002.

XU, L.; KRZYZAK, A.; SUEN, C. Y. Methods of Combining Multiple Classifiers and Their Applications to Handwriting Recognition. **IEEE Transactions on Systems Man and Cybernetics**, v. 22, n. 3, p. 418-435, May-Jun 1992.

ZHANG, W. et al. **Information fusion in face identification**. In: Proceedings of the 17th International Conference on Pattern Recognition (ICPR'2004), 2004. p. 950-953.

ZHAO, W. et al. Face recognition: A literature survey. **ACM Computing Surveys**, v. 35, n. 4, p. 399-459, Dec 2003.

ZHAO, Z.-Q.; HUANG, D.-S.; SUN, B.-Y. Human face recognition based on multi-features using neural networks committee. **Pattern Recognition Letters**, v. 25, n. 12, p. 1351-1358, 2004a.

_____. **Human facial recognition based on multiple feature domains**. In: Proceedings of the Fifth World Congress on Intelligent Control and Automation (WCICA'2004), 2004b. p. 4150-4155.

ZHOU, J.; ZHANG, D. **Face recognition by combining several algorithms**. In: Proceedings of the 16th International Conference on Pattern Recognition (ICPR'2002). , 2002. p. 497-500.

ZUO, W. M. et al. Combination of two novel LDA-based methods for face recognition. **Neurocomputing**, v. 70, n. 4-6, p. 735-742, Jan 2007.