

Anderson Rodrigo dos Santos

IDENTIFICAÇÃO DE FACES HUMANAS ATRAVÉS DE PCA-LDA E
REDES NEURAIS SOM

Dissertação apresentada à Escola de Engenharia de São Carlos da Universidade de São Paulo, como parte dos requisitos para a obtenção do Título de Mestre em Engenharia Elétrica.

Orientador: Professor Dr. Adilson Gonzaga

São Carlos

2005

Livros Grátis

<http://www.livrosgratis.com.br>

Milhares de livros grátis para download.

Este trabalho é dedicado a duas pessoas importantes e fundamentais na minha trajetória, sem as quais eu não teria alcançado êxito: aos meus pais Ana e Alceu por me incentivar a sonhar com objetivos grandes, e por me motivar a busca pela perfeição com críticas e ensinamentos sábios.

Agradecimentos

A vida é um laboratório constante para as descobertas individuais e interações sociais. Cada dia é um novo dia para aprender coisas novas, para enfrentar novos desafios, para superar medos antigos, para se conhecer melhor, para se reinventar ... para ser feliz.

E para ser feliz o homem deve ser capaz de sonhar com alguma coisa, deve ser capaz de realizá-la, ou seja, deve ser a mudança que deseja ver. Somente o desafio provoca o crescimento. Somente o desafio irá testar nossas habilidades e nos tornar melhores. Somente o desafio e a automotivação para encararmos o desafio podem nos transformar. E cada desafio que aceitamos cria um "eu" mais habilidoso.

Tudo o que aprendemos a fazer, somente aprendemos quando realmente fazemos. Embora uma dissertação seja pela sua finalidade acadêmica um trabalho individual, há contribuições de natureza diversa que não podem nem devem deixar de ser realçadas. Por essa razão, é difícil criar uma lista contendo todas as pessoas que me apoiaram e contribuíram direta ou indiretamente para o desenvolvimento dessa dissertação. No momento, eu gostaria de ressaltar alguns colaboradores importantes e me desculpar por todos os nomes que eu omiti neste momento, mas que estão presentes de alguma forma nas entrelinhas desse projeto de mestrado.

Início meus agradecimentos citando o Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) pelo apoio financeiro (processo 130120/ 2003-0). Ao meu orientador e amigo, Prof. Dr. Adilson Gonzaga, pelo constante incentivo sempre indicando à direção a ser tomada nos momentos de maior dificuldade, apoiando as minhas decisões nos momentos de afloração intelectual, e por despertar em mim o interesse pela

área de Visão Computacional. Agradeço, principalmente, pela confiança depositada no meu trabalho de mestrado.

Agradeço aos professores: Evandro L. L. Rodrigues, Ivan N. Silva, Maria S. V. Paiva e Roberto M. Cesar Jr. pela contribuição na correção da dissertação, e por apresentarem idéias e críticas que enriqueceram o trabalho.

Agradeço as secretárias de pós-graduação do departamento de Engenharia Elétrica, Denise e Marisa, pela agilidade nos processos burocráticos relacionados ao bom andamento de um trabalho de mestrado.

Agradeço aos meus familiares - pai, mãe e irmão, por me apoiarem e incentivarem diariamente. A minha atenciosa amiga Gisele por me oferecer equilíbrio e entender a minha ausência nos momentos especiais de nossas vidas.

Agradeço aos meus amigos - Aline, Celsinho, Celso, Deise, Eduardo, Juliana, Junior, Lilian, Nilvana, Paulo, Rocha, Rubinho, Tarallo, Thiago, Zé (técnico) - e colegas do laboratório LAVI por sua ajuda ora silenciosa, ora com bons conselhos e incentivo. Ao departamento de Engenharia Elétrica - USP - São Carlos, onde fiz o curso de graduação, por me creditar confiança para a realização desta pesquisa. E ainda a todos que fizeram parte da minha formação pessoal, que de forma intencional ou não me emprestaram parte de seu conhecimento, de sua visão de mundo, que me foram e ainda serão necessários neste e em desafios vindouros.

Obrigado a todos.

"A Vida sem um objetivo é igual a um navio sem um capitão. Para todo objetivo traçado existem inúmeros problemas a serem resolvidos, mas os grandes problemas e desafios nos parecem grandes porque estamos de joelhos diante deles. Tornar o simples complicado é fácil; tornar o complicado simples, inacreditavelmente simples, isso é criatividade e persistência.

A Vida é um jogo. Fingir que não vai morrer é prejudicial para a forma como se leva a vida. Sem fim, não existe jogo. Se não tiver consciência da morte, não é possível ter consciência da vida. Aquele que não está ocupado tentando nascer, está ocupado tentando morrer. Mostre-me um homem que tem medo de desafios, e eu lhe mostrarei um homem que pode ser derrotado o tempo todo. A Vida é um reflexo da perfeição e do poder Divino. Faça de cada dia uma obra de arte."

Anderson R. Santos

Resumo

SANTOS, A. R. (2005). Identificação de Faces Humanas Através de Pca-Lda e Redes Neurais SOM. Dissertação (Mestrado) - Escola de Engenharia de São Carlos, Universidade de São Paulo, São Carlos, 2005.

O uso de dados biométricos da face para verificação automática de identidade é um dos maiores desafios em sistemas de controle de acesso seguro. O processo é extremamente complexo e influenciado por muitos fatores relacionados à forma, posição, iluminação, rotação, translação, disfarce e oclusão de características faciais.

Hoje existem muitas técnicas para se reconhecer uma face. Esse trabalho apresenta uma investigação buscando identificar uma face no banco de dados ORL com diferentes grupos de treinamento. É proposto um algoritmo para o reconhecimento de faces baseado na técnica de subespaço LDA (PCA + LDA) utilizando uma rede neural SOM para representar cada classe (face) na etapa de classificação/identificação.

Aplicando o método do subespaço LDA busca-se extrair as características mais importantes na identificação das faces previamente conhecidas e presentes no banco de dados, criando um espaço dimensional menor e discriminante com relação ao espaço original. As redes SOM são responsáveis pela memorização das características de cada classe. O algoritmo oferece maior desempenho (taxas de reconhecimento entre 97% e 98%) com relação às adversidades e fontes de erros que prejudicam os métodos de reconhecimento de faces tradicionais.

Palavras-chave: Reconhecimento de Faces; Análise de Componentes Principais (PCA); Autofaces; Análise Discriminante Linear (LDA); Subespaço LDA; Rede Neural SOM; Rede Neural de Kohonen; Rede Neural PCA Adaptativo.

Abstract

SANTOS, A. R. (2005). Identification of Human Faces Based on PCA - LDA and SOM Neural Networks. M.Sc. Dissertation - Escola de Engenharia de São Carlos, Universidade de São Paulo, São Carlos, 2005.

The use of biometric technique for automatic personal identification is one of the biggest challenges in the security field. The process is complex because it is influenced by many factors related to the form, position, illumination, rotation, translation, disguise and occlusion of face characteristics.

Now a days, there are many face recognition techniques. This work presents a methodology for searching a face in the ORL database with some different training sets . The algorithm for face recognition was based on sub-space LDA (PCA + LDA) technique using a SOM neural net to represent each class (face) in the stage of classification/identification.

By applying the sub-space LDA method, we extract the most important characteristics in the identification of previously known faces that belong to the database, creating a reduced and more discriminated dimensional space than the original space. The SOM nets are responsible for the memorization of each class characteristic. The algorithm offers great performance (recognition rates between 97% and 98%) considering the adversities and sources of errors inherent to the traditional methods of face recognition.

Keywords: Face Recognition; Principal Component Analysis (PCA); Eigenface; Linear Discriminant Analysis (LDA); LDA Sub-space; SOM Neural Networks; Kohonen Neural Networks; PCA Neural Networks.

Lista de Figuras

2.1	Sistema genérico de reconhecimento de padrões.	14
2.2	Biometria facial.	19
2.3	a) Características geométricas (em branco) utilizadas no reconhecimento de faces, b) Diferentes regiões utilizadas na estratégia <i>Template Matching</i> (BRUNELLI; POGGIO, 1993).	20
2.4	Imagens de Faces utilizadas para formar o banco de dados (Sirovich; Kirby, 1987)	25
2.5	Erro percentual versus número de autovetores (Sirovich; Kirby, 1987). curva contínua - face completa curva pontilhada - face recortada.	26
2.6	Primeiros nove autovetores, na ordem da esquerda para direita e de cima para baixo (KIRBY; SIROVICH, 1990).	27
2.7	Sistema de localização e <i>tracking</i> da cabeça (TURK; PENTLAND, 1991a).	29
2.8	Taxa de Reconhecimento para os métodos autofaces, autocaracterística (olhos, nariz e boca) e representação modular combinada (PENTLAND et al., 1994).	31
2.9	Reta indicando a posição de maior separabilidade dos conjuntos.	33
2.10	Separabilidade entre os dois eixos reais.	34
2.11	Reta ótima para a separabilidade dos conjuntos.	35
2.12	Comparação entre PCA e LDA para um problema de duas classe, onde os dados de cada classe são projetados em um subespaço linear.	38
2.13	Sistema Sub-espaço LDA generalizado para o reconhecimento de faces.	40

2.14	Imagens de uma pessoa no <i>AR face database</i> . As imagens de (a)-(m) foram tiradas durante uma seção e as imagens de (n)-(z) em uma seção diferente.	41
2.15	Experiência - a) duas imagens de treinamento e cinco para teste b) 13 imagens de treinamento e 13 para teste. Fonte: Martinez e Kak (2001). . .	42
2.16	Imagens eletronicamente modificadas que foram corretamente identificadas (MARTINEZ; KAK, 2001).	43
2.17	Rede de auto-associação e classificativa para reconhecimento de faces. Fonte: Cottrell e Fleming (1990).	44
2.18	Descrição do processo de amostragem local da imagem. Uma janela é deslizada passo-a-passo sobre a imagem e um vetor é criado em cada passo referente a uma localização	45
2.19	Diagrama geral do sistema de reconhecimento de faces.	46
2.20	Método para extração das características de expressão facial (KATOH; FUKUI, 1998).	48
3.1	<i>Cluster</i> de Face no Espaço de Imagem (cor de fundo preto).	53
3.2	a) Exemplo de um conjunto de dados correlacionados, b) Dados normalizados e seus autovetores	55
3.3	Caracterização do vetor imagem.	57
3.4	Autovetores - exemplo de bases do sub-espço de autofaces.	58
3.5	Exemplo de face formada pela combinação linear da base autofaces. . . .	60
3.6	Versão simplificada do espaço de faces ilustrando as quatro possibilidades para a projeção de uma imagem no espaço de faces. Neste caso, há dois autovetores (μ_1 e μ_2) e três indivíduos conhecidos (Ω_1 , Ω_2 e Ω_3).	63
3.7	Teste com o PCA efetuando recuperação de informação.	63
4.1	Neurônios do sistema nervoso central dos vertebrados: (a) Neurônio motor da célula espinhal; (b) Célula de Purkinje encontrada no cerebelo. Esta célula é notável pela extensa ramificação da sua árvore dendrital, da qual apenas uma pequena parte é mostrada na figura, podendo ultrapassar várias dezenas de vezes as dimensões do corpo celular (HAYKIN, 1998).	66
4.2	Modelo matemático não-linear de um neurônio.	66
4.3	Diagrama de blocos representativos da aprendizagem supervisionada (HAYKIN, 1998).	68
4.4	Diagrama de blocos da aprendizagem por reforço (HAYKIN, 1998). . . .	69
4.5	Diagrama de blocos da aprendizagem não-supervisionado (HAYKIN, 1998).	70

4.6	Aprendizagem: Erro em relação ao conjunto de aprendizado e de teste.	71
4.7	RNA de uma única unidade, a ser treinada pela aprendizagem Hebbiana.	73
4.8	Classificação estrutural e funcional de redes neurais artificiais.	76
4.9	RNA de uma única unidade, a ser treinada pela aprendizagem Hebbiana.	78
4.10	RNA progressiva com uma única camada de neurônios lineares, a ser treinada pela aprendizagem Hebbiana Generalizada.	80
4.11	Rede neural PCA Adaptativa - APEX.	83
4.12	(a) Rede neural PCA proposta por Rubner e Tavan (1989). (b) Rede neural LDA proposta por Mao e Jain (1995)(1993).	87
4.13	Exemplo de estrutura da Rede de Kohonen com neurônios de entrada e saída.	90
4.14	Exemplos de arranjos. (a) arranjo linear; (b) arranjo retangular; (c) arranjo hexagonal.	91
4.15	Neurônio vencedor no processo competitivo.	91
4.16	Atualização dos pesos na vizinhança do neurônio vencedor.	92
4.17	Visualização geométrica da atualização dos pesos na rede de Kohonen.	93
5.1	Diagrama em blocos de um sistema de reconhecimento de faces.	97
5.2	Diagrama em blocos do modelo proposto - fase de treinamento.	100
5.3	Diagrama em blocos do modelo proposto - fase de teste.	101
5.4	Exemplo de faces presentes no banco de faces ORL.	102
5.5	Equalização do Histograma de uma face.	104
5.6	Convergência - gráfica da soma dos pesos laterais da rede (8 faces de treino).	105
5.7	Face média.	105
5.8	Conjunto de autofaces (autovetores) colocados em ordem de 1 a 20.	106
5.9	Autovetores de 311 a 320.	106
5.10	Visualização dos Autovalores.	107
5.11	Composição de uma face através de seus pesos e respectivas autofaces.	107
5.12	Diagrama das possíveis combinações PCA e LDA.	109
5.13	Desempenho PCA e máximo desempenho LDA (5 faces de treino).	110
5.14	Desempenho PCA e máximo desempenho LDA (6 faces de treino).	111
5.15	Desempenho PCA e máximo desempenho LDA (7 faces de treino).	111
5.16	Desempenho PCA e máximo desempenho LDA (8 faces de treino).	112
5.17	Desempenho PCA e máximo desempenho LDA (9 faces de treino).	112
5.18	Desempenho PCA e máximo desempenho LDA (10 faces de treino).	113
5.19	Desempenho LDA (5 faces de treino).	115

5.20	Desempenho LDA (6 faces de treino).	115
5.21	Desempenho LDA (7 faces de treino).	116
5.22	Desempenho LDA (8 faces de treino).	116
5.23	Desempenho LDA (9 faces de treino).	117
5.24	Desempenho LDA (10 faces de treino).	117
5.25	Desempenho das redes SOM (8 faces de treino e 4 neurônios).	119
5.26	Desempenho das redes SOM (8 faces de treino e 8 neurônios).	120
5.27	Desempenho das redes SOM (8 faces de treino e 12 neurônios).	120
5.28	Desempenho das redes SOM (8 faces de treino e 3x3 neurônios).	121
5.29	Desempenho das redes SOM (8 faces de treino e 4x4 neurônios).	122
5.30	Evolução da energia acumulada (5 faces de treino).	123
5.31	Evolução da energia acumulada (6 faces de treino).	123
5.32	Evolução da energia acumulada (7 faces de treino).	124
5.33	Evolução da energia acumulada (8 faces de treino).	124
5.34	Evolução da energia acumulada (9 faces de treino).	124
5.35	Evolução da energia acumulada (10 faces de treino).	125
5.36	Desempenho PCA, energia acumulada e desempenho LDA (8 faces de treino).	125
5.37	Pessoa com a maior quantidade de erros no reconhecimento.	129
5.38	Pessoa com erro zero no reconhecimento.	129
5.39	Atualização dos neurônios na rede SOM.	130
5.40	Faces modificadas.	131
5.41	Faces similares recuperadas para a imagem (a).	131
5.42	Faces similares recuperadas para a imagem (b).	131
5.43	Faces similares recuperadas para a imagem (c).	131
5.44	Faces similares recuperadas para a imagem (d).	131
5.45	Faces similares recuperadas para a imagem (e).	132
B.1	Pessoa A.	150
B.2	Pessoa B.	151
B.3	Distância entre os 8 neurônios (linha) e as 12 faces de teste (colunas) antes do treinamento.	152
B.4	Distância entre os 8 neurônios (linha) e as 12 faces de teste (colunas) após o treinamento.	152
B.5	Pessoa B com as imagens mais claras.	153

B.6	Distância entre os 8 neurônios (linha) e as 12 faces de teste (colunas) antes do treinamento.	153
B.7	Distância entre os 8 neurônios (linha) e as 12 faces de teste (colunas) após o treinamento.	154
B.8	Modificação da região de identificação para o neurônio vencedor com a face de teste X.	154

Lista de Tabelas

1.1	Aplicações das tecnologias de reconhecimento de faces	3
1.2	Sistemas de reconhecimento de face (ZHAO et al., 2003).	4
2.1	Técnicas de reconhecimento de faces (fotografias) desenvolvidas nos últimos anos.	18
2.2	Desempenho geral das técnicas utilizadas.	47
2.3	Resultado da pesquisa (Katoh; Fukui, 1998).	47
5.1	Taxas de reconhecimento	127
5.2	Erros na identificação	128

Lista de Siglas

APEX	Adaptive Principal Components Extraction (português: Extrator Adaptativo de Componentes Principais)
BPNN	Back Propagation Neural Network
CNN	Convolution Neural Network (português: Rede Neural Convolutacional)
CSOM	Concurrent Self Organize Map (português: Mapa Auto-organizável Concorrente)
DCT	Discrete Cosine Transform (português: Transformada Discreta Cosseno)
DRT	Discrete Randon Transform
DWT	Discrete Wavelet Transform (português: Transformada Discreta Wavelet)
Eigenfaces	(português: Autofaces)
Eigenvalues	(português: Autovalores)
Eigenvectors	(português: Autovetores)
FFT	Fast Fourier Transform (português: Transformada Rápida de Fourier)
FT	Fourier Transform (português: Transformada de Fourier)
FLD	Fisher's Linear Discriminant (português: Discriminante Linear de Fishers)
ICA	Independent Component Analysis (português: Análise do Componente Independente)
LDA	Linear Discriminant Analysis (português: Análise de Discriminante Linear)
MATLAB	Matrix Laboratory
MCNN	Multilayer Cluster Neural Networks (português: Rede Neural de Cluster Multicamadas)
MLP	Multilayer Perceptron (português: Perceptron Multicamadas)
ORL	Olivetti Research Laboratory
PDBNN	Probabilistic Decision Based Neural Network (português: Decisão Probabilística Baseada em Rede Neural)
PCA	Principal Component Analysis (português: Análise do Componente Principal)
SOM	Self Organize Map (português: Mapa Auto-organizável)
RBFs	Radial Based Functions (português: Funções de Base Radial)
SVM	Support Vector Machines (português: Máquinas de Vetor de Suporte)
VLSI	Very Large-Scale Integration (português: Alta Escala de Integração)

Sumário

1	Introdução	1
1.1	Considerações Iniciais.	1
1.2	Aplicações.	3
1.3	Visão Natural e o Reconhecimento Automático de Faces Humanas.	4
1.4	Objetivos.	7
1.5	Estrutura do Texto.	8
2	Reconhecimento de Faces Humanas	9
2.1	Considerações Iniciais.	9
2.2	A Capacidade Humana para o Reconhecimento de Faces.	10
2.3	Reconhecimento de Padrões.	14
2.4	Reconhecimento Automático de Faces.	15
2.5	Métodos Baseados em Características.	18
2.6	Métodos Holísticos (e Estatísticos).	20
2.6.1	Análise de Componentes Principais (PCA).	22
2.6.2	Evolução no Reconhecimento de Faces Aplicando PCA.	24
2.6.3	Análise Discriminante Linear (Linear Discriminant Analysis - LDA).	32
2.6.4	Sub-espço LDA.	39
2.7	Redes Neurais Aplicadas ao Reconhecimento de Faces Humanas.	42
2.8	Outros Métodos para o Reconhecimento de Faces.	48
2.8.1	Transformada de Fourier (FT).	48
2.8.2	Transformada Cosseno Discreta (DCT).	49
2.8.3	Imagens de Perfil.	49
2.9	Considerações Finais.	50

3	Análise de Componentes Principais (PCA - Principal Component Analysis)	51
3.1	Considerações Iniciais.	51
3.2	Características das Imagens Faciais.	52
3.3	Identificação de Faces Utilizando PCA.	53
3.4	Calculando as Autofaces.	57
3.5	Projetando as Faces de Treinamento no Sub-espço de Faces.	60
3.6	Classificação e Identificação de uma Nova Imagem.	61
3.7	Considerações Finais.	63
4	Redes Neurais para o Reconhecimento de Padrões	64
4.1	Considerações Iniciais.	64
4.2	Redes Neurais Artificiais (RNA).	64
4.3	Características dos Processos de Aprendizagem.	67
4.3.1	Aprendizagem Supervisionada.	68
4.3.2	Aprendizagem por Reforço.	69
4.3.3	Aprendizagem Não-Supervisionada.	69
4.3.4	Generalização das Redes Neurais.	70
4.3.5	Aprendizagem Hebbiana.	72
4.3.6	Aprendizagem Competitiva.	74
4.4	Redes Neurais para Reconhecimento de Padrões.	75
4.4.1	Redes Auto-Organizadas.	76
4.4.2	Rede Neural PCA Auto-Organizante.	77
4.4.3	Rede Neural PCA Adaptativa.	82
4.4.4	Rede Neural LDA.	85
4.4.5	Rede Neural de Kohonen.	88
4.5	Considerações Finais.	95
5	Implementação do Algoritmo e Resultados	96
5.1	Considerações Iniciais.	96
5.2	Modelo Proposto.	97
5.3	Banco de Imagens.	102
5.4	Pré-Processamento das Imagens.	103
5.5	Descrição da Fase de Treinamento.	103
5.6	Testes Realizados.	108
5.6.1	Avaliação de Desempenho (PCA e LDA).	108
5.6.2	Avaliação do Desempenho LDA (ou sub-espço LDA).	114

5.6.3	Avaliação do Desempenho da Rede Neural SOM (Kohonen). . . .	118
5.6.4	Definição dos Subespaços PCA e LDA.	122
5.6.5	Aplicação do Algoritmo.	126
5.6.6	Atualização Constante dos Neurônios.	129
5.6.7	Busca por Similaridade.	130
5.7	Considerações Finais.	132
6	Conclusões	134
6.1	Considerações Finais.	134
6.2	Perspectivas Gerais do Trabalho.	136
6.3	Linhas de Pesquisas Futuras.	136
	Referências Bibliográficas	137
A	Apêndice: Algoritmos	147
B	Apêndice: Testes - Atualização Constante dos Neurônios	150

Introdução

1.1 Considerações Iniciais.

A verificação automática da identidade de uma pessoa é um assunto de grande relevância, pois pode ser aplicado tanto em ambientes físicos sociais com conotação privada como em ambientes virtuais, como exemplo, podemos considerar: vigilância, controle de acesso seguro, interfaces inteligentes (para controles de acessos de mercadorias ou informações), comércio eletrônico (World Wide Web), etc... Atualmente, nas redes de computadores espalhadas pelo mundo, está se tornando cada vez mais importante e cada vez mais difícil manter a segurança da informação, ou da propriedade física vinculada às entidades que utilizam os sistemas de informação modernos.

As redes de computadores revolucionaram as telecomunicações, as transferências e armazenamentos de dados, a logística, as operações financeiras e comerciais das empresas, mas apresentaram novos problemas a serem resolvidos com relação à segurança da informação e dos pontos de acesso, e como exemplo pode-se citar as fraudes nos cartões de crédito, as invasões de computadores por *hackers*¹, ou a ruptura no sistema de segurança de edifícios privado ou governamental. No ano de 1998, segundo publicado na

¹As palavras de outra língua serão colocadas em itálico no texto

Reuters (1999), os crimes fraudulentos relacionados aos sistemas de informação somaram \$100 milhões em perdas nos EUA. Na maioria desses crimes, os criminosos aproveitam uma falha fundamental nos sistemas de controle convencionais de acesso: os sistemas não concedem o acesso por "quem nós somos", mas por "o que nós temos", como cartões de identificação, chaves e senhas. Nenhum desses meios nos definem individualmente, pois são meramente meios de autenticação. Lembrando, que se alguém roubar, duplicar, ou adquirir estes meios de identidade poderá alcançar nossos dados ou nossa propriedade pessoal quando desejar. Recentemente, os avanços tecnológicos buscam tornar disponível a verificação da identidade individual "verdadeira". Esta tecnologia é baseada em um campo chamado "biometria"(LIN,2000).

Nos últimos anos, acompanhando a revolução dos computadores foram propostas muitas linhas de projetos para a autenticação/reconhecimento da identidade de uma pessoa no campo da biometria, como reconhecimento de impressão digital, da retina, da voz, geometria de mão, reconhecimento de íris e também análise de imagem facial. Todos estes métodos são caracterizados por limitações intrínsecas e muitos requerem uma cooperação ativa do usuário, e isto têm como consequência uma baixa aceitabilidade social. Em princípio, a análise de imagens de face parece ser o melhor modo para executar autenticação de identidade, e também o mais aceitável pelas pessoas: isto é o que todos os seres humanos fazem todos os dias de suas vidas, pois estão sempre associando as características sociais e pessoais a uma determinada face guardada na memória, que posteriormente será lembrada para o reconhecimento. Por outro lado, surgem muitas dificuldades com relação à enorme dimensionalidade do espaço relacionado ao armazenamento e análise dos dados encontrados em imagens naturais de faces obtidas por câmeras (TISTARELLI; GROSSO, 1999).

No reconhecimento facial pode-se estabelecer duas categorias, que norteiam as aplicações:

- Encontrar uma pessoa dentro de um grande banco de dados de face (ex, banco de dados da polícia). Estes sistemas tipicamente retornam uma lista das pessoas mais parecidas no banco de dados com relação à imagem de entrada. Não é necessário o reconhecimento em tempo real.
- Identificar uma pessoa em tempo real (ex, monitoramento em sistemas de segurança), ou permitir o acesso a um grupo de pessoas, e negar a outros (ex, acesso a edifícios, computadores, etc).

1.2 Aplicações.

Os sistemas de reconhecimento de faces descritos na tabela 1.1 abrangem aplicações em imagens estáticas com ambiente controlado até imagens de vídeo em ambiente não controlado, portanto os desafios a serem superados para efetuar o reconhecimento diferem com relação às aplicações. Desse modo, os sistemas podem ser classificados em dois grupos: estáticos (sem movimento) e dinâmicos (vídeo). Existem significantes diferenças entre as aplicações descritas nos dois grupos; as diferenças são com relação à qualidade das imagens, a segmentação das faces na cena, e a necessidade de reconhecimento em tempo real. Recentes avanços estão sendo feitos no desenvolvimento de novas metodologias de pré-processamento das imagens e detecção das faces para autenticação/reconhecimento de pessoas, reduzindo os efeitos de posição (translação, rotação e escala), sombras ou outras ambigüidades, que podem surgir entre o banco de dados e as imagens capturadas em tempo real (vídeo), ou por fotografias. Comercialmente já existem alguns sistemas sendo vendidos e testados em diversas aplicações (ver tabela 1.2).

Tabela 1.1: Aplicações das tecnologias de reconhecimento de faces

Aplicações	Vantagens	Desvantagens
Cartão de crédito, licença de motorista, Passaporte e Identificação Pessoal	Imagem controlada, segmentação controlada e boa qualidade nas imagens	Não existe banco de dados grande, dificuldade na busca quando o banco de dados aumenta
Retrato falado	Imagem padronizada e mais de uma imagem disponível	Não existe banco de dados grande, dificuldade na busca quando o banco de dados aumenta
Segurança em Bancos/ Lojas	Localização geográfica conhecida	Segmentação não-controlada e baixa qualidade nas imagens
Vigilância em locais públicos	Sistemas de câmeras instalados (Shopping, praças...)	Segmentação não-controlada, baixa qualidade nas imagens e processamento em tempo real
Identificação inteligente	mercado em expansão	Baixa qualidade nas imagens

Na etapa de identificação muitos pesquisadores estão dedicando seus esforços para limitar, ou evitar a complexidade de processamento na autenticação/reconhecimento das imagens associadas às informações contidas no banco de dados dos sistemas de verificação. Diferentes técnicas de reconhecimento foram propostas recentemente, tais como: métodos utilizando redes neurais, *template matching* elásticos, *Support Vector Machine* (SVM), expansão de Karhunen-Loeve (PCA) e extração de características baseada nas

Tabela 1.2: Sistemas de reconhecimento de face (ZHAO et al., 2003).

Produtos Comerciais	Websites
FaceIt from Visionics	http://www.FaceIt.com
Viisage Technology	http://www.viisage.com
FaceVACS from Plettac	http://www.plettac-eletronics.com
FaceKey Corp.	http://www.facekey.com
Cognitec Systems	http://www.cognitec-systems.de
Keyware Technologies	http://www.keywareusa.com
Passfaces from ID-arts	http://www.id-arts.com
ImageWare Software	http://www.iwsinc.com
Eyematic Interfaces Inc.	http://www.eyematic.com
BioID sensor fusion	http://www.bioid.com
Visionshere Technologies	http://www.visionsheretech.com/menu.htm
Biometric Systems, Inc.	http://www.biometrica.com
FaceSnap Recoder	http://www.facesnap.de/htdocs/english/index2.html
SpotIt for face composite	http://spotit.itc.it/SpotIt.html

distâncias dos elementos locais da face. Estas técnicas produzem bons resultados, limitados às restrições impostas pelas condições de aquisição das imagens (iluminação, posição e cena de fundo bem controlada).

1.3 Visão Natural e o Reconhecimento Automático de Faces Humanas.

O Reconhecimento é possivelmente a motivação final para muitos sistemas de visão artificial ou natural. Uma das razões principais para a grande complexidade de tarefas de reconhecimento/ verificação é a quantidade de informação a ser processada. Segundo Tistarelli e Grosso (1999) para realizar qualquer tarefa visual, todos os sistemas de percepção natural são capazes de interagir com o ambiente e adquirir tantas informações quantas forem precisas, propositalmente controlando o fluxo de dados de entrada, mas também limitando a quantidade de informação adquirida do sistema sensorial. A anatomia do sistema visual humano é um exemplo claro: apesar da boa acuidade no centro da fóvea (1 minuto de arco) e o campo largo de visão (aproximadamente 140 a 200 graus de ângulo sólido), o nervo ótico é composto de somente 1 milhão de fibras nervosas. A distribuição de espaço-variável das células na retina permite uma redução no fluxo de dados. Na realidade, a mesma resolução resultaria dentro de um espaço-invariante de

sensores cerca de 600.000.000 pixels (ou um nervo ótico mais largo que o próprio olho). Outro mecanismo perceptual importante relacionado ao processo de aquisição de dados é o mecanismo de atenção. Novamente, como nem todos os dados de contribuição são pertinentes à uma determinada tarefa, o sistema perceptual deve ser capaz de fazer uma seleção do sinal de contribuição em várias dimensões: sinais espaciais, profundidade, movimento, etc... A seleção é controlada por um mecanismo próprio de atenção limitando a faixa de visão ou focalizando processos. O paradigma de visão ativa leva em conta conceber sistemas visuais artificiais que podem executar uma determinada tarefa partindo de suposições gerais.

O controle de acesso e identificação através do reconhecimento de faces é uma tecnologia promissora no campo da identificação pessoal, pois este sistema traz algumas vantagens com relação a outras formas de identificação biométrica. Exemplo:

- Com relação à impressão digital, pesquisas mostram que 5% das pessoas não possuem uma "boa" impressão digital. As razões para isto incluem: pele cortada, dedo com curativo, dedo com calo, pele seca, úmida, doente, dedo estreito, e sensor de leitura manchado ou engordurado pelo toque. A área da impressão digital é muito menor que a face, portanto a face é mais resistente a cortes e contusões.
- O reconhecimento por íris possui um grau de acerto de praticamente 100%, pois é uma característica única de cada ser humano. Mas o problema está relacionado à captura da íris, visto que a câmera deve estar perfeitamente alinhada em relação à íris contando com a cooperação do indivíduo.
- A análise de voz para reconhecimento é um meio natural e pouco importuno, mas o reconhecimento através da face é o mais prático e cômodo, pois permite que o usuário seja identificado simplesmente passando na frente de uma câmera.

Um sistema genérico de reconhecimento de faces deve abordar diversas questões técnicas relativas tanto ao estímulo tratado (a face humana), quanto aos processos envolvidos na tarefa de reconhecimento em si. São várias as etapas associadas ao processo de reconhecimento automático de faces:

1. Detecção: dada uma imagem qualquer, identificar o local onde aparece a face.
2. Segmentação: identificar as partes constituintes de uma face, e separá-las das outras informações presentes na cena.

3. Representação: deve-se selecionar as informações da face que serão utilizadas para representá-la.
4. Identificação: verificar se uma face pertence ao conjunto de faces do banco de dados.

No reconhecimento automático de faces, um conjunto de faces é primeiramente registrado no sistema e codificado para uma futura busca. Uma face de teste é apresentada ao sistema e comparada com o banco de dados; o reconhecimento é efetuado quando a face teste se aproxima de uma face do banco de dados de acordo com um limiar definido. O desafio para os sistemas é executar o reconhecimento da face apesar das transformações: mudança na rotação, translação, escala, iluminação, e mudanças também de expressão, idade, uso de óculos e disfarces, que são inerentes às faces. Assim, a necessidade é encontrar uma codificação apropriada para as imagens de faces (2-D ou 3-D) derivadas de uma mesma pessoa para que se possa efetuar o reconhecimento com eficiência.

A eficiência no reconhecimento está relacionada à tarefa executada pelo sistema nas aplicações. Tomando-se uma base de dados que consiste em um conjunto de treinamento T , de faces de c pessoas conhecidas, sendo Ω o conjunto de todas as classes (ou pessoas) existentes e $\omega_1, \omega_2, \omega_3 \dots \omega_c$ classes de padrões (pessoas). Considerando-se também, que x é um padrão originário de uma face cuja classificação é desconhecida, Gong et al. (2000) definiram no mínimo quatro tarefas relacionadas com a identificação:

1. Classificação: consiste na identificação de uma face x assumindo-se que ela é de uma pessoa do conjunto Ω . Em outras palavras, assumindo-se que x pode ser classificado como um padrão de alguma classe ω_i , tal que $\omega_i \in \Omega$, a tarefa de classificação consiste em determinar o valor de i .
2. Conhecido-desconhecido: objetiva decidir se a face é ou não um membro de Ω , ou seja, se x pode ser classificado como um padrão de alguma classe de Ω .
3. Verificação: dado que a identidade ω_i de uma face x foi determinada através de um outro meio não visual, essa tarefa busca confirmar a identidade dessa pessoa usando imagens de face, ou seja, confirmar se x é da classe ω_i .
4. Reconhecimento Completo: Visa determinar se uma face é de alguma classe de Ω e, em caso positivo, determinar sua identidade ω_i .

1.4 Objetivos.

O objetivo principal deste trabalho foi avaliar um novo algoritmo para o reconhecimento de faces em aplicações (*one-to-many*). Mais especificamente, o trabalho foi desenvolvido enfatizando as técnicas de redução de dimensionalidade estatística utilizando redes neurais artificiais. O reconhecimento das faces baseadas no grupo de treinamento foi efetuado aplicando-se técnicas holísticas e estatísticas de reconhecimento de faces buscando-se extrair as características mais importantes na identificação das faces da base de dados, criando um espaço dimensional propício ao reconhecimento.

Após a detecção e pré-processamento das imagens de face ainda existem problemas com relação as pequenas divergências na iluminação, rotação, e escala entre as faces de treinamento e teste. A meta foi desenvolver um algoritmo flexível com relação à pequenas variações na iluminação, escala, orientação, uso de óculos, feições, oclusões de características, e envelhecimento de uma face. Além de apresentar um algoritmo com taxas de reconhecimento acima de 95%, e rápido², que não tenha um grande custo computacional, ou seja, que a dimensionalidade dos dados não seja grande.

A pesquisa não tratou de dados tridimensionais ou geometricamente detalhados da face, mas sim, características estatísticas da distribuição dos pixels na imagem de face bidimensional em escala de cinza. Portanto, o objetivo é implementar um modelo computacional para o reconhecimento de faces, que seja dinâmico e eficiente no reconhecimento baseando-se nos resultados apresentados por técnicas recentes, avaliando-se o desempenho utilizando o algoritmo do Sub-espaço LDA para a codificação das imagens, e rede neural de Kohonen para a classificação do conjunto de faces.

Outros objetivos secundários e importantes para o projeto são descritos abaixo:

- Abordar os pontos principais relacionados aos métodos PCA e LDA aplicados no reconhecimento de faces para elaborar um texto didático e explicativo.
- Avaliar a interação entre os parâmetros PCA e LDA no algoritmo Subespaço LDA.
- Efetuar os cálculos dos coeficientes PCA através da Rede Neural PCA Adaptativa.
- Realizar os testes com o banco de dados ORL testando 6 diferentes combinações para as faces de teste e treinamento (5, 6, 7, 8, 9, 10 faces de treino).

²Um algoritmo rápido em aplicações reais significa identificar, ou associar uma face com o banco de dados (menos de 1000 imagens) em aproximadamente 1s utilizando computadores pessoais.

- Apresentar os resultados obtidos através de uma seleção aleatória das faces para formar os agrupamentos, e estabelecer um intervalo de confiança para os resultados de 95%.

1.5 Estrutura do Texto.

Este primeiro capítulo mostrou algumas considerações iniciais e o contexto no qual se insere este trabalho buscando apresentar as aplicações e desafios, que envolvem a área relacionada ao reconhecimento de faces. O texto a seguir está organizado da seguinte forma.

O capítulo 2 apresenta alguns temas relevantes ao reconhecimento de faces de forma automática presentes no sistema cognitivo humano. Na seqüência serão ilustrados as principais técnicas utilizadas para o reconhecimento automático de faces visando ressaltar os métodos de maior interesse para a elaboração e sustentação deste trabalho.

O capítulo 3 introduz os conceitos teóricos envolvidos no reconhecimento de faces utilizando a técnica de Análise de Componentes Principais (PCA). São apresentadas a modelagem matemática e a estrutura geral do sistema para efetuar a identificação de uma face.

No capítulo 4 são apresentados os conceitos relacionados às redes neurais e seus processos de aprendizado visando introduzir uma gama de conceitos importantes para na seqüência apresentar as redes neurais PCA, Análise Discriminante Linear (LDA) e SOM, que serão utilizadas em algumas estruturas do sistema proposto para o reconhecimento.

O capítulo 5 refere-se a implementação do algoritmo apresentando detalhadamente as suas estruturas, e mostrando os resultados e avaliações obtidas através dos testes efetuados com o algoritmo.

O capítulo 6 finaliza o projeto com as considerações finais, apresenta as perspectivas e contribuições gerais do trabalho, e define algumas linhas de pesquisas futuras.

Reconhecimento de Faces Humanas

2.1 Considerações Iniciais.

A identificação de uma pessoa através da sua face é um dos mais efetivos métodos não-intrusivos em biometria, no entanto, também consiste em um dos maiores desafios para os pesquisadores na área; envolvendo pesquisas em psicofísica (um ramo da psicologia que estuda a relação entre os estímulos físicos e sua percepção em tarefas envolvendo o processamento mental), neurociência, engenharia, reconhecimento de padrões, análises e processamento de imagens, e visão computacional aplicadas ao reconhecimento de faces por humanos e máquinas.

Reconhecimento de faces em imagens é uma sub-área dos problemas de reconhecimento de objetos gerais, e está entre as aplicações de destaque na análise e compreensão de imagens biométricas. O reconhecimento de faces tem recebido recentemente uma atenção significativa, especialmente durante os últimos anos, e ao menos duas razões esclarecem esta tendência: o primeiro é a larga escala de aplicações comerciais, militares e de segurança civil, e o segundo é a disponibilidade de tecnologias praticáveis após 30

anos de pesquisas. Mas mesmo que os sistemas atuais de reconhecimento automático alcancem um determinado nível de maturidade, seu sucesso é limitado pelas circunstâncias impostas por muitas aplicações reais. Por exemplo, o reconhecimento de imagens de faces adquiridas em um ambiente ao ar livre com mudanças na iluminação e/ou na pose da face é um problema que ainda não pode ser resolvido. Ou seja, os sistemas atuais estão ainda distantes da potencialidade do sistema de percepção humana (ZHAO et al., 2003), e são relativamente pobres quando comparados aos sistemas de reconhecimento de íris ou impressão digital, pois sua robustez está intimamente ligada ao ambiente controlado da aplicação, apresentando taxas de erros de 2 a 25%.

Como na maioria dos outros sistemas de medidas biométricas, um objetivo geral do reconhecimento de faces é conseguir um nível elevado de desempenho na identificação de uma face apresentada ao sistema com referência ao banco de dados conhecido. O desempenho de um algoritmo de reconhecimento de faces é julgado pela combinação de diversos fatores, como: precisão na resposta (com relação ao falso negativo e falso positivo), robustez contra fatores adversos do meio, velocidade no reconhecimento, custo baixo do equipamento, etc. Os fatores adversos incluem condições de luminosidade, ruído na imagem, variações na expressão facial, presença de óculos, mudanças no cabelo, translação e rotação da face com relação à postura (BARRETT, 1998). Estes fatores estão presentes nas aplicações estáticas relacionadas às fotografias em passaportes, cartões de crédito, carteiras de identidade, carteiras de motorista, e retrato falado, sendo cabível destacar todos estes fatores somados à complexidade de processamento de imagens para as aplicações de vigilância em tempo-real.

2.2 A Capacidade Humana para o Reconhecimento de Faces.

Os conceitos de reconhecimento de faces envolvendo o sistema perceptual e cognitivo humano são um bom ponto de referência para cientistas e pesquisadores da área, que buscam construir um sistema de reconhecimento de faces artificial com a mesma capacidade ou superior ao sistema visual humano, sendo necessário tentar compreender a arquitetura envolvendo os estímulos e processos cognitivos presentes nos seres humanos para efetuar o reconhecimento de faces.

Com uma análise superficial em relação à arquitetura do sistema de reconhecimento de faces humanas nota-se uma grande quantidade de informações a serem processadas,

pois os estímulos visuais, auditivos, táteis, etc, são informações preciosas armazenadas de forma individual ou coletiva pelo cérebro para possivelmente recuperar estas informações e estabelecer um reconhecimento futuro. Além disso, em muitas situações o conhecimento contextual com relação ao ambiente em que o indivíduo está inserido torna-se um estímulo importante para o reconhecimento, pois caracteriza o reconhecimento de faces em lugares onde são prováveis a sua existência. Portanto, com a tecnologia atual é difícil desenvolver um sistema que esteja atento a todos estes estímulos, mas o cérebro humano possui uma desvantagem: sua capacidade de armazenamento para o reconhecimento preciso é pequena (CHELLAPPA et al., 1995).

Há evidências sugestivas de que a capacidade visual humana para o reconhecimento de faces é um processo dedicado envolvendo informações holísticas (características globais) e traços do rosto (características locais), não meramente uma aplicação geral de reconhecimento de objetos (ELLIS, 1986). Isso encoraja as aplicações em reconhecimento de faces a serem também específicas, sendo capazes de armazenar uma grande quantidade de imagens de faces, e trabalharem apenas com o estímulo visual caracterizado por imagens 2-D (em escala de cinza) únicas ou em múltiplas vistas de uma face.

Com base nos conceitos anteriores são apresentados alguns estudos em psicologia e neurociência, que têm relevâncias diretas aos cientistas e projetistas interessados em desenvolver algoritmos para o reconhecimento de faces. Por exemplo, características psicológicas importantes para o reconhecimento são encontradas em (BRUCE, 1988) e (SHEPHERD et al., 1981), estudos comparando a importância relativa de diferentes características faciais foram pesquisadas por (ETEMAD; CHELLAPPA, 1997).

Em uma outra abordagem, os sistemas de máquina fornecem ferramentas para conduzir estudos, que comprovem ou expliquem alguns efeitos encontrados na psicologia ou neurociência descritos em (HANCOCK et al., 1998) e (KALOCSAI et al., 1998), como exemplo, uma possível explicação da engenharia para os efeitos registrados em testes com iluminação inferior em Johnston et al. (1992) são os seguintes: quando o sentido real da iluminação é oposto ao sentido geralmente suposto (de cima para baixo ou frontal), um algoritmo forma o mapeamento das possíveis sombras presentes na face, e de posse desta informação modifica a intensidade dos pixels de acordo com um método pré-estabelecido para posteriormente recuperar esta face em um banco de dados com iluminação padronizada. Também, um estudo efetuado por Bruce (1999) com relação a direção da iluminação mostrou a importância da iluminação superior; é mais fácil para os seres humanos reconhecerem faces iluminadas de cima para baixo, que faces iluminadas de baixo para cima. Isto ocorre também com algoritmos de reconhecimento de faces, uma explicação

plausível é a presença de maior informação na região superior da face relevante para o reconhecimento .

Algumas características potencialmente relevantes para o desenvolvimento de sistemas de reconhecimento de faces são apresentadas abaixo segundo a importância dada por Zhao et al. (2003):

- Análise holística e/ou de características (traços do rosto)(BRUCE et al., 1998): ambas as informações holísticas e características são cruciais para a percepção e o reconhecimento das faces. Os estudos sugerem que as descrições globais podem servir como uma aproximação ou redução de possibilidades para uma percepção mais fina, baseada em características. Mas se as características (traços do rosto) forem dominantes, então as descrições holísticas não serão usadas. Por exemplo, na recordação de faces, estudos mostram que os seres humanos focalizam rapidamente em características ímpares tais como as orelhas grandes, um nariz curvado, um olho com qualidades físicas diferentes do normal, etc.
- Escala das características faciais mais significantes (BRUCE, 1988)(SHEPHERD et al., 1981): o cabelo, o contorno da face, os olhos, e a boca (não necessariamente nesta ordem) foram determinados como sendo importantes para perceber e reconhecer as faces (SHEPHERD et al., 1981). Diversos estudos mostraram que o nariz apresenta um papel insignificante; isto pode ser devido ao fato que quase todos estes estudos foram feitos usando imagens frontais. No reconhecimento de faces usando os perfis (que podem ser importantes nas aplicações de retrato falado, onde os perfis podem ser extraídos das vistas laterais), uma forma distintiva do nariz poderia ser mais importante do que os olhos ou a boca (BRUCE, 1988). Também se encontrou que a parte superior da face é mais útil para o reconhecimento de faces do que a parte mais inferior. Os atributos relacionados à estética, tais como beleza, atração, e charme também foram estudados, com a conclusão de que quanto mais atrativas forem as faces, melhores serão suas taxas de reconhecimento; e as menos atrativas vêm em seguida, deixando as faces habituais por último, em termos de facilidade para serem reconhecidas.
- Caricaturas (PERKINS, 1975): um caricatura pode formalmente ser definida como "um símbolo que exagera algumas medidas relativas a qualquer medida que variar de uma pessoa a outra (características que chamam a atenção e destacam a pessoa das demais)." Assim o comprimento de um nariz é uma medida que varia de pessoa

para pessoa, e poderia ser útil como um símbolo para caricaturar alguém. As caricaturas não contêm tanta informação quanto fotografias, mas podem capturar as características mais importantes de uma face.

- Distinção (BRUCE et al., 1994): os estudos mostram que as faces distintas são melhores retidas pela memória e tem um melhor e mais rápido reconhecimento do que as faces típicas. Entretanto, se uma decisão tiver que ser feita a respeito se um objeto é uma face ou não, é notado um maior tempo para reconhecer uma face atípica do que uma face típica. Isso pode ser explicado pelos mecanismos diferentes que estão sendo usados para a detecção e para a identificação realizada pelo cérebro.
- Estudo na análise de frequência espacial (GINSBURG, 1978)(HARMON, 1973)(SERGENT, 1986): estudos realizados por GINSBURG (1978) e HARMON (1973) concluíram que as informações contidas nas faixas de baixa frequência espacial constituem um fator dominante no reconhecimento de faces. Os estudos realizados por SERGENT (1986) mostraram que, dependendo da tarefa específica de reconhecimento, os componentes de baixa e alta frequência podem estabelecer relações diferentes. A classificação de sexo, por exemplo, pode com sucesso ser realizada usando componentes de baixa frequência, enquanto que a identificação requer o uso de componentes de alta frequência. Os componentes de baixa frequência contribuem para a descrição global, e os componentes de alta frequência contribuem para os detalhes mais finos necessários na identificação.
- O reconhecimento com relação ao ponto de vista (invariante ou não)(BIEDERMAN, 1987)(TARR; BULTHOFF, 1995): os desenvolvimentos no reconhecimento visual de objetos foram moldados dentro de uma estrutura teórica introduzida por MARR (1982) em que as vistas diferentes dos objetos são analisadas em uma maneira que permita o acesso (pela maior parte) às descrições invariantes ao ponto de vista. Algumas experiências sugerem que a memória para o reconhecimento das faces é altamente dependente do ponto de vista. A generalização, mesmo de um ponto de vista de perfil a outro, é pobre mas a generalização de uma vista de três quartos (esquerda) à outra (direita) é muito boa.
- Expressões faciais (BRUCE, 1988): baseado em estudos neuropsicológicos, há evidências de que a análise de expressões faciais são realizadas em paralelo com o reconhecimento.

- Reconhecimento de faces em raças diferentes (CHELLAPPA et al., 1995): os humanos reconhecem as pessoas de sua própria raça melhor que pessoas de outras raças. Isto pode ser devido ao fato, de que humanos codificam uma face média com atributos médios, e estas características podem ser diferentes para raças diferentes.

2.3 Reconhecimento de Padrões.

Os sistemas de reconhecimento de faces são baseados nos sistemas de reconhecimento de padrões, que podem ser compostos pelas seguintes partes (DUDA; HART, 1973) , (JAIN et al., 2000) (figura 2.1):

- aquisição de dados: câmeras fotográficas ou de vídeo.
- pré-processamento para eliminar ruídos ou distorções.
- extrator de características (ou atributos): cria um vetor de características com dados extraídos dos objetos adquiridos, reduzindo os dados para atributos, propriedades ou características
- seletor de características; analisa o conjunto de características e elimina as mais redundantes.
- classificador; analisa um padrão obtido e toma uma decisão.

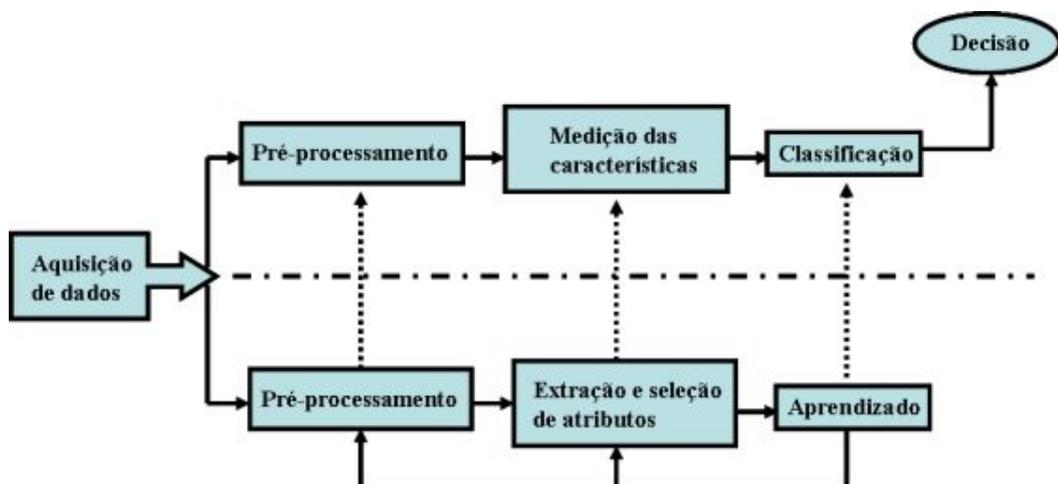


Figura 2.1: Sistema genérico de reconhecimento de padrões.

O classificador toma decisões baseando-se no aprendizado realizado a partir de um conjunto de treinamento, o qual contém exemplos de padrões de todas as classes existentes no sistema. Em reconhecimento estatístico de padrões, a classificação é realizada utilizando estimativas de distribuições probabilísticas. O reconhecedor de padrões é avaliado através de um conjunto de testes, preferencialmente composto por padrões de todas as classes, mas que não estejam no conjunto de treinamento. Além do classificador, o pré-processamento, o extrator e o seletor de características podem ser dependentes dos dados de treinamento. Em reconhecimento de faces cada classe e suas variações representam uma mesma pessoa, portanto o problema de classificação é expresso pela seguinte idéia: dada uma pessoa desconhecida, extrair suas características para obter seu vetor de características x e determinar a qual classe essa pessoa pertence.

Os padrões são tratados como vetores aleatórios, pois um padrão desconhecido pode ser o representante de uma classe conhecida que sofreu alterações aleatórias proporcionadas por ruídos oriundos do método de aquisição (câmeras), da influência de outros fatores externos (iluminação, posição, escala, etc.) ou mesmo dos mecanismos de extração de características intrínsecos ao sistema de reconhecimento.

2.4 Reconhecimento Automático de Faces.

Há vinte e cinco anos atrás, o problema do reconhecimento de faces foi considerado entre os mais difíceis nas áreas de inteligência artificial e visão computacional, mas surpreendentemente na última década, uma série de sucessos em sistemas de reconhecimento de faces mostraram-se tecnicamente viáveis. O desenvolvimento de computadores mais rápidos, a evolução do hardware para aquisição de imagens, e finalmente a inovação nos algoritmos de reconhecimento de faces aplicando conceitos de reconhecimento de padrões, redução de dimensionalidade, redes neurais, detecção eficiente das faces, e etc contribuíram para fortalecer as pesquisas na área de visão computacional restrita ao reconhecimento de faces.

Hoje, os pesquisadores estão trabalhando em relaxar alguns limites impostos pelos algoritmos existentes para o reconhecimento de faces, sendo assim, a meta é conseguir a robustez sob as mudanças devido à iluminação, ao envelhecimento, à rotação em profundidade, e a mudanças de expressão. Está sendo estudado também como tratar das variações na aparência devido a mudanças no cabelo, presença de óculos, e maquiagem.

As pesquisas em reconhecimento de faces através de imagens 2-D em escala de cinza apresentam-se geralmente em três categorias: métodos baseados em características (locais), métodos holísticos (globais), métodos híbridos (soma dos dois métodos anteriores). Os métodos baseados em características confiam na identificação de determinados pontos fiduciais na face tais como os olhos, o nariz, a boca, etc. A posição desses pontos pode ser determinada e usada para computar os relacionamentos geométricos entre os pontos para analisar localmente a região circunvizinha. Assim, o processamento independente dos olhos, do nariz, e de outros pontos fiduciais são executados e combinados em um classificador para então produzir o reconhecimento. Desde que a detecção de pontos das características precede a análise, tal sistema é robusto a variações de translação da face na imagem (ZHAO et al., 2003).

Os métodos holísticos tratam os dados da imagem simultaneamente sem tentar localizar pontos individuais. A face é reconhecida como uma entidade sem regiões diferentes explicitamente isoladas nela. As técnicas holísticas utilizam a análise estatística, redes neurais e transformações lineares ou não. Também requerem geralmente amostras grandes de dados de treinamento. A vantagem dos métodos holísticos é que eles utilizam a face como um todo e não destroem nenhuma informação exclusivamente processando determinados pontos fiduciais. Assim, geralmente fornecem resultados mais exatos de reconhecimento. Entretanto, tais técnicas são sensíveis às variações em posição, escala e assim por diante, o que restringem seus usos às imagens com ambientes controlados.

Uma fusão dos métodos baseados em características e holísticos é chamada de método Híbrido: as idéias consolidadas nestes sistemas estão embasadas no sistema de percepção humana, que usa características locais e a região inteira da face para efetuar o reconhecimento, e portanto um sistema de reconhecimento artificial deveria usar ambos. Assim, pode-se discutir a combinação das melhores características definidas pelos dois métodos anteriores para formalizar esse método.

Observando a pesquisa em reconhecimento de faces, talvez o exemplo mais famoso de um sistema de reconhecimento de faces com possibilidades técnicas aplicativas foi o proposto por Teuvo Kohonen (1988c), que demonstrou, que uma simples rede neural poderia executar o reconhecimento de faces para imagens alinhadas e normalizadas. O sistema de Kohonen não foi um sucesso prático, porque confiou na precisão do alinhamento e da normalização das faces. Nos anos seguintes, muitos pesquisadores tentaram sistemas de reconhecimento de faces baseados em bordas, em distâncias entre as características (olho, nariz, boca, etc), e em outras aproximações com redes neurais. Enquanto diversos sistemas eram bem sucedidos com bases de dados pequenas de imagens alinha-

das, nenhum conseguiu sucesso em ambientes reais com bases de dados grandes, onde a posição e a escala da face são desconhecidas. Sirovich e Kirby (1987) introduziram uma manipulação algébrica para reduzir a dimensionalidade das imagens de faces, e tornar o problema mais generalista (imune a rotação, translação, escala,..., etc) com boa precisão no reconhecimento. Turk e Pentland (1991a)(1991b) do MIT demonstraram, que o erro residual presente na codificação de faces proposta por Kirby e Sirovich (1990) poderia ser usado para detectar as faces, e determinar a posição e a escala precisa delas em uma imagem. Então, detectar e localizar as faces juntamente com o método proposto para o reconhecimento poderia aumentar a confiança nas aplicações em tempo real definindo um ambiente com características limitadas. Esta demonstração culminou em uma explosão de interessados pela área (PENTLAND; CHOUDHURY, 2000).

É válido lembrar que os trabalhos realizados nas décadas de 70 e 80 tinham como métodos tradicionais as estratégias baseadas em análise das características e *Templating Matching*, enquanto que a partir dos anos 90 surgiram métodos baseados em transformações lineares (e também não lineares) e redes neurais, que apresentaram melhores desempenhos com relação às bases de dados grandes (superiores a 100 imagens). Portanto enfatizando as técnicas em reconhecimento de faces através de fotografias desenvolvidas nos últimos anos pode-se recorrer à tabela 2.1.

As técnicas usadas para o reconhecimento de faces são muitas e variadas, mas todas consistem nas mesmas três etapas básicas (SULAIMAN et al.,2000):

1. Segmentação/ detecção da face na cena:

- Mapeamento das bordas
- Análise de pigmentação da pele
- Redes neurais

2. Extração das características presentes na região da face:

- Características holísticas (toda a região da face é entrada para o sistema)
- Características locais (olhos, nariz, boca, etc.)

3. Decisão:

- Identificação
- Reconhecimento de pessoas

Tabela 2.1: Técnicas de reconhecimento de faces (fotografias) desenvolvidas nos últimos anos.

Técnicas	Trabalhos Representativos
Métodos Holísticos	
<i>Principal Component Analysis (PCA)</i>	Aplicações diretas do PCA (KIRBY; SIROVICH, 1990)(TURK; PENTLAND, 1991)
<i>Eigenfaces</i> Probabilístico	Medidas de Probabilidade (MOGHADDAM; PENTLAND, 1997)
<i>Fisherfaces</i> /sub-espço	LDA (BELHUMEUR et al., 1997)(SWETS; WENG, 1996)(ZHAO et al., 1998)
LDA	
SVM	Problema baseado em separação de classes (PHILLIPS, 1998)
ICA	Análises de características com ICA (BARTLETT et al., 1998)
Outros Métodos	
LDA	LDA aplicado diretamente nas faces (ETEMAD; CHELLAPPA, 1997)
PDBNN	Rede neural baseada em decisão probabilística (LIN et al., 1997)
Métodos Baseados em Características	
Método Geométrico Puro	um dos primeiros métodos (KANADE, 1973), métodos recentes (COX et al., 1996)
Arquitetura de Link Dinâmico	(OKADA et al., 1998)(WISKOTT et al., 1997)
Modelo Hidden Markov	(NEFIAN; HAYES, 1998)(SAMARIA, 1994)
Rede Neural de Convolução	(LAWRENCE et al., 1997)
Métodos Híbridos	
<i>Eigenfaces</i> Modular	<i>Eigenface e Eigenmódulos</i> (PENTLAND et al., 1994)
Híbrido LFA	método local de características (PENEV; ATICK, 1996)
Baseado em Componentes	região da face e componentes (HUANG et al., 2003)

- Categorização (raça, sexo, idade, expressão facial, etc.)

Algumas das técnicas mais importantes na área serão brevemente relatadas nos próximos sub-capítulos.

2.5 Métodos Baseados em Características.

Os sistemas que utilizam as feições da face, ou as características geométricas da face usam a priori informações baseadas nas características locais das faces para montar um vetor característico com o intuito de identificar um indivíduo associado ao banco de dados. A face pode ser vista como uma estrutura construída de diversos componentes, tais como os olhos, a testa, o nariz, a boca, etc. Então as variações das posições, das distâncias e de outros parâmetros pertencentes a estes componentes definem a aparência de uma face

e estas variações podem também ser usadas para classificar e distinguir diferentes faces. Por exemplo, estes parâmetros podem ser as distâncias entre os olhos, o tamanho da boca, e a distância entre a boca e o nariz, como na figura 2.2. Todos esses tipos de medidas são exemplos de biometria da face humana.

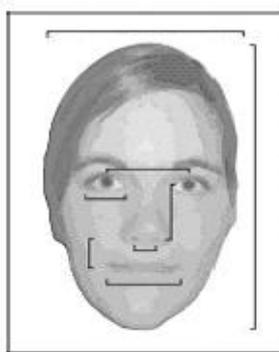


Figura 2.2: Biometria facial.

As primeiras implementações em reconhecimento de faces foram feitas utilizando os conceitos de biometria facial. Kanade (1973) descreve um dos primeiros sistemas construídos para essa tarefa. O sistema localiza automaticamente características tais como cantos dos olhos, narinas, a boca etc., então um conjunto de dezesseis parâmetros faciais, que corresponde a essas características é usado para definir um vetor característico, e para o reconhecimento é feita uma medida euclidiana simples da distância entre o vetor teste e os vetores do banco de dados para definir o grau de similaridade entre eles e possivelmente associar um indivíduo. O melhor desempenho encontrado para este sistema foi 15 identificações corretas para 20 faces de teste.

Brunelli e Poggio (1992) fizeram inúmeras experiências com um banco de dados de 47 pessoas com quatro imagens frontais de cada uma utilizando a análise geométrica da face. No processamento para formar o vetor característico, as bordas são extraídas com filtros de borda horizontal e vertical formando duas imagens, que na seqüência são utilizadas para detectar a posição da boca e do nariz, enquanto os olhos são detectados através de cinco *templates* com distâncias intra-oculares diferentes. Com as características locais das faces detectadas (olhos, nariz e boca) e com as distâncias geométricas referentes às bordas da face e seus elementos característicos, os autores discutiram quais medidas seriam mais apropriadas para efetuar uma classificação e identificação automática. A figura 2.3a mostra as características geométricas analisadas. Para complementar o trabalho, Brunelli e Poggio (1993) compararam o algoritmo do trabalho anterior baseado em características geométricas com um baseado em *Template Matching* das imagens em escala de cinza. Os

Templates foram definidos armazenando a face toda, os olhos, o nariz e a boca de cada imagem de treinamento com vistas diferentes para cada indivíduo, conforme figura 2.3b. Segundo os autores o resultado foi 90% de reconhecimento correto para o método baseado em características geométricas e 100% para o reconhecimento utilizando *Template Matching*.



Figura 2.3: a) Características geométricas (em branco) utilizadas no reconhecimento de faces, b) Diferentes regiões utilizadas na estratégia *Template Matching* (BRUNELLI; POGGIO, 1993).

Os métodos baseados em características geométricas são intimamente dependentes da forma como ocorre o pré-processamento para extrair as características, pois a resolução das imagens, a iluminação do ambiente, as rotações e translações pequenas das faces frontais, a profundidade da face na cena, o *threshold* adotado para efetuar a binarização, e até mesmo a quantidade de características levantadas para formar o vetor característico podem influenciar diretamente na performance do sistema. Desta forma, as mais infundáveis restrições impostas ao bom funcionamento do método tornam sua aplicação no mundo real pouco robusta e eficiente. A estratégia utilizando *Templates* tem uma boa performance no reconhecimento de faces, além de ser generalista, mas os *Templates* ocupam muito espaço nos bancos de dados, e o tempo de processamento é muito alto quando comparado com métodos de reconhecimento característico (BRUNELLI; POGGIO, 1993).

2.6 Métodos Holísticos (e Estatísticos).

O reconhecimento de faces aplicando os métodos holísticos utiliza as informações globais das faces para executar o reconhecimento. A informação global das faces é representada por um vetor de face em um sistema de coordenadas reduzidas do normal, e este ve-

tor é derivado diretamente da informação dos pixels das imagens da face. Este vetor no sistema de coordenadas reduzidas captura distintamente a variação entre as faces individuais diferentes da média e conseqüentemente é usado para identificar excepcionalmente os indivíduos.

Para análises matemáticas, os dados da imagem são representados como vetores, isto é, como pontos em um espaço dimensional elevado. Por exemplo, uma imagem 2-D em escala de cinza descrita pela matriz $P(m, n)$ pode ser representada por um vetor $x \in R^{m \cdot n}$, com os seus pixels concatenados em linhas ou colunas para formar um só vetor no espaço $R^{m \cdot n}$. Apesar deste espaço altamente dimensional representar o mundo real e natural das imagens de face, ele não é amigável para efetuar o reconhecimento, então observando as redundâncias presentes nas imagens pode-se caracterizá-las em um espaço dimensional menor, onde apenas as informações mais relevantes para o reconhecimento estão presentes. O objetivo preliminar da análise para construir o sub-espaço é identificar, representar, e parametrizar uma transformação linear de acordo com alguns critérios ótimos que viabilizem o reconhecimento.

Um dos fatores mais importantes encontrados nos métodos estatísticos é a compressão das informações contidas nas imagens (redução de dimensionalidade) para posteriormente ocorrer a classificação, e os métodos mais relevantes na área são Análise de Componente Principal (PCA) (TURK; PENTLAND, 1991a)(TURK; PENTLAND, 1991b), Análise de Componente Independente (ICA) (BARTLETT et al., 1998), e Análise Discriminante Linear (LDA) (BELHUMEUR et al., 1997) (SWETS; WENG, 1996a)(SWETS; WENG, 1996b). Os três métodos são baseados em análises estatísticas dos vetores de face, e em transformações lineares e não lineares dos sistemas de coordenadas onde se encontram estes vetores. Cada método tem sua própria forma de atuação e representação (vetores da base) para formar um espaço dimensional menor baseado em conceitos estatísticos diferentes. Projetando o vetor de face nos vetores da base do sistema, têm-se os coeficientes da projeção definidos como a nova representação da característica de cada imagem de face. O reconhecimento da face teste é efetuado através de um classificador, que deve analisar os coeficientes dos protótipos das faces de treinamento comparando-os com os coeficientes da face teste, e de acordo com um *threshold* previamente estabelecido definir se a face teste está no banco de treinamento ou não. Uma das formas mais simples de efetuar a classificação é utilizar as medidas de distâncias entre vetores, como por exemplo, distância Euclidiana, do Cosseno, Mahalanobis, etc.

Sendo assim, definindo X como o conjunto de todas as faces de treinamento, ou seja, $X = (x_1, x_2, \dots, x_i, \dots, x_N)$ uma matriz $pt \cdot N$, onde cada x_i é um vetor face de dimensão pt

(concatenação da imagem $m.n = pt$). pt representa o número total de pixels na imagem de face, e N a quantidade de faces diferentes presente no grupo de treinamento. Pode-se definir as três representações (PCA, LDA e ICA) como uma transformação linear do vetor original da imagem em um vetor caracterizado pela projeção, isto é:

$$Y = W^T.X \quad (2.1)$$

Onde Y é a matriz com a representação dos vetores originais no novo espaço vetorial ($d.N$), $X(pt.N)$ são os vetores do espaço original, e $W(pt.d)$ é a matriz transformação. Note que ' d ' é a dimensão do espaço Y e pt o tamanho dos vetores, ressaltando que d sempre é menor que a dimensão original representada por pt .

2.6.1 Análise de Componentes Principais (PCA).

A técnica de Análise de Componentes Principais (PCA), também conhecida como expansão de Karhunen-Loève consiste em reduzir a dimensionalidade dos dados efetuando uma análise estatística baseada na variância, e na redundância presente nos dados. Quando a técnica PCA é utilizada em aplicações de faces recebe o nome de *Eigenfaces*(Autofaces), que é derivado do prefixo alemão *eigen* (auto), significando próprio ou individual.

Um dos problemas básicos analisando e visualizando conjunto de dados multidimensional, como o caso das faces, é a representação dos dados e a sua estruturação em um sub-espaço de dimensão mais baixa. A implementação geral para resolver este problema é representar os dados de modo que as distâncias entre os pontos, que representam as amostras no sub-espaço de dimensão mais baixa corresponda o tanto quanto possível às dissimilaridades entre os pontos no espaço dimensional original. Um dos mais famosos métodos nessa tarefa é a técnica PCA, que objetiva encontrar um sub-espaço contendo como bases ortonormais, os vetores que definem as principais direções de distribuição dos dados no espaço original otimizando o critério de correlação entre os dados, ou seja, o sub-espaço das principais componentes minimiza a correlação cruzada entre as amostras dos dados (GROB, 1994).

Teoricamente é desejável extrair as informações relevantes contidas na imagem da face, codificando a imagem de forma eficiente, e comparando uma imagem codificada com um banco de dados similar. Uma implementação básica pode basear-se na captura da variação dos níveis de cinza em uma coleção de imagens de face, independente de qualquer análise dos traços do rosto. Portanto, estatisticamente podemos aplicar a técnica

PCA encontrando os *eigenvectors*(autovetores), que constituem as bases ortonormais do sub-espaço de faces e seus respectivos *eigenvalues*(autovalores) obtidos através da matriz de covariância formada pela correlação entre os pixels. Em resumo, cada imagem de face no conjunto de treinamento pode ser exatamente representada em termos de uma combinação linear dos autovetores, e os coeficientes desta combinação serão os novos representantes da face no sub-espaço dimensional.

Matematicamente, para aplicações em reconhecimento de faces este método pode ser formulado como segue:

Considerando-se a equação 2.1 referente à representação de uma transformação linear e visando-se encontrar a matriz transformação W constituída pelos autovetores têm-se: cada imagem de face definida por uma matriz de pixels com intensidades na escala de cinzas, então definindo os vetores $\Gamma_1, \Gamma_2, \Gamma_3, \dots, \Gamma_M$ associados aos pixels das imagens $Im(x, y)$ concatenados por linhas formando vetores $p.p$, e utilizados no treinamento do sistema. A média é calculada da seguinte forma:

$$\Psi = \frac{1}{M} \cdot \sum_{n=1}^M \Gamma_n \quad (2.2)$$

Através do valor definido pela média, são calculados os desvios de cada imagem com relação à média:

$$\phi_n = \Gamma_n - \Psi \quad (2.3)$$

O conjunto de vetores definido por ϕ está normalizado com relação à média, e pronto para ser utilizado no cálculo dos autovetores μ_k e autovalores λ_k da matrix de covariância da equação 2.4.

$$C = \frac{1}{M} \cdot \sum_{n=1}^M \phi_n \phi_n^T = A \cdot A^T \quad (2.4)$$

onde a matriz $A = [\phi_1, \phi_2, \dots, \phi_M]$

Os autovetores calculados através da matriz de covariância formam as bases ortonormais do sub-espaço dimensional, portanto cada face representada por um vetor no espaço original deve ser identificada no sub-espaço, mas primeiramente deve-se escolher de forma heurística os M' primeiros autovalores associados aos primeiros autovetores, que

possuem valores numericamente expressivos, e conseqüentemente estão relacionados às maiores variâncias no conjunto de dados formando assim um espaço vetorial comprimido com relação ao original. Sendo assim, as faces de treinamento são projetadas do seguinte modo no sub-espaço definido pelos autovetores selecionados:

$$\omega_n = \mu_n^T \cdot (\Gamma - \psi) \quad \text{para } n = 1, \dots, M' \quad (2.5)$$

Portanto os ω_n definem um vetor de pesos $\Omega^T(i) = [\omega_1^i, \omega_2^i, \dots, \omega_{M'}^i]$, $i = 1, \dots, M$ (onde M é a quantidade de imagens), que expressa cada face de treinamento no sub-espaço definido pelos autovetores formando uma nuvem de pontos em uma determinada região do sub-espaço caracterizando um espaço de faces, ou seja, temos uma região que limita e engloba todos os pontos representados pelos vetores $\Omega^T(i)$. Este vetor Ω^T descreve a contribuição de cada autovetor na representação da imagem i .

Na identificação, o sistema encontra os pesos ω que representam a face de entrada no sub-espaço, isto é conseguido projetando-se o vetor Γ_{nova} no sub-espaço, ou seja:

$$\omega_n = \mu_n^T \cdot (\Gamma_{nova} - \Psi) \quad \text{para } n = 1, \dots, M' \quad (2.6)$$

Os pesos formam um vetor $\Omega^T(nova) = [\omega_1, \omega_2, \dots, \omega_{M'}]$, assim pode-se utilizar qualquer algoritmo de reconhecimento de padrões para descrever as classes referentes às faces. Portanto, uma face de entrada pode ser identificada com alguma classe de treinamento (onde cada pessoa é uma classe), ou como uma face desconhecida, ou simplesmente como uma imagem não face.

2.6.2 Evolução no Reconhecimento de Faces Aplicando PCA.

Sirovich e Kirby (1987) foram os primeiros a aplicar a expansão de Karhunen-Loève (ou PCA) no intuito de investigar os problemas de caracterização, identificação, e distinção das faces humanas com uma abordagem estatística para o reconhecimento dos padrões referentes às classes de faces. Eles demonstraram que qualquer face particular pode ser comprimida e representada através dos coeficientes referentes a esta face no sub-espaço de coordenadas descritas pelos autovetores da matriz de covariância associada aos pixels de um conjunto de imagens de treinamento. Sendo assim, quaisquer conjuntos de faces podem ser classificados armazenando um pequeno conjunto de coeficientes para cada

face e um pequeno conjunto de faces padrões conhecidas como autofaces, que são as bases ortonormais do sub-espço.

As imagens de faces utilizadas para esta investigação inicial eram compostas de 115 faces frontais de indivíduos do sexo masculino atentando ao cuidado de selecionar pessoas claras caucasianas para formar um banco de dados homogêneo. As imagens foram digitalizadas em 128x128 pixels com 2⁸ níveis de cinza e pequenas variações de iluminação, tendo como prioridade o ajuste da profundidade, o alinhamento vertical (linha de simetria da face) e horizontal (linha cruzando a pupila dos olhos) na aquisição das faces. Após a aquisição das imagens foi efetuado manualmente um recorte para extrair somente a região dos olhos e nariz para formar um novo banco de dados, e posteriormente analisar as interferências causadas pelo cabelo e pela boca na reconstrução e identificação das faces. Na figura 2.4 é apresentado um exemplo das faces do banco de dados, e um gráfico relacionando a porcentagem de erro na reconstrução das imagens com o número de autovetores selecionados, na figura 2.5.



Figura 2.4: Imagens de Faces utilizadas para formar o banco de dados (Sirovich; Kirby, 1987)

Para investigar a robustez e generalização na reconstrução das faces, também foram testadas duas imagens femininas recortadas e uma masculina com baixa luminosidade. A experiência foi realizada com 40 autovetores para a codificação encontrando uma taxa de erro de 3.9% e de 2.4% respectivamente para as duas faces femininas. Analisando este resultado concluíram que o erro estava próximo ao encontrado para as faces masculinas, então o sistema de codificação é independente com relação ao sexo, mas vale ressaltar que se as imagens femininas fossem completas com cabelo e boca talvez o erro de reconstrução seria maior, pois o cabelo feminino geralmente ocupa um volume maior que o masculino absorvendo uma grande quantidade de pixels na imagem, e sendo os autovetores calculados através das faces masculinas possivelmente haveria um erro maior na reconstrução para as faces femininas. Na experiência com a face masculina foi encontrada uma taxa de erro de 7.8% confirmando uma pequena vulnerabilidade do sistema

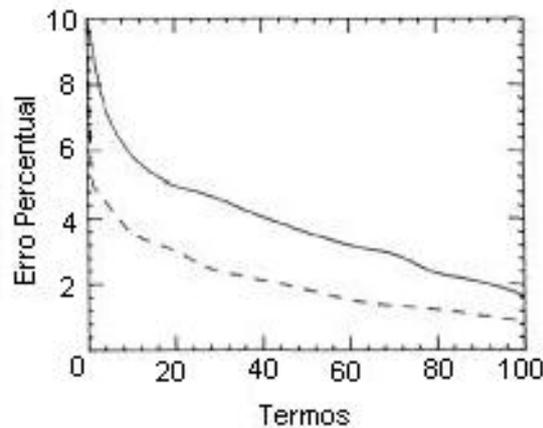


Figura 2.5: Erro percentual versus número de autovetores (Sirovich; Kirby, 1987).

curva contínua - face completa

curva pontilhada - face recortada.

com relação à luminosidade, lembrando que se a face fosse completa este erro seria ainda maior, pois o relevo da face apresentaria um maior contraste nos níveis de cinza entre os picos e vales da mesma, e possivelmente o cabelo também contribuiria para o aumento no erro de reconstrução. Mas enquanto esta investigação se limitou ao uso das faces parciais, em um outro artigo investigaram as vantagens de se usar as faces completas para o reconhecimento.

Kirby e Sirovich (1990) estenderam sua investigação anterior analisando o comportamento da expansão de Karhunen-Loève usando faces frontais completas e faces refletidas (em torno do eixo y) explorando as simetrias naturais presentes na face humana. Para a experiência foram selecionadas 100 imagens de face adquiridas pelo trabalho anterior, sendo as faces completas definidas por uma estrutura oval sobreposta à face, que eliminava o cabelo e o fundo das imagens. As faces refletidas foram definidas a partir das imagens selecionadas; considerando cada imagem como sendo uma função $\varphi(x, y)$ representando a intensidade de cada pixel com relação à sua coordenada espacial, e definindo o alinhamento vertical (linha de simetria da face) centrado em $x = 0$, foram obtidas as imagens refletidas com relação à simetria par $\varphi(x, y) = \varphi(-x, y)$ e simetria impar $\varphi(x, y) = -\varphi(-x, y)$ formando um novo conjunto de faces com 200 imagens.

A informação mais relevante apresentada por esta experiência foi obtida através da análise dos primeiros autovetores com relação à distribuição dos autovalores representantes das simetrias pares e impares. Não surpreendentemente, a maioria dos autovetores (cinco dos primeiro seis) corresponderam aos maiores autovalores de simetria par. Isto é

compreensível, visto que a face é simétrica, mas o terceiro autovetor tem simetria impar possibilitando analisá-lo como sendo um forte representante das variações de intensidade luminosa no conjunto de treinamento. Então, examinando os coeficientes relacionados aos autovetores para a projeção de qualquer face no sub-espço se tem uma medida relativa de sua simetria. Por exemplo, se há um coeficiente relativamente grande que corresponde ao autovetor número três haverá grandes chances da face ser mais assimétrica que a média, ou possuir uma distribuição luminosa diferente da média. Na figura 2.6 são apresentados os nove primeiros autovetores obtidos na experiência.



Figura 2.6: Primeiros nove autovetores, na ordem da esquerda para direita e de cima para baixo (KIRBY; SIROVICH, 1990).

Motivado pelos trabalhos de Sirovich e Kirby (1987)(1990), Turk e Pentland (1991a) aplicaram a expansão de Karhunen-Loève, para alcançar o reconhecimento de faces. Turk e Pentland buscaram executar o reconhecimento de faces através de uma análise global e holística em contraposição aos outros métodos, que estavam sendo implementados visando aproveitar as características faciais individuais e locais. A grande contribuição deste trabalho para a área foi a descrição de uma metodologia visando definir um mapa de faces para viabilizar a detecção, e um espaço de faces para efetuar o reconhecimento utilizando o sub-espço Autofaces. Para avaliar a viabilidade do método proposto para reconhecimento de faces, eles fizeram experimentos (treinamento) com banco de imagens

de faces estáticas, e construíram um sistema para localizar e reconhecer as faces em um ambiente dinâmico para efetuar o teste.

O sub-espço de faces foi definido pelos autovetores encontrados através do conjunto de treinamento. Assim, a imagem de teste deveria ser projetada no sub-espço originando um vetor de coeficientes para representá-la, usando a Distância Euclidiana para encontrar a classe de face mais próxima da amostra de teste determinando se a imagem teste era um individuo conhecido ou não. Além disso, também foi possível determinar se a imagem teste era uma face ou não. Para propósitos ilustrativos, serão esquematizadas as possibilidades envolvendo as duas métricas descritas:

A métrica para a classe é definida como:

$$\varepsilon_k^2 = \|(\Omega_{nova} - \Omega_k)\|^2 \quad (2.7)$$

$$\text{Onde } \|(\Omega_{nova} - \Omega_k)\|^2 = (\Omega_{nova} - \Omega_k)^T \cdot (\Omega_{nova} - \Omega_k)$$

E Ω_{nova} é a projeção da face de teste no sub-espço Autofaces e Ω_k é o vetor médio representante da classe k associada as pessoas presentes no banco de dados.

A métrica para o espço de faces é definida como:

$$\varepsilon_{Fespace}^2 = \|(\phi_{nova} - \phi_f)\|^2 \quad (2.8)$$

$$\text{Onde } \|(\phi_{nova} - \phi_f)\|^2 = (\phi_{nova} - \phi_f)^T \cdot (\phi_{nova} - \phi_f)$$

E ϕ_{nova} é a imagem de entrada ajustada pela média ($\phi = \Gamma - \Psi$) e ϕ_f é a sua imagem recuperada após a projeção.

Obtendo a medida entre a imagem e a classe mais próxima, e entre a imagem e o espço de faces pode-se determinar através de um *threshold* pré-definido, se a imagem é uma face e a que classe pertence. Por exemplo, se as distâncias encontradas em ε_k e $\varepsilon_{Fespace}$ estão abaixo do limiar para a classe de face e espço de face, então a face foi reconhecida.

A detecção das faces em imagens dinâmicas foi implementada em duas etapas; na primeira etapa a pessoa é detectada na imagem através do seu movimento no ambiente onde está fixada a câmera, sendo que um algoritmo de *tracking* encontra o corpo e posteriormente a cabeça analisando a seqüência dos frames. Além de encontrar a cabeça,

o algoritmo permite estimar a escala referente à face (ou objeto). Veja o diagrama em blocos da primeira etapa (detecção - figura 2.7):

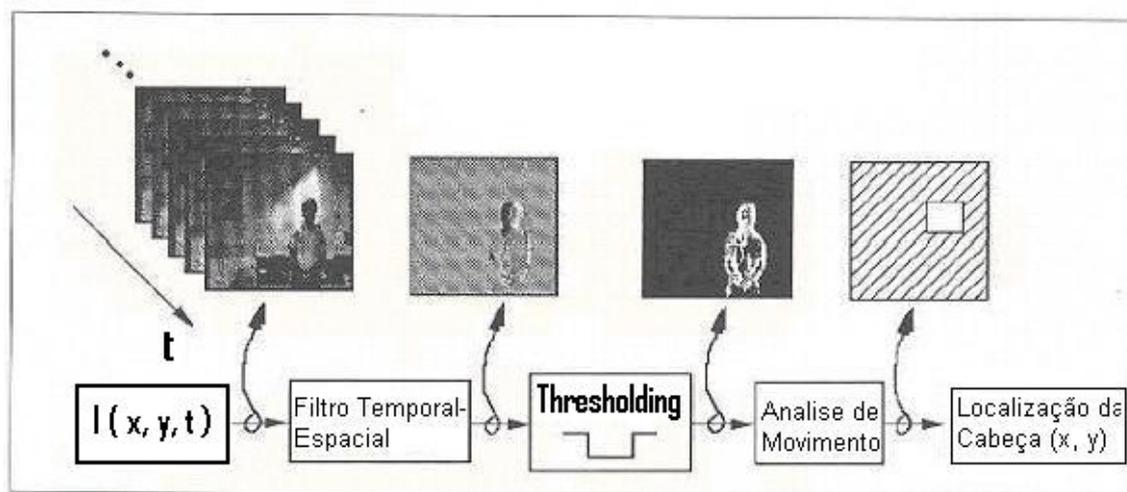


Figura 2.7: Sistema de localização e *tracking* da cabeça (TURK; PENTLAND, 1991a).

Na segunda etapa, o objeto em movimento, que pode ser uma face é re-escalado para ter a mesma dimensão das imagens de faces utilizadas no treinamento, e posteriormente esta imagem do objeto é projetada no espaço de Autofaces e calculada a sua distância com relação ao espaço de faces. De acordo com um *threshold* esta imagem é classificada como face e possivelmente identificada, ou é classificada como um objeto e descartada.

Turk e Pentland (1991a)(1991b) foram os primeiros a demonstrar a viabilidade de um sistema para reconhecimento de faces trabalhando com um banco de dados grande. Eles digitalizaram 2500 imagens de faces sob condições controladas; 16 pessoas foram digitalizadas com todas as combinações possíveis para três diferentes orientações da cabeça, três escalas de foco diferentes, e três condições diferentes de luminosidade. Para finalizar, a resolução das imagens foram modificadas para formar seis grupos com resolução de 512x512 pixels até 16x16. Os resultados médios obtidos em um primeiro experimento com a presença de apenas uma face por pessoa para o treinamento foram de 96% de classificação sob mudanças na condição de luminosidade, 85% sob mudanças na orientação, e 64% sob variação no tamanho das imagens. Uma melhor avaliação dos efeitos condicionados às mudanças de iluminação, orientação e escala aplicando o método Autofaces podem ser encontrados em (LUO; YUWAPAT, 2003).

No início dos trabalhos aplicando Autofaces surgiu a seguinte dúvida: quais autovetores seriam os mais importantes para efetuar o reconhecimento, então O'Toole et al. (1993)

detalharam uma investigação buscando analisar se a seleção de autovetores com diferentes variâncias melhoraria o reconhecimento. Nas suas experiências, os autores exploraram os efeitos de selecionar autovetores com variâncias mais altas comparando-os aos resultados obtidos com os autovetores com variâncias mais baixas para o reconhecimento. Eles descobriram que os primeiros autovetores com maiores variâncias capturam as características gerais das faces (formato do rosto, tamanho e local dos olhos, nariz, boca), enquanto que os autovetores de variância mais baixa capturam as características particulares inerentes às faces que divergem muito da média (cabelo diferente, nariz grande, boca larga, etc).

Visando demonstrar a possibilidade das aplicações referentes ao sistema Autofaces no mundo real, Pentland et al.(1994) extrapolaram o tamanho dos bancos de dados utilizados até então, que eram restritos a algumas centenas de faces com poucas pessoas; eles trabalharam com milhares de imagens (7562 imagens de 3000 pessoas) em diferentes vistas, pois transformações na posição, orientação, escala e iluminação causam uma variação substancial na aparência das faces e assolam os problemas envolvendo o reconhecimento de faces no mundo real. Para aumentar a performance do método Autofaces proposto por Turk e Pentland (1991a)(1991b) em bancos de dados grandes, os autores generalizaram a aplicação para um espaço de faces baseado em vistas e modularizado. A formulação baseada em vistas permite o reconhecimento baseado em várias orientações da cabeça, e a descrição modular permite a incorporação das características faciais como olhos, nariz, boca, ..., etc de forma isolada para a análise.

O sistema Autofaces baseado em vistas foi construído através de M sub-espacos de autofaces separados, onde cada um capturava a variação dos N indivíduos em um ponto de vista e escala comuns. Esta arquitetura define um processo paralelo de observação, portanto cada sub-espaco de autofaces possui os seus autovetores, que explanam uma orientação e escala de acordo com o treinamento. No teste, para a imagem de entrada é calculada a distância desta para os sub-espacos de autofaces, e o sub-espaco mais próximo é usado para efetuar o reconhecimento. Já a aplicação do autofaces modularizado é baseado na formulação de vários sub-espacos de autofaces definidos pelas características individuais da face (boca, olhos, nariz), ou sub-espaco *eigeneyes*(autoolhos), *eigennose*(autonariz) e *eigenmouth*(autoboca), portanto as faces são recortadas formando regiões estanques para o reconhecimento(CAMPOS et al., 2000). Adicionalmente, sugeriram também que uma estratégia partindo de um nível grosseiro para um refinamento no reconhecimento poderia ser empregada, onde no nível grosseiro foi implementado o método de Autofaces, e no refinamento foram implementados os sub-espacos autoolhos, autonariz e autoboca para reconhecer indivíduos holisticamente, e por suas características faciais locais (sistema

conhecido como representação modular combinada). Para demonstrar o desempenho do Autofaces modularizado, os autores selecionaram 45 indivíduos com expressões faciais neutras e sorrindo; o treinamento foi efetuado com as imagens neutras e os testes com as imagens contendo expressões de sorriso. O resultado encontra-se no gráfico da figura 2.8.

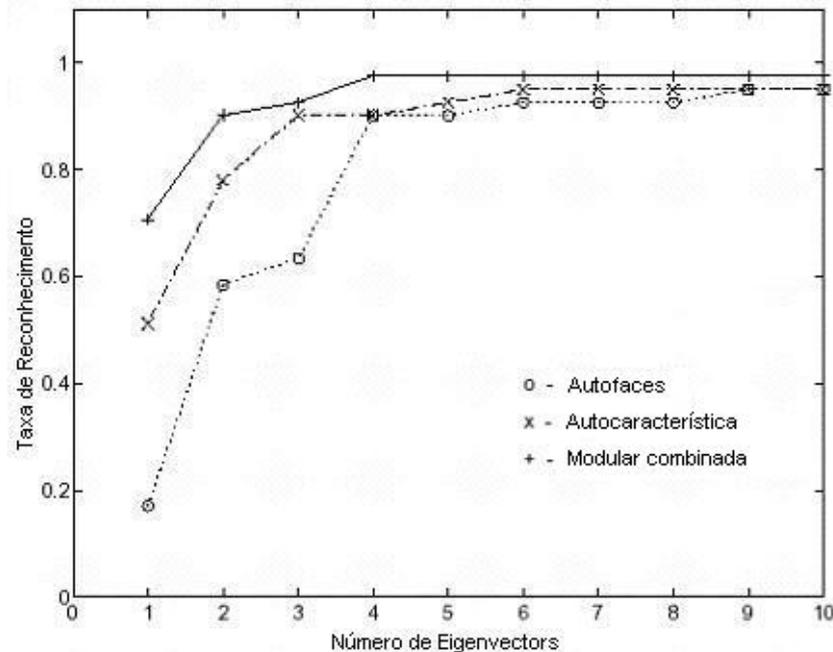


Figura 2.8: Taxa de Reconhecimento para os métodos autofaces, autocaracterística (olhos, nariz e boca) e representação modular combinada (PENTLAND et al., 1994).

Analisando todo o processo envolvido no reconhecimento de faces utilizando Autofaces surge uma dúvida; será que a Distância Euclidiana utilizada como classificador para efetuar a medida de similaridade entre uma face de entrada e o grupo de treinamento é viável? Respondendo esta pergunta Yambor et al. (2000) estudaram diferentes medidas de distâncias como classificador visando aumentar a performance no estágio de reconhecimento das faces utilizando Autofaces. As medidas de distâncias implementadas são apresentadas a seguir.

L1 - Distância City Block

$$d(x, y) = \|x - y\| = \sum_{i=1}^k \|x_i - y_i\| \quad (2.9)$$

L2 - Distância Euclidiana

$$d(x, y) = \|x - y\|^2 = \sum_{i=1}^k \|x_i - y_i\|^2 \quad (2.10)$$

Angular - Ângulo entre os vetores de imagem

$$d(x, y) = -\frac{x \cdot y}{\|x\| \cdot \|y\|} = -\frac{\sum_{i=1}^k x_i \cdot y_i}{\sqrt{\sum_{i=1}^k (x_i)^2 \sum_{i=1}^k (y_i)^2}} \quad (2.11)$$

Mahalanobis - Distância de Mahalanobis

$$d(x, y) = -\sum_{i=1}^k \frac{1}{\sqrt{\lambda_i}} \cdot x_i \cdot y_i \quad (2.12)$$

Na equação 2.12 λ_i é o i -ésimo autovalor.

Como resultado, Yambor et al. (2000) observaram uma melhora significativa utilizando a distância de Mahalanobis, quando comparada com as outras três.

2.6.3 Análise Discriminante Linear (Linear Discriminant Analysis - LDA).

Existem muitas técnicas para a classificação de dados na literatura estatística, mas duas delas são muito eficientes na compressão e reconhecimento de faces: PCA e Análise Discriminante Linear (LDA), que são comumente usadas para a classificação de dados e na redução de dimensionalidade. PCA já foi anteriormente discutido, e Análise Discriminante Linear pode ser entendido como um método, que busca maximizar o raio de variância entre as classes (*between-class*), e paralelamente minimizar a variância dentro das classes (*within-class*) garantindo a máxima separabilidade entre as classes de um conjunto particular de dados, sendo aplicada uma transformação linear visando encontrar um sistema de coordenadas ótimas, que melhor represente os dados com máxima separabilidade. O uso do LDA para classificação de dados foi primeiramente aplicado ao problema de classificação e reconhecimento da fala, e posteriormente aplicado em faces visando prover maior performance que o PCA.

A principal diferença entre LDA e PCA é que: a aproximação PCA proporciona características que capturam as direções principais com relação ao espalhamento dos dados

observando as diferenças significantes entre as imagens, mas não tenta reduzir o espalhamento dos pontos característicos dentro da classe (intra-classe), pois trata cada imagem como única, sem agregar informações sobre a sua classe. Em outras palavras, desde que nenhuma informação de agrupamento das classes é usada, exemplos da mesma classe e de classes diferentes são tratados do mesmo modo. Porém, LDA usa a informação de agrupamento das classes e nos permite achar autovetores, ou *eigenfeatures* (autocaracterísticas) nas quais são enfatizadas as variações entre faces de indivíduos diferentes, enquanto as variações na mesma face do indivíduo provocada pelas mudanças de iluminação, expressão facial, orientação etc., é minimizada. Portanto PCA faz classificação de características e LDA faz classificação de dados. Em PCA, a forma e local do conjunto de dados originais mudam para um sub-espço diferente, enquanto que o LDA não muda o local, mas apenas tenta prover mais separabilidade entre as classes e estabelecer uma melhor região de decisão entre elas (CHELLAPPA; ETEMAD, 1997).

LDA é uma técnica clássica em reconhecimento de padrões desenvolvida primeiramente por Robert Fisher em 1936 (FISHER, 1936), por isso também é chamada de *Fisher's Linear Discriminante* (FLD). Para compreender claramente a forma com que os dados são tratados, é viável iniciar a abordagem do problema de reconhecimento de faces com LDA através de um exemplo com duas classes distintas em um ambiente 2-D.

Assumindo um conjunto de amostras bi-dimensional igual a $x^1, x^2, x^3, \dots, x^n$ distribuídas entre duas classes, sendo a classe ω_1 com N_1 amostras e a classe ω_2 com N_2 amostras, é desejável obter um escalar y projetando as amostras x em uma reta, que maximize a separabilidade dos escalares (veja figura 2.9).

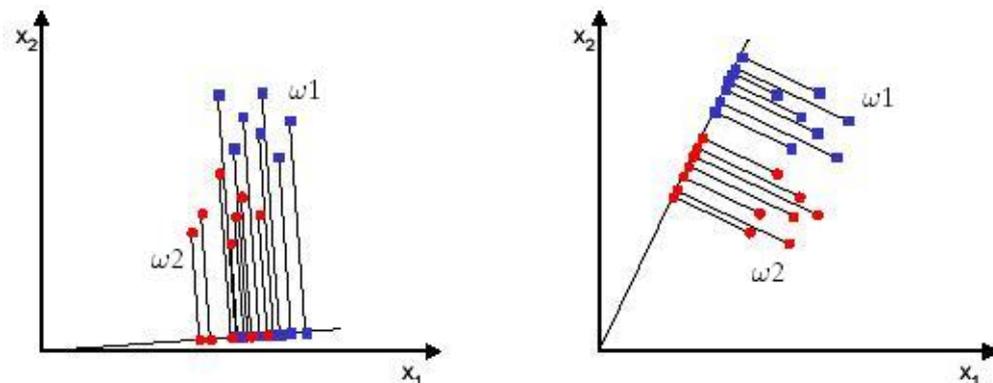


Figura 2.9: Reta indicando a posição de maior separabilidade dos conjuntos.

Observando o problema definido acima, busca-se encontrar um "vetor ótimo" para a projeção, onde a distância de separação entre as classes seja maximizada, ou seja o vetor médio de cada classe no espaço original e no espaço LDA é:

$$\mu_i = \frac{1}{N_i} \cdot \sum_{x \in \omega_i} x \quad (2.13)$$

e

$$\tilde{\mu}_i = \frac{1}{N_i} \cdot \sum_{y \in \omega_i} y = \frac{1}{N_i} \cdot \sum_{x \in \omega_i} W^T \cdot x = W^T \cdot \mu_i \quad (2.14)$$

Pode-se então escolher a distância entre a média das projeções para formar uma função com o intuito de maximizar esta distância.

$$J(W) = \|\tilde{\mu}_1 - \tilde{\mu}_2\| = \|W^T \cdot (\mu_1 - \mu_2)\| \quad (2.15)$$

Mas a distância entre as médias das projeções não é uma boa medida, pois não leva em consideração o desvio padrão presente em cada classe (figura 2.10).

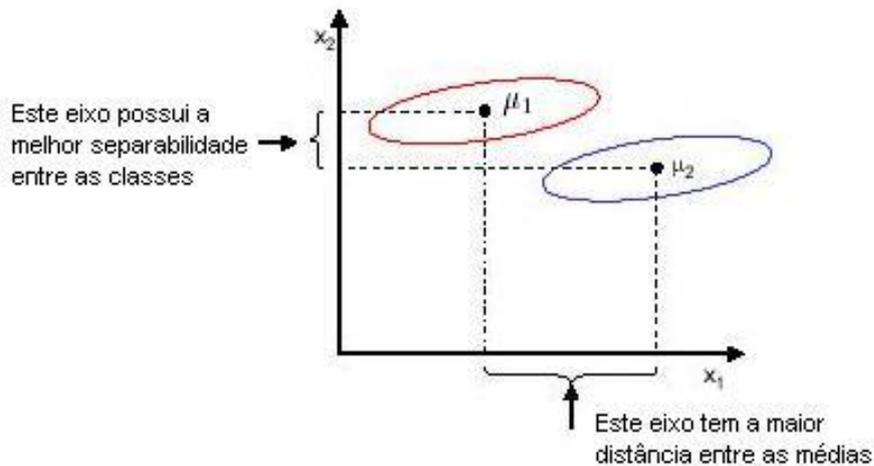


Figura 2.10: Separabilidade entre os dois eixos reais.

A solução proposta por Fisher é maximizar uma função que representa a diferença entre as médias, normalizada pela medida de dispersão inerente a cada classe. Sendo assim, para cada classe define-se a dispersão como sendo equivalente à variância:

$$\widetilde{S}_i^2 = \sum_{y \in \omega_i} (y - \widetilde{\mu}_i)^2 \quad (2.16)$$

Desse modo, o LDA é definido como uma função linear $W^T \cdot X$ que maximiza a seguinte função critério:

$$J(W) = \frac{\|\widetilde{\mu}_1 - \widetilde{\mu}_2\|^2}{\widetilde{S}_1^2 + \widetilde{S}_2^2} \quad (2.17)$$

Na equação 2.17 W definirá um vetor para a projeção das amostras, onde exemplos de mesma classe são projetados muito próximos uns dos outros, e ao mesmo tempo as médias dos conjuntos são colocadas o mais distante possível. Veja a solução na figura 2.11 para o caso mostrado na figura 2.10.

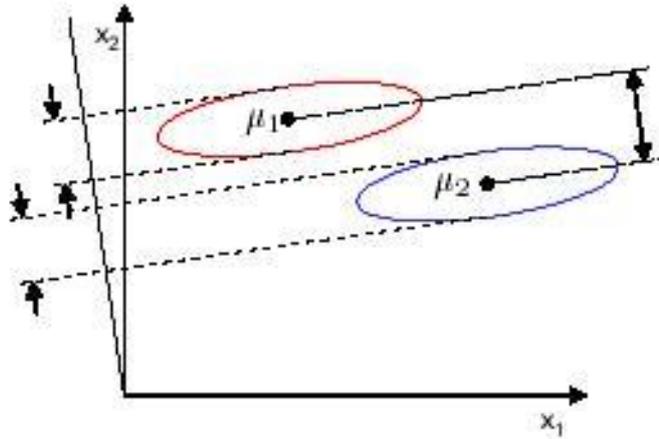


Figura 2.11: Reta ótima para a separabilidade dos conjuntos.

Visando encontrar um vetor W , que defina a projeção ótima é necessário expressar $J(W)$ como uma função explícita de W , então:

$$S_i = \sum_{x \in \omega_i} (x - \mu_i) \cdot (x - \mu_i)^T \quad (2.18)$$

Neste caso um conjunto de dados com duas classes definirá $S_W = S_1 + S_2$

onde S_W é chamado de matriz de dispersão dentro da classe (*within-class scatter matrix*), e a sua projeção no espaço LDA pode ser definida em função das variáveis no espaço original como:

$$\widetilde{S}_i^2 = \sum_{y \in \omega_i} (y - \widetilde{\mu}_i)^2 = \sum_{x \in \omega_i} (W^T \cdot x - W^T \cdot \mu_i)^2 = \sum_{x \in \omega_i} W^T \cdot (x - \mu_i) \cdot (x - \mu_i)^T \cdot W = W^T \cdot S_i \cdot W \quad (2.19)$$

$$\widetilde{S}_1^2 + \widetilde{S}_2^2 = W^T \cdot S_W \cdot W \quad (2.20)$$

Similarmente a diferença entre as projeções da média podem ser definidas como:

$$(\widetilde{\mu}_1 - \widetilde{\mu}_2)^2 = (W^T \cdot \mu_1 - W^T \cdot \mu_2)^2 = W^T \cdot \underbrace{(\mu_1 - \mu_2) \cdot (\mu_1 - \mu_2)^T}_{S_B} \cdot W = W^T \cdot S_B \cdot W \quad (2.21)$$

onde a matriz S_B é chamada de dispersão entre classes (*between-class scatter*). Então finalmente o critério de Fisher pode ser expresso em termos de S_W e S_B como sendo:

$$J(W) = \frac{W^T \cdot S_B \cdot W}{W^T \cdot S_W \cdot W} \quad (2.22)$$

O único critério preponderante para encontrar W é maximizar a função $J(W)$, então derivando, igualando a zero, e efetuando algumas manipulações matemáticas encontra-se:

$$\begin{aligned} \Rightarrow S_B \cdot W - J \cdot S_W \cdot W &= 0 \Rightarrow \\ \Rightarrow S_W^{-1} \cdot S_B \cdot W - J \cdot W &= 0 \end{aligned} \quad (2.23)$$

Onde J é um escalar resultante das manipulações matemáticas relacionadas a função critério $J(W)$ e, finalmente, para encontrar o vetor de projeção deve-se resolver um simples problema de autovetor para encontrar W em:

$$S_w^{-1} \cdot S_b \cdot W = J \cdot W \quad (2.24)$$

Com W encontrado, basta efetuar a projeção das amostras e definir um classificador para designar as classes no espaço LDA.

O problema de reconhecimento de faces utilizando LDA é uma generalização da abordagem 2-D definida acima, pois as amostras estão em um ambiente $n.n$ dimensional e as classes são definidas pela quantidade de pessoas no treinamento. Belhumeur et al. (1997)

declararam que se o conjunto de treinamento está rotulado, faz sentido usar esta informação para construir um método mais seguro na redução de dimensionalidade do espaço de características. Portanto, usando métodos lineares embasados em informações das classes para redução de dimensionalidade e classificadores simples no espaço de características reduzido, pode-se melhorar as taxas de reconhecimento obtidas pelo método de Autofaces. Visualizando este contexto, o método LDA é um exemplo de método específico por classe, e aplicado em faces visa selecionar um conjunto de vetores W para formar um novo espaço vetorial, de modo que a razão da dispersão entre as classes e a dispersão dentro de cada classe seja maximizada.

A matriz de dispersão entre classes é definida como:

$$S_B = \sum_{i=1}^c N_i \cdot (\mu_i - \mu) \cdot (\mu_i - \mu)^T \quad (2.25)$$

E a matriz de dispersão dentro de cada classe é definida como:

$$S_W = \sum_{i=1}^c \sum_{x_k \in X_i} (x_k - \mu_i) \cdot (x_k - \mu_i)^T \quad (2.26)$$

Onde μ é a média de todo o conjunto, μ_i é a média das imagens de classe X_i , e N_i é o número de amostras na classe X_i . Se S_W é não-singular, a projeção ótima W_{otimo} é escolhida como a matriz com colunas ortonormal que maximiza a razão entre o determinante da matriz de espalhamento (entre-classes) com relação ao determinante da matriz de espalhamento (intra-classe) das amostras projetadas, isto é:

$$W_{otimo} = \underbrace{\operatorname{argmax}}_W \left(\frac{W^T \cdot S_B \cdot W}{W^T \cdot S_W \cdot W} \right) = [w_1, w_2 \dots w_m] \quad (2.27)$$

Onde $\{W_i \mid i = 1, 2, 3 \dots m\}$ é o conjunto generalizado de autovetores de S_B e S_W correspondente aos m maiores autovalores $\{\lambda_i \mid i = 1, 2, \dots m\}$, isto é:

$$S_B \cdot W_i = \lambda_i \cdot S_W \cdot W_i \quad i = 1, 2, \dots m \quad (2.28)$$

Note, que haverá no máximo $c - 1$ autovalores não-zero, portanto o limite para m é $c - 1$, onde c é o número de classes. Para maiores explicações teóricas sobre esta limitação presente no LDA com relação aos seus autovalores, veja Duda e Hart (1973).

Para fixar os conhecimentos relativos à técnica LDA é viável ilustrar os benefícios de uma projeção linear específica por classe, quando aplicada a um conjunto de dados, onde se conhecem as classes de treinamento destinadas à classificação. Como um recurso visual, a figura 2.12 é uma comparação entre PCA e LDA para um problema de duas classes em que as amostras de cada classe estão distribuídas randomicamente em uma determinada região do espaço, onde as amostras de cada classe estão próximas a uma linha, que passa pela origem em um espaço característico 2-D. Para este exemplo tem-se $N = 20$, $n = 2$, e $m = 1$. Ambos, PCA e LDA são usados para projetar os pontos de 2-D para um sub-espaço 1-D, que será utilizado para a classificação. Comparando as duas projeções na figura é factual notar que: o PCA mistura as duas classes, então elas não são separáveis no sub-espaço ao longo da projeção 1-D, mas é evidente notar que o LDA alcança uma melhor performance com relação ao espalhamento entre as classes, e conseqüentemente a classificação é simplificada.

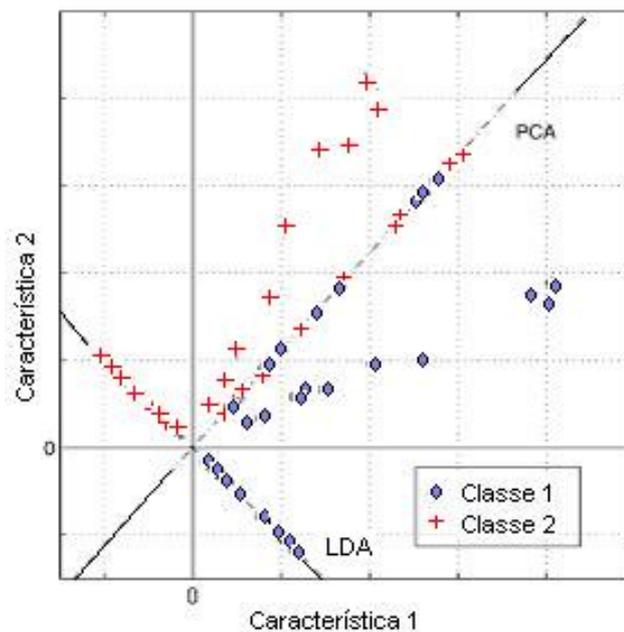


Figura 2.12: Comparação entre PCA e LDA para um problema de duas classe, onde os dados de cada classe são projetados em um subespaço linear.

Enquadrando o problema de reconhecimento de faces com a técnica LDA pode-se dizer que a análise é similar ao método PCA proposto por Pentland et al. (1994), mas a diferença fundamental está presente no cálculo dos autovalores executada através das matrizes de dispersão entre-classes e intra-classe. Lembrando que as faces humanas como padrões em três dimensões (x , y , níveis de cinza) tem muitas coisas em comum e são

espectralmente muito similares, portanto considerando a imagem de face como um todo tem-se alguns aspectos importantes para a discriminação presentes nas características (autovalores) de baixa energia, e estas características de baixa energia podem não ser capturadas em uma redução de dimensionalidade e extração de características baseada em PCA (CHELLAPPA; ETEMAD, 1996).

O método LDA parece ser eficaz para resolver problemas onde se conhecem as classes de treinamento e se deseja maximizar a concentração de energia no sub-espaço, mas no problema de reconhecimento de faces a matriz de dispersão intra-classe $S_W \in R^{n,n}$ é quase sempre singular. Isto origina do fato que o *rank* de S_W é no máximo $N - c$, e em geral, o número de imagens no conjunto de treinamento (N) é bem menor que o número de pixels em cada imagem n , que conseqüentemente formará a dimensão das matrizes de dispersão. Portanto, isto significa que é possível escolher a matriz W em que a matriz de dispersão intra-classe seja exatamente zero, e conseqüentemente W não será uma matriz estável para o sistema (BELHUMEUR et al., 1997).

2.6.4 Sub-espaço LDA.

O Sub-espaço LDA pode ser definido como a união entre PCA e LDA visando aproveitar o potencial e complementar a deficiência de cada um no reconhecimento de faces. O PCA atua no sentido de reduzir a dimensionalidade dos dados sem perder a representatividade das faces e mantendo a generalidade do sistema, enquanto o LDA atua discriminando as classes de treinamento possibilitando uma maior separabilidade entre os dados para o classificador.

No sub-capítulo anterior foi relatada uma deficiência do método LDA, quando aplicado no reconhecimento de faces, pois a matriz S_W pode se tornar singular, e segundo Fisher (1938) é necessário ao menos $t + c$ amostras para garantir, que S_W não se torne singular (que é praticamente impossível em qualquer aplicação real, pois t é a dimensão dos dados originais e c é o número de classes ou pessoas de treinamento). Para resolver este problema, Swets e Weng (1996a)(1996b) e Belhumeur et al. (1997) propuseram o uso de um espaço intermediário. Em ambos os trabalhos, este espaço intermediário foi escolhido como sendo o espaço PCA, portanto o espaço original t -dimensional é projetado em um espaço intermediário g -dimensional usando PCA e posteriormente em um espaço final f -dimensional usando LDA. Esta seqüência de passos é alcançada usando o PCA para reduzir a dimensão do espaço de características para $N - c$, e então aplicando o

LDA padrão reduzindo a dimensionalidade para $c - 1$. Mais formalmente, W_{otimo} é dado por:

$$W_{otimo}^T = W_{lda}^T \cdot W_{pca}^T \quad (2.29)$$

Onde

$$W_{pca} = \underbrace{argmax}_W (W^T \cdot S_T \cdot W) \quad (2.30)$$

$$W_{otimo} = \underbrace{argmax}_W \left(\frac{|W^T W_{pca}^T \cdot S_B \cdot W_{pca} \cdot W|}{|W^T \cdot W_{pca}^T \cdot S_W \cdot W_{pca} \cdot W|} \right) \quad (2.31)$$

Combinando PCA e LDA tem-se uma projeção linear que mapeia a imagem de entrada x em um sub-espço de face y , e posteriormente em um espaço de classificação z :

$$Y = W_{pca} \cdot (x - \Psi)$$

$$Z = W_{lda} \cdot Y$$

$$Z = W_{otimo} \cdot (x - \Psi)$$

Na figura 2.13 é mostrado um diagrama do sistema Sub-espço LDA para o reconhecimento de faces:

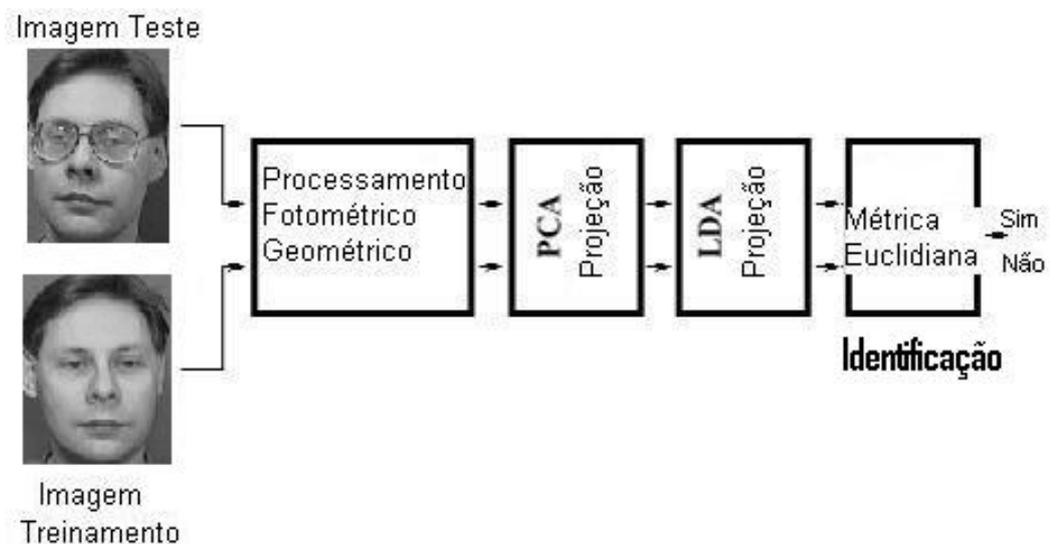


Figura 2.13: Sistema Sub-espço LDA generalizado para o reconhecimento de faces.

A performance do sub-espço LDA pode ser prejudicada se o conjunto de treinamento for pequeno, pois o LDA não terá um conjunto de características semelhantes extraídos das faces da mesma pessoa para realizar a separabilidade. Martinez e Kak (2001) realizaram um vasto trabalho analisando o comportamento dos métodos PCA, PCA (sem os três primeiros autovetores) e Sub-espço LDA, quando aplicados em um conjunto de treinamento pequeno com poucas imagens por classe, e em um conjunto com muitas imagens.

O banco de dados utilizado foi o *AR-face database*, que possui mais de 3200 imagens coloridas de faces frontais de 126 pessoas com 26 imagens diferentes para cada uma (figura 2.14).

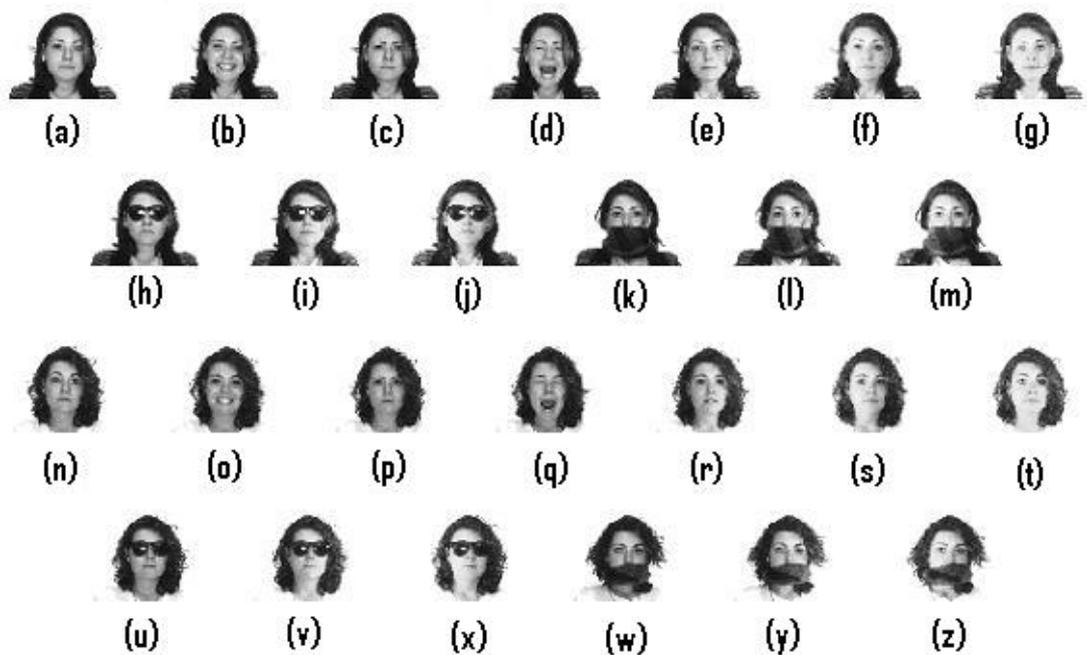


Figura 2.14: Imagens de uma pessoa no *AR face database*. As imagens de (a)-(m) foram tiradas durante uma seção e as imagens de (n)-(z) em uma seção diferente.

Na experiência relatada foram utilizadas 50 pessoas diferentes (25 homens e 25 mulheres). Para analisar a performance dos métodos diante de um grupo de treinamento pequeno foram selecionadas duas imagens de treinamento e cinco para testes entre as imagens de "a" até "g" de cada pessoa, e o resultado é mostrado na figura 2.15a evidenciando a baixa performance do Sub-espço LDA para este caso. Mas quando o conjunto de treinamento é constituído de 13 imagens e as outras 13 utilizadas para teste tem-se como resultado uma alta performance do Sub-espço LDA comparado aos outros dois métodos (veja figura 2.15b).

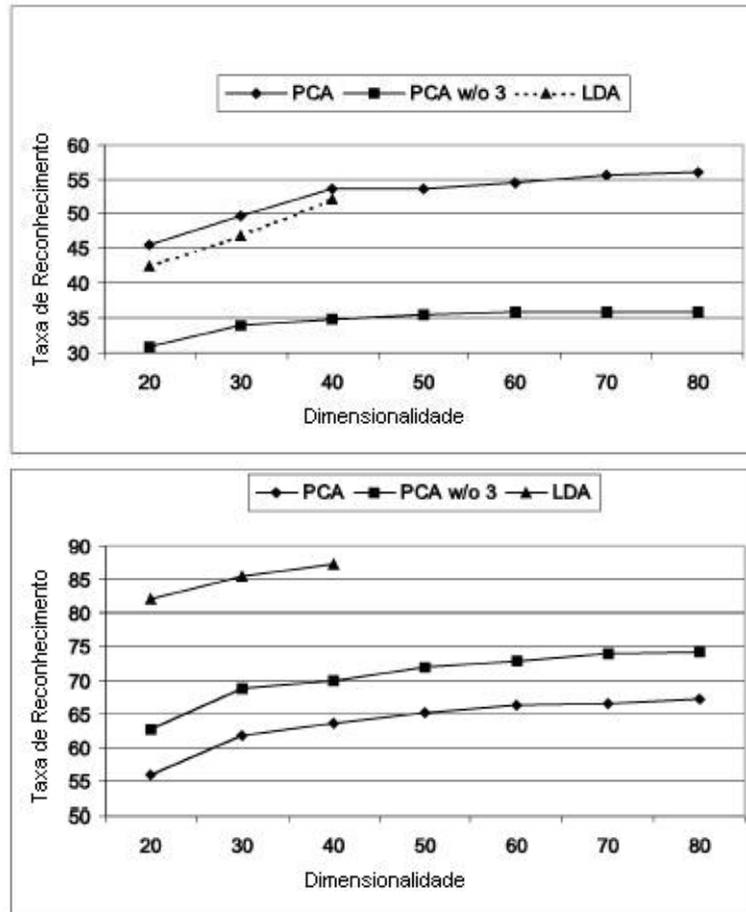


Figura 2.15: Experiência - a) duas imagens de treinamento e cinco para teste b) 13 imagens de treinamento e 13 para teste. Fonte: Martinez e Kak (2001).

Zhao et al. (entre 1995 e 2000) demonstraram também a eficácia do método Sub-espaço LDA na identificação de imagens com ruído evidenciando a generalidade do sistema, (veja figura 2.16).

2.7 Redes Neurais Aplicadas ao Reconhecimento de Faces Humanas.

As aplicações utilizando redes neurais para efetuar o reconhecimento de faces, abordam na maioria dos casos, a extração de características geométricas locais para estabelecer a classificação, ou seja, o vetor característico de entrada da rede é formado por variáveis de distância entre olho, nariz, boca, etc. Mas há alguns métodos onde as redes neurais são

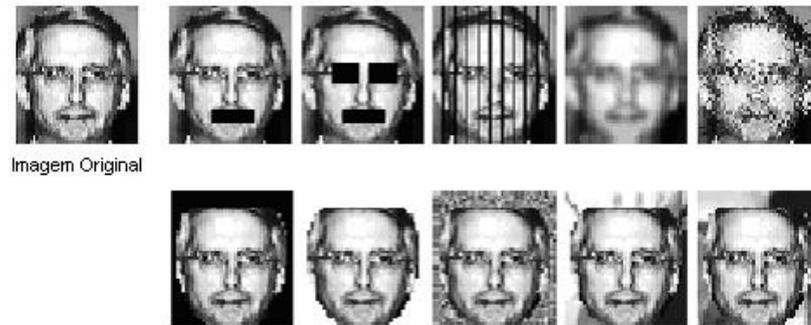


Figura 2.16: Imagens eletronicamente modificadas que foram corretamente identificadas (MARTINEZ; KAK, 2001).

aplicadas de forma holística utilizando as características globais da face. Como exemplo, uma rede neural *perceptron back-propagation* pode ser treinada para reconhecer imagens de faces, portanto uma imagem típica de face requer $N = m.n$ neurônios de entrada, um para cada pixel da imagem, sendo que uma imagem de baixa resolução, com 128×128 pixels, necessita de $N = 16385$ neurônios de entrada. A camada de saída deve ter ao menos um neurônio de saída para cada pessoa, e a camada intermediária (hidden-layer) com p neurônios é a responsável por extrair as características dos dados de entrada, portanto a camada intermediária deve ser escolhida de forma adequada, pois forma um vetor característico comprimido com relação ao vetor de entrada (BARRETT, 1998).

Uma simples rede neural *perceptron back-propagation* aplicada a um banco de dados pequeno pode ser muito complexa e difícil de treinar. Visando reduzir a complexidade Cottrell e Fleming (1990) utilizaram duas redes *perceptron back-propagation*, como ilustradas na figura 2.17. A primeira rede opera no modo auto-associativo descrito por Bourlard e Kamp (1998) extraindo características para a segunda rede, que opera no modo comum de classificação. De forma descritiva, a rede auto-associativa tem n entradas, n saídas, e p nós intermediários (hidden layer) com $p \ll n$. A rede toma o vetor de face x como entrada, e é treinada para obter uma saída y , que é a "melhor aproximação" da entrada x , então a saída da camada intermediária p constitui uma versão comprimida da entrada x , ou um vetor característico que pode ser usado como entrada para a rede de classificação. Mas o sistema continua exigindo uma alta capacidade computacional para bancos de dados médios e grandes, além de não apresentar uma boa generalização, sendo na melhor das hipóteses, igual ao PCA simples.

Um dos grandes desafios a ser superado nas aplicações de redes neurais em reconhecimento de faces é a grande dificuldade em alinhar as características geométricas

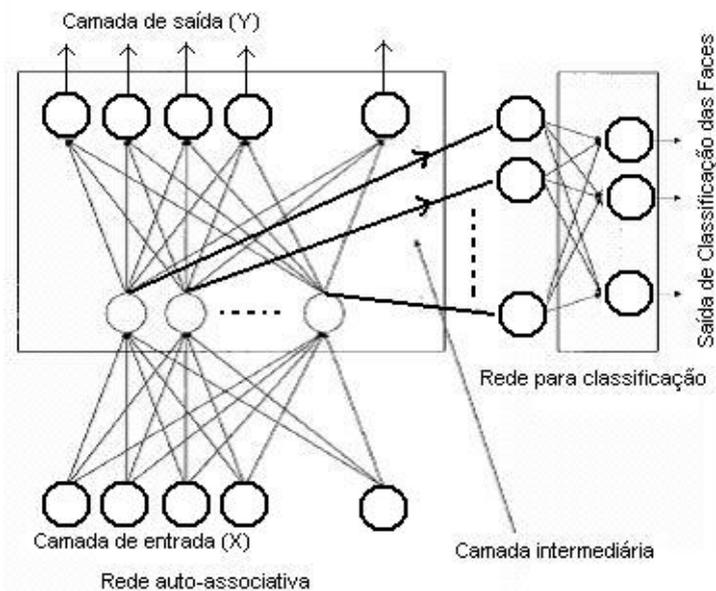


Figura 2.17: Rede de auto-associação e classificativa para reconhecimento de faces. Fonte: Cottrell e Fleming (1990).

das imagens de teste com o banco de dados, pois as redes são altamente dependentes dos valores de cinza dos pixels, portanto muito sensíveis à rotação, translação e escala. Tendo em mente superar estes desafios, Lawrence et al. (1997) apresentaram um sistema, que combinava uma amostragem local da imagem, uma rede neural auto-organizante ou *Self-Organizing Map* (SOM), e uma rede neural de convolução ou *Convolution Neural Network*. A SOM oferece uma quantização das amostras da imagem em um espaço topológico, onde os dados de entrada próximos no espaço original são mapeados próximos no espaço de saída da rede, desse modo, providenciando uma redução de dimensionalidade e invariância com relação às pequenas mudanças nas amostras, e a *Convolution Neural Network* providencia uma invariância parcial para translação, rotação, escala, e deformação.

As imagens de treinamento (96x112) são amostradas com uma janela de 5x5 pixels deslocada sobre a imagem com passos de 4 pixels, (veja figura 2.18), e as sub-amostras das imagens alimentam uma rede SOM de três dimensões com cinco nós por dimensão, perfazendo um total de 125 nós. Agora as imagens de treinamento são novamente sub-amostradas e apresentadas à rede SOM para identificar o neurônio vencedor em cada local de sub-amostra. Dessa forma, cada imagem de entrada é representada por três mapas relacionados com as dimensões da SOM, e cada mapa tem o tamanho de 23x28 (quan-

tidade de locais sub-amostrados na imagem devido ao janelamento 5x5 com passos de 4 pixels). Estes mapas são utilizados para alimentar e treinar a rede de convolução, que aplica a média local para a operação de sub-amostragem e o método *backpropagation gradient-descent* para treinar os neurônios da rede. Os autores descrevem que a rede apresentou erro de 3.83% para o banco de dados da *ORL database*¹ encontrado na *Olivetti Research Laboratory* em Cambridge, U.K., com cinco faces de treinamento e cinco faces de teste para cada pessoa com a estrutura da rede e do sistema como descrito na figura 2.19.

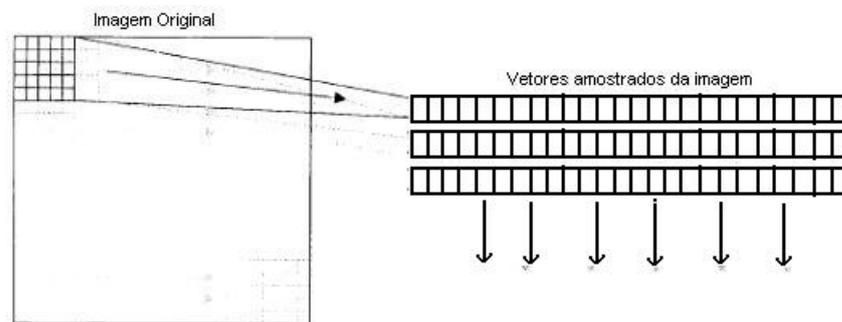


Figura 2.18: Descrição do processo de amostragem local da imagem. Uma janela é deslizada passo-a-passo sobre a imagem e um vetor é criado em cada passo referente a uma localização .

Rizk e Taha (2002) fizeram uma ampla análise com as redes Perceptron Multicamadas utilizando o algoritmo de treinamento *Backpropagation (Multilayer Perceptron Backpropagation Neural Networks - MLP/ BP NN)*, rede neural de Função Base Radial (*Radial Basis Function Neural Networks - RBF NN*), e rede neural de *Cluster Multicamadas (Multilayer Cluster Neural Networks - MCNN)* para aplicações no reconhecimento de faces. As redes neurais foram alimentadas com um vetor característico extraído das imagens através de vários métodos de extração e compressão de dados, eles são: *Discrete Wavelet Transform (DWT)*, *Discrete Radon Transform (DRT)*, *Discrete Cosine Transform (DCT)* e *Principal Component Analysis (PCA)*. O banco de dados utilizado foi o *ORL database* . O teste foi realizado de acordo com as seguintes características (a tabela 2.2 apresenta os resultados):

1. Uma transformada é utilizada para reduzir o espaço de entrada de 10304 (92x112 pixels) para um espaço intermediário reduzido de acordo com o método adotado.
2. Os vetores das imagens obtidos através da transformação são normalizados.

¹Este banco de dados possui 10 imagens diferentes de 40 pessoas

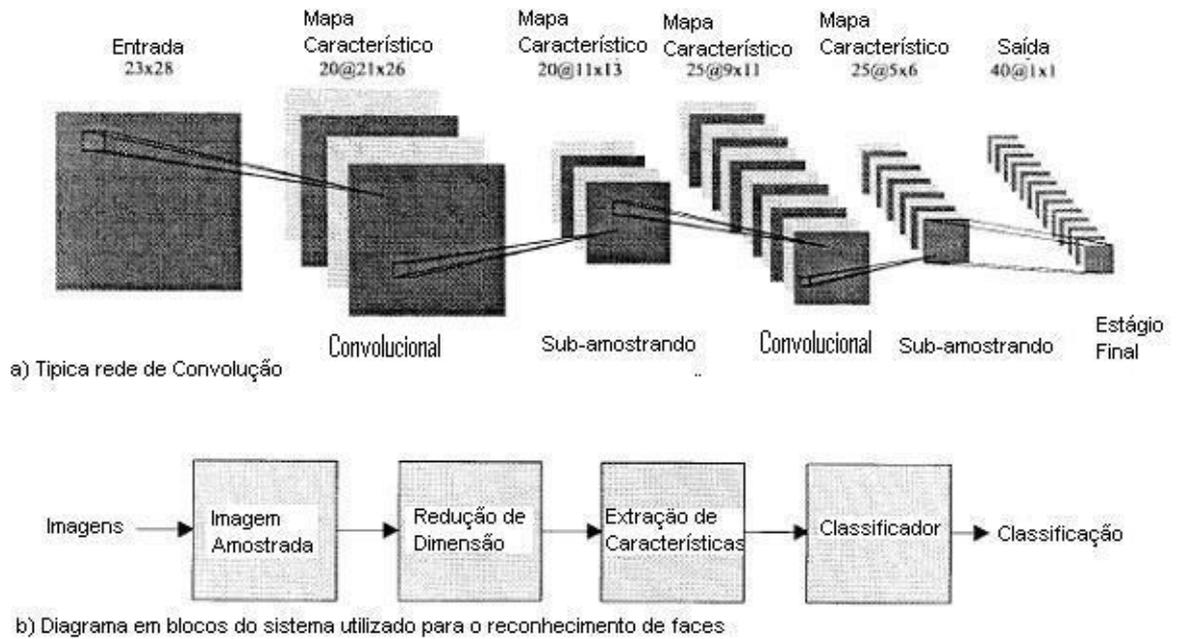


Figura 2.19: Diagrama geral do sistema de reconhecimento de faces.

3. O conjunto de treinamento usado contém 5 amostras por pessoa.
4. O conjunto de testes usado contém 10 amostras por pessoa, ou seja, inclui as amostras de treino também.
5. Classificador MLP/ BP, RBF ou MCNN.

Além das aplicações das redes neurais no reconhecimento de faces existem algumas investigações com relação ao reconhecimento da emoção humana pela expressão facial. Katoh e Fukui (1998) implementaram um conjunto de redes neurais SOM para reconhecer a expressão facial identificando mudanças nos segmentos faciais mais importantes, tais como a sobrancelha, olhos e boca, portanto cada segmento facial alimentava uma rede SOM (veja figura 2.20). As imagens para o treinamento foram tomadas através de uma câmera de vídeo, que registrou a seqüência de seis tipos de expressões faciais de quatro pessoas diferentes (A, B, C e D), os quadros foram selecionados para treinar as redes SOM. Em cada imagem foi definido um retângulo para segmentar as sobrancelhas, os olhos e a boca, e no teste foi avaliado o movimento do neurônio vencedor com relação às expressões faciais efetuadas.

As rede neurais foram elaboradas utilizando a topologia linear com 40 neurônios cada uma, e o resultado encontra-se na tabela 2.3, provando que as redes SOM possuem uma

Tabela 2.2: Desempenho geral das técnicas utilizadas.

Tipo de classificador	Tipo de compressão	Taxa de reconhecimento
MLP/ BP NN	DWT	93.25%
MLP/ BP NN	CI	90.25%
MLP/ BP NN	DRT	93.75%
MLP/ BP NN	DCT	97.50%
MLP/ BP NN	PCA	90.25%
RBF NN	DWT	93.75%
RBF NN	CI	83.75%
RBF NN	DRT	92.75%
RBF NN	DCT	91.0%
RBF NN	PCA	81.0%
Distância Euclidiana	PCA	83.0%
MCNN	DWT	91.0%

boa performance para representar as expressões faciais relacionando o número do neurônio vencedor no mapa topológico. Assim, é possível considerar mudanças nas imagens como mudanças no número do neurônio vencedor.

Tabela 2.3: Resultado da pesquisa (Katoh; Fukui, 1998).

Taxa de Reconhecimento facial					
Expressão facial	Indivíduos				Média
	A	B	C	D	
Feliz	100	100	100	60	90%
Zangado	47	47	53	27	43%
Surpreso	93	80	87	60	80%
Desgosto	67	80	47	67	65%
Medo	20	33	20	0	18%
Triste	13	20	27	13	18%

No trabalho de Sato et al. (2003) uma rede SOM é utilizada como um instrumento secundário, mas importante no reconhecimento de faces, pois ela é aplicada nas imagens como um pré-processamento, definindo o grupo com relação à rotação em que a face se enquadra para posteriormente ser efetuado o reconhecimento com o método LDA treinado com imagens, que apresentam rotações da face próximas a vista frontal. O sistema projetado capta a imagem do individuo em uma sala através de quatro câmeras, e define através da rede SOM qual das imagens é a mais próxima da posição frontal para efetuar o reconhecimento.

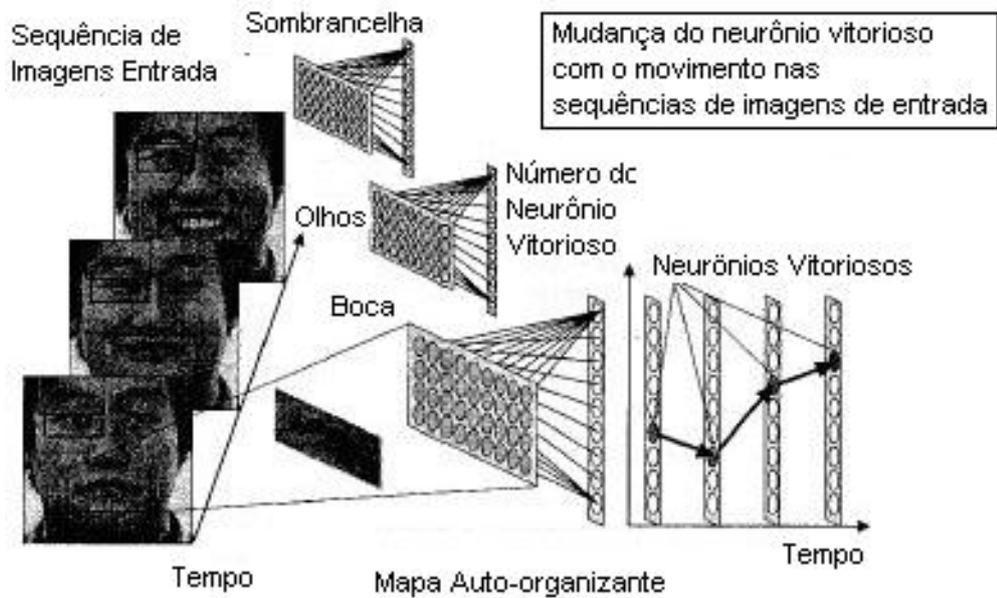


Figura 2.20: Método para extração das características de expressão facial (KATOH; FUKUI, 1998).

2.8 Outros Métodos para o Reconhecimento de Faces.

A área de reconhecimento de faces abrange uma vasta quantidade de conhecimentos com inúmeros algoritmos e métodos, mas a comparação e descrição de vários métodos são na maioria das vezes uma tarefa muito complexa, pois os resultados são obtidos usando diferentes bancos de dados.

2.8.1 Transformada de Fourier (FT).

A Transformada de Fourier é um método muito conhecido nas aplicações de engenharia, que envolvem análises do espectro de frequências de um sinal. Buscando analisar os espectros de frequências, que constituem uma imagem de face, Spies e Ricketts (2000) desenvolveram um sistema de reconhecimento de faces baseado na Transformada de Fourier para comprimir e identificar uma face em um banco de dados visando minimizar os efeitos de translação e rotação presentes no reconhecimento.

O reconhecimento é alcançado encontrando no banco de dados o vetor característico, que contém os coeficientes de Fourier nas frequências selecionadas mais próximos do vetor da imagem teste. Todas as informações globais e locais presentes nos pixels da

face contribuem para formar o espectro de frequências, pois a transformada de Fourier plota as energias das frequências presentes nas imagens, e estas frequências espaciais estão relacionadas com a distribuição espacial dos níveis de cinza na imagem. No caso de reconhecimento de face isto se traduz a distâncias entre as áreas de brilho particular da imagem, como o tamanho global da cabeça, ou à distância dos olhos e a variação destas intensidades na escala de cinza. Frequências mais altas descrevem os detalhes finos da imagem, e os autores relatam que estes são menos úteis para a identificação de uma pessoa. Eles também sugerem que, os humanos podem reconhecer uma face com um breve olhar sem focalizar em detalhes pequenos. Para o experimento prático foi utilizado o banco de dados da ORL com cinco faces de treinamento e cinco de teste para cada pessoa, e foram escolhidas 27 frequências do espectro de Fourier para formar o vetor característico rendendo um resultado de 95% a 98%. Além disso, este pequeno vetor de característica combinado com a Transformada Rápida de Fourier (FFT) faz este sistema melhorar em velocidade.

2.8.2 Transformada Cosseno Discreta (DCT).

Podilchuk e Zang (1996) propuseram um método, que define um vetor característico para as imagens de face usando a Transformada Cosseno Discreta (DCT). O sistema tenta detectar as áreas mais críticas da face relacionadas com o reconhecimento (olhos, boca, nariz, etc) para calcular os coeficientes da DCT e formar um vetor de atributos faciais invariantes associados com as áreas específicas. Um vetor característico é montado para cada pessoa do grupo de treinamento, e os autores relatam que os coeficientes DCT reduzem a dimensionalidade dos dados apresentando um bom desempenho no reconhecimento. Além de trazer a possibilidade de executar o reconhecimento de face diretamente em um "bitstream"(com relação a imagens digitais) comprimido baseado em DCT sem ter que decodificar a imagem. Portanto o método é atrativo, pois os padrões de compressão para foto e vídeo são na sua maioria baseados no método DCT.

2.8.3 Imagens de Perfil.

Em muitas aplicações, as informações de perfil das faces estão disponíveis para serem utilizadas como informação adicional no reconhecimento, como por exemplo: banco de dados da policia. Liposcak e Loncaric (1999) trabalharam com imagens de perfil em vez de imagens frontais. O método está baseado nas informações da face de perfil analisando

a forma do rosto e o formato do cabelo. Para formar o vetor característico, a imagem da face em escala de cinza é binarizada com um *threshold* automático ressaltando a região da face, que é normalizada usando dilatação e erosão. Os autores simularam o perfil de cada indivíduo com cabelo comprido e curto produzindo duas novas imagens para cada um. Portanto essas três imagens de perfil de cada pessoa formam um vetor característico de treinamento, que será utilizado para efetuar o reconhecimento de uma imagem de teste através da medida de distância Euclidiana.

2.9 Considerações Finais.

Nesse capítulo foi descrito em linhas gerais como o sistema visual humano juntamente com o cérebro faz o reconhecimento de faces, além de apresentar todas as variáveis físicas e psicológicas que formalizam o reconhecimento de faces. Existem inúmeras características intrínsecas ao problema de reconhecimento automático de faces, que devem ser levadas em consideração no modelamento dos algoritmos.

Na literatura científica é apresentado 3 abordagens principais para o reconhecimento de faces: métodos holísticos, característicos e híbridos. Nesse capítulo foram discutidos superficialmente os métodos característicos e híbridos, mas foi amplamente discutido o método holístico enfatizando a técnica LDA que embasa todo o projeto proposto nos próximos capítulos. No texto foi apresentado um histórico com relação aos principais trabalhos realizados nos últimos anos aplicando as técnicas holísticas, e alguns trabalhos significantes utilizando redes neurais.

No próximo capítulo serão apresentadas com detalhes as características faciais holísticas e a técnica PCA aplica no reconhecimento de faces, pois a técnica PCA é utilizada no projeto para fazer a compressão inicial dos dados.

Análise de Componentes Principais (PCA - Principal Component Analysis)

3.1 Considerações Iniciais.

Um dos problemas a serem resolvidos nos Sistemas de Visão Computacional está na etapa de reconhecimento de padrões, ou seja, na classificação dos vetores de características de acordo com as classes conhecidas, e é referente a alta dimensionalidade dos dados. Em métodos de reconhecimento de faces, que visam analisar as características globais da face como vetor de características tem-se a presença de um espaço dimensional alto representado pelas informações contidas nos pixels das imagens.

A alta dimensionalidade presente no trabalho de reconhecer faces, mostra-se um ambiente propício para a aplicação da técnica estatística PCA (Principal Component Analysis). A idéia central do método PCA é reduzir a dimensionalidade do conjunto de dados, que possui como característica fundamental conter um grande número de variáveis correlacionadas, enquanto mantém a maioria das variações representativas presentes. Isto

é alcançado transformando os dados em um novo conjunto de variáveis, os *Principal Components* (PCs), que são descorrelacionados e ordenados de forma que os primeiros mantenham a maioria das variações presentes em todo o conjunto de variáveis originais (JOLLIFFE, 1986).

PCA é essencialmente uma técnica estatística, que transforma linearmente um conjunto original de variáveis em um outro conjunto substancialmente menor, onde as novas variáveis são os coeficientes relacionados à projeção das variáveis originais no sub-espço representado pelos PCs derivados da matriz de covariância do conjunto original. Um conjunto menor de variáveis descorrelacionadas é muito mais fácil de se entender e analisar suas interações do que um conjunto com um grande número de variáveis correlacionadas. Esta técnica foi originalmente concebida por Pearson (1901) e independentemente desenvolvida por Hotelling (1933), mas somente com o advento da computação tornou-se popular com livros e artigos citando suas aplicações na agricultura, biologia, química, climatologia, demografia, ecologia, economia, geologia, meteorologia, psicologia, controle de qualidade e finalmente visão computacional (DUNTEMAN, 1989).

A compressão das imagens de face é efetuada aplicando o método PCA no desenvolvimento do algoritmo proposto por esse trabalho para realizar o reconhecimento de faces. Esse capítulo descreve com detalhes a aplicação do método PCA (mais conhecido como eigenface) na redução de dimensionalidade dos dados referentes à caracterização das faces humanas. Alguns conceitos abordados nesse capítulo já foram descritos sutilmente no capítulo 2, mas nesse momento é necessário detalhar os conceitos matemáticos e a estrutura do algoritmo aplicado na compressão e classificação das faces para que no capítulo 5 seja possível apresentar o nosso algoritmo com os resultados obtidos.

3.2 Características das Imagens Faciais.

As imagens de faces humanas presentes na maioria dos bancos de dados estão em escala de cinza e verticalmente orientadas para a visão frontal. Analisando-se as imagens em um gráfico 3-D com as coordenadas x e y sendo a referência espacial de cada pixel, e z representando o valor em escala de cinza de cada pixel, pode-se notar que todas as imagens são localizadas em um *cluster* muito estreito no espaço de imagens caracterizados pelos eixos x , y e z , como mostra a figura 3.1. Mas é necessário concluir que qualquer modificação na iluminação, rotação, translação, ou escala de uma das faces presentes no

banco de dados pode excluí-la do cluster, levando-a a ter sua representação em uma outra região do espaço compreendido pelos eixos x , y e z .

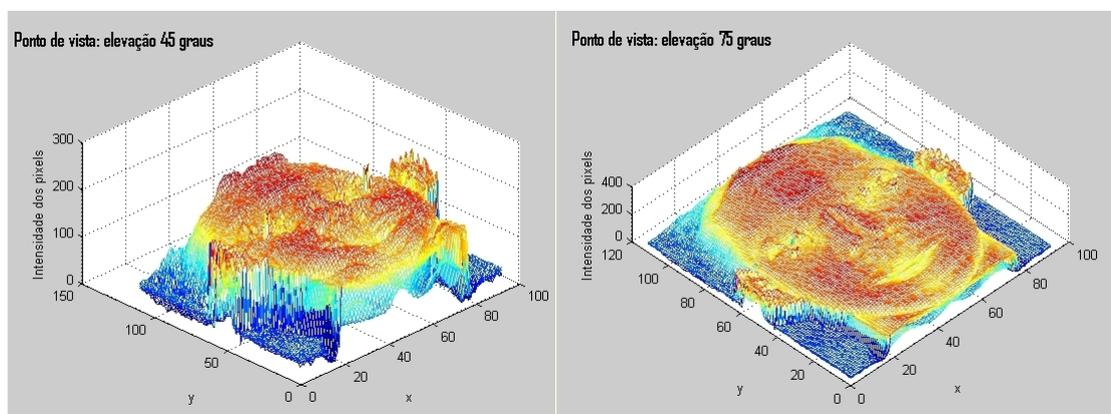


Figura 3.1: *Cluster* de Face no Espaço de Imagem (cor de fundo preto).

Globalmente, todas as faces se parecem umas com as outras, pois todas elas têm dois olhos, uma boca e um nariz. Além disso, todos estes elementos característicos de uma face são localizados relativamente no mesmo lugar no espaço de faces em x , y e z , portanto apenas pequenas variâncias locais distinguem uma face das outras. Mas o espaço de faces não é um ótimo espaço para descrição das faces devido a sua alta dimensionalidade. Para resolver este problema Sirovich e Kirby (1987) trataram a imagem de face como sendo um vetor de características altamente dimensional, onde cada pixel seria mapeado como um componente no vetor, e conseqüentemente formalizando um ambiente propício para a aplicação da técnica PCA na redução da dimensionalidade, pois os vetores de características de cada face são correlacionados uns com os outros.

3.3 Identificação de Faces Utilizando PCA.

A aplicação da técnica PCA na identificação de faces permite encontrar vetores que melhor estimam a distribuição das imagens de faces no espaço de imagens definido pelo vetor de características que representa os pixels da imagem. Portanto, a técnica *Eigenfaces* (Autofaces) não relata as características geométricas da face (posição e tamanho dos olhos, boca, nariz, etc) individualmente, pois efetua uma análise estatística nas imagens ressaltando a correlação presente entre os níveis de cinza dos pixels sem discriminar sua posição, ou representação na imagem.

Um dos grandes desafios envolvidos no reconhecimento de faces é descobrir quais aspectos da face são estímulos importantes para a identificação, e como funciona o sistema cognitivo humano. Turk e Pentland (1991a) sugerem que uma implementação teórica para codificação e decodificação das imagens de faces deve dar introspecção nas informações contidas nas imagens de faces, enfatizando os traços significantes locais e globais do rosto. Teoricamente é desejável extrair as informações relevantes contidas na imagem da face, codificando a imagem de forma eficiente, e comparando uma imagem codificada com um banco de dados similar. Uma implementação básica pode basear-se na captura da variação dos níveis de cinza em uma coleção de imagens de face, independente de qualquer julgamento com relação aos traços do rosto.

Em condições matemáticas, deseja-se achar os componentes principais da distribuição das faces, ou seja, os autovetores da matriz de covariância do conjunto de imagens de faces. Esses autovetores podem ser pensados como um conjunto de características que juntos representam as variações entre as imagens de face e a imagem média do conjunto. Cada local da imagem contribui mais ou menos a cada autovetor, de forma que se pode exibir o autovetor como um tipo de face fantasmagórica que recebe o nome de *eigenface* (autoface). Cada imagem de face no conjunto de treinamento pode ser exatamente representada em termos de uma combinação linear dos autovetores. O número de possíveis autovetores representativos é igual ou menor ao número de imagens de faces no conjunto de treinamento. Porém as faces também podem ser aproximadas usando-se somente os "melhores" autovetores: e estes estão associados aos maiores autovalores, os quais, representam os maiores valores de discrepância dentro do conjunto de imagens de faces. A razão primária para se usar menos autovetores é a eficiência computacional (TURK; PENTLAND, 1991b).

A meta é representar uma imagem de face em termos de um sistema de coordenadas ótimas. Dentre as propriedades ótimas destaca-se o fato de que o erro-quadrático-médio devido ao truncamento da expansão de autofaces (autovetores) é um mínimo global. E os vetores que estabelecem o sistema de coordenadas são os autovetores definidos em função da matriz de covariância do conjunto de faces (TURK; PENTLAND, 1991a).

Geometricamente, a primeira componente principal (ou primeiro autovetor) é uma reta, que melhor se aproxima da distribuição dos vetores de características das imagens no espaço de dimensão d (equivalente à quantidade de pixels da imagem) visando o mínimo erro-quadrático-médio entre os vetores (que representam um ponto no espaço) e a reta. A segunda componente principal é um plano, que melhor se aproxima da distribuição dos vetores de características no espaço d dimensional. Equivalentemente a d principal

componente é o próprio espaço dimensional caracterizado por um hiperplano, que define um sistema de coordenadas ortogonais, onde todos os vetores de características estão contidos (DUNTEMAN, 1989).

Visando compreender melhor as características intrínsecas ao método de Autofaces são apresentados na figura 3.2 amostras de imagens com dois pixels correlacionados, onde a primeira componente é um novo eixo de coordenadas no espaço das variáveis, e é orientada na direção que maximiza a variação das projeções (linha perpendicular ao eixo) dos pontos no novo eixo de coordenadas. A segunda componente está orientada na direção da segunda maior variação das projeções dos pontos, é ortogonal à primeira e as duas principais componentes formam um plano, que contém todos os pontos.

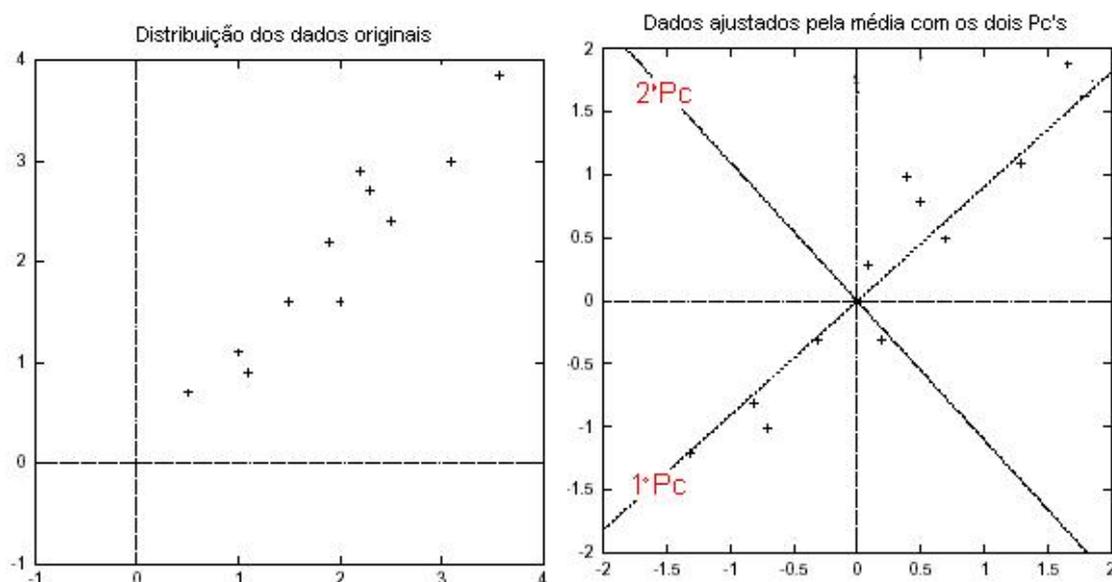


Figura 3.2: a) Exemplo de um conjunto de dados correlacionados, b) Dados normalizados e seus autovetores

Cada autovetor desvia-se da distribuição uniforme de cinzas da imagem, onde alguma característica facial difere entre o conjunto de faces de treinamento; eles são um mapa ordenado das variações entre as faces (LEE et al., 1999). Então um conjunto de imagens faciais pode ser aproximadamente reconstruído armazenando-se uma coleção pequena de pesos para cada face e um conjunto pequeno de imagens "padrões" (*eigenfaces*). Portanto, se uma grande variedade de imagens de faces pode ser reconstruída por somas de pesos de uma coleção pequena de imagens "padrões", então um modo eficiente para armazenar e reconhecer faces poderia ser baseado em construir as características de imagens de faces conhecidas e reconhecer faces particulares através da comparação dos pesos carac-

terísticos necessários (aproximadamente) para reconstruí-los com os pesos associados aos indivíduos conhecidos (TURK; PENTLAND, 1991b).

Os seguintes passos resumem o processo de reconhecimento (TURK; PENTLAND, 1991b):

- **Treinamento**

1. Adquira o conjunto de imagens de faces para o treinamento.
2. Calcule as autofaces (ou autovetores) do conjunto de treinamento mantendo somente os M autovetores que correspondem aos mais altos autovalores. Esses M autovetores definem o *face space* (sub-espço de faces), e quando uma nova face é inserida faz-se necessário recalculá-lo.
3. Calcule a correspondente distribuição dos pesos no espaço M -dimensional para cada indivíduo conhecido através da projeção de suas imagens faciais (vetor de características) no sub-espço de faces.
4. Armazenar os autovetores, e os pesos correspondentes a cada face do conjunto de treinamento.

Esta operação exige um alto custo computacional, mas é efetuada apenas no treinamento.

- **Reconhecendo Novas Faces**

1. Calcule o conjunto de pesos que representam a imagem de entrada no sub-espço de faces projetando a imagem em cada um dos M autovetores.
2. Determine se a imagem é uma face observando se a sua projeção no sub-espço de faces se aproxima dos pontos que representam as faces de treinamento.
3. Se a imagem é uma face, classifique os seus padrões de pesos como uma face conhecida ou desconhecida de acordo com as faces de treinamento.
4. (Opcional) Atualize as autofaces e os pesos recalculando o sub-espço de faces para um conjunto de treinamento com as novas faces inseridas.
5. (Opcional) Se a mesma face desconhecida é apresentada várias vezes ao sistema, calcule os seus padrões de pesos e insira-os como face conhecida.

3.4 Calculando as Autofaces.

O cálculo das autofaces (ou autovetores) tem como bases teóricas os conceitos fundamentais presentes na técnica estatística PCA aplicada na redução de dimensionalidade de um conjunto de dados. Para apresentar a técnica PCA aplicada às imagens de faces serão tomadas como referência os desenvolvimentos matemáticos sustentados por Turk e Pentland (1991b) para a formação do sub-espaço de faces caracterizado pelas autofaces.

Embasado nos conceitos de Visão Computacional, a imagem de uma face, pode ser visualizada como uma matriz com largura w e altura h formada pelos seus pixels, que carregam um valor de intensidade na escala de cinzas da imagem em função da sua posição x e y . Cada imagem $I(x, y)$ é uma matriz bi-dimensional ($w.h$) de intensidades na escala de cinzas, portanto uma imagem pode ser considerada um vetor de dimensão $w.h$ executando uma simples concatenação das linhas ou colunas da imagem (figura 3.3). Para uma imagem de 256×256 pixels o vetor terá 65.536 elementos, ou seja, equivale a um ponto contido em um espaço de dimensão 65.536.

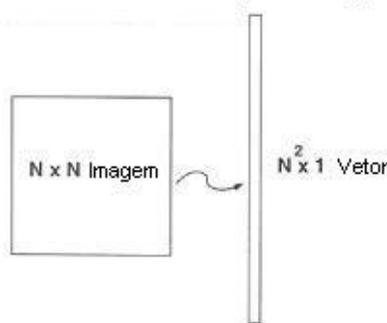


Figura 3.3: Caracterização do vetor imagem.

Como visto anteriormente, as imagens de faces apresentam características semelhantes na sua forma global (todas têm boca, nariz, olhos, etc) e os pontos formados pelos vetores não são aleatoriamente distribuídos no espaço dimensional, mas sim, agrupados em um *cluster*. Portanto esta análise sugere, que os pixels da face devam manter uma forte correlação (ou seja, os valores de intensidade referentes às M imagens distribuídos ao longo da dimensão referente a um pixel, devem manter correlação com os seus pixels vizinhos, que fazem referência a outras dimensões).

A idéia principal da técnica PCA é achar os vetores de imagens que melhor representam a distribuição de imagens de face dentro do espaço de imagens originais inteiro. Esses vetores definem um sub-espaço de imagens de faces. Cada vetor de comprimento

$w.h$, descreve uma imagem $w.h$, e é uma representação da variância das faces originais com relação à face média (figura 3.4).

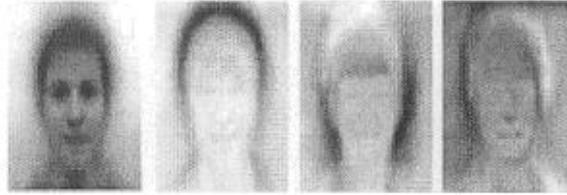


Figura 3.4: Autovetores - exemplo de bases do sub-espaco de autofaces.

Os vetores $\Gamma_1, \Gamma_2, \Gamma_3, \dots, \Gamma_M$ serão associados aos pixels das imagens $Im(x, y)$ concatenados por linhas formando vetores $p.p$ (ou $w.h$) concatenados por linhas, e utilizados no treinamento do sistema. Visando encontrar a matriz de covariância deve-se lembrar que: as dimensões envolvidas devem ter o mesmo número de dados, que devem ser padronizados com média igual a "zero" e variância igual a "um" para cada dimensão. Para padronizar os dados deve-se calcular a média das imagens como sendo:

$$\Psi = \frac{1}{M} \cdot \sum_{n=1}^M \Gamma_n \quad (3.1)$$

Através do valor definido pela média, são calculados os desvios de cada imagem com relação à média:

$$\phi_n = \Gamma_n - \Psi \quad (3.2)$$

O conjunto de vetores definido por ϕ está normalizado com relação a média, e pronto para ser utilizado no cálculo dos autovetores restritos pelas condições dadas nas equações 3.3e 3.4, que encontrará um conjunto de M vetores ortonormais, μ_n , que melhor descreve a distribuição dos dados. Então o k -ésimo vetor, μ_k é escolhido como sendo:

$$\lambda_k = \frac{1}{M} \cdot \sum_{n=1}^M (\mu_k^T \cdot \phi_n)^2 \quad (3.3)$$

Onde λ_k é um máximo, sujeito à seguinte restrição:

$$\mu_l^T \mu_k = \begin{cases} 1, & l=k \\ 0, & \text{caso contrário} \end{cases} \quad (3.4)$$

Os vetores μ_k e escalares λ_k são autovetores e autovalores, respectivamente, da matriz de covariância

$$C = \frac{1}{M} \cdot \sum_{n=1}^M \phi_n \phi_n^T = A \cdot A^T \quad (3.5)$$

onde a matriz $A = [\phi_1, \phi_2, \dots, \phi_M]$.

A matriz C , no entanto, é $N^2 \cdot N^2$, e determina os N^2 autovetores e autovalores, mas é uma tarefa intratável para o tamanho das imagens típicas. Um método computacionalmente viável é necessário para encontrar esses autovetores. Se o número de pontos dos dados no espaço de imagens (ou seja, o número de imagens) for menor que a dimensão do espaço ($M < N^2$), haverá somente $M - 1$, um pouco menos que N^2 , autovetores significantes (o restante dos autovetores serão associados aos autovalores zero). Felizmente, há um método algébrico para resolver este problema N^2 -dimensional. Neste caso resolvendo para os primeiros autovetores da matriz M por M referente a $A^T \cdot A$.

Considere os autovetores V_n de $A^T \cdot A$ semelhantes a:

$$A^T \cdot A \cdot V_n = \lambda_n \cdot V_n \quad (3.6)$$

multiplicando ambos os lados por A , obtém-se:

$$A \cdot A^T \cdot A \cdot V_n = \lambda_n \cdot A \cdot V_n \quad (3.7)$$

logo, conclui-se que $A \cdot V_n$ são os autovetores de $C = A \cdot A^T$

Seguindo esta análise, constrói-se a matriz $L = A^T \cdot A$ (M por M), onde $L_{m,n} = \Phi_m^T \cdot \Phi_n$, e encontra-se os M autovetores V_n de L^1 . Estes vetores determinam uma combinação linear das M imagens faces do conjunto de treinamento para formar os autovetores μ_n :

$$\mu_n = \sum_{k=1}^M V_{nk} \phi_k = A \cdot V_n \quad n = 1, 2, \dots, M \quad (3.8)$$

Com esta análise os cálculos são reduzidos, da ordem dos números de pixels nas imagens (N^2) para a ordem do número de imagens no conjunto de treinamento (M). Na

¹ L e C são matrizes simétricas não negativas

3.5. PROJETANDO AS FACES DE TREINAMENTO NO SUB-ESPAÇO DE FACES

prática, o conjunto de faces de treinamento será relativamente pequeno ($M < N^2$), e o cálculo torna-se completamente gerenciável. Os autovalores associados permitem-nos colocar em ordem os autovetores de acordo com a sua utilidade na caracterização da variação entre as imagens, pois os primeiros autovalores possuem valores altos caracterizando a sua representatividade com relação aos dados.

3.5 Projetando as Faces de Treinamento no Sub-espaço de Faces.

Uma vez criado o sub-espaço de faces tendo como bases os autovetores, a identificação de cada face no espaço torna-se uma tarefa padrão, onde primeiramente deve-se escolher de forma heurística os primeiros autovalores, que possuem valores numericamente expressivos, associados aos primeiros autovetores, e conseqüentemente estão relacionados às maiores variâncias no conjunto de dados. Dessa forma, diminuindo-se o sub-espaço de faces de M para M' ($M' < M$), o custo computacional é reduzido como conseqüência, e desde que a reconstrução da imagem não seja importante, M' deve ser suficiente para efetuar apenas a identificação. Sendo assim, as faces de treinamento são transformadas em combinações lineares dos vetores bases (autovetores) do sub-espaço de faces (ou espaço autofaces) associados a coeficientes (veja figura 3.5), que são calculados projetando-as no sub-espaço definido pelas autofaces, do seguinte modo:

$$\omega_n = \mu_n^T \cdot (\Gamma - \Psi) \quad \text{para } n = 1, \dots, M \quad (3.9)$$



Figura 3.5: Exemplo de face formada pela combinação linear da base autofaces.

Portanto os ω_n definem um vetor de pesos $\Omega^T(i) = [\omega_1^i, \omega_2^i, \dots, \omega_n^i]$, $i = 1, \dots, M$ (onde M é a quantidade de imagens), que expressa cada face de treinamento (i) no sub-espaço definido pelos autovetores. Este vetor Ω^T descreve a contribuição de cada autovetor (autoface) na representação da imagem.

Para efetuar a reconstrução da imagem segundo Sirovich e Kirby (1987) basta efetuar alguns cálculos de multiplicação e soma entre os pesos ω_n e os autovetores como mostrado na equação 3.10.

$$\phi(i) \simeq \Psi + \sum_{n=1}^{M'} \omega_n(i) \cdot \mu^n = \phi^{M'}(i) \quad (3.10)$$

A reconstrução perfeita da imagem pode ser obtida através da equação 3.11:

$$\phi(i) = \Psi + \sum_{n=1}^M \omega_n(i) \cdot \mu^n \quad (3.11)$$

O erro no truncamento de M para M' é:

$$\varepsilon_{M'}(i) = \frac{\|\phi(i) - \phi^{M'}(i)\|}{\|\phi(i)\|} \quad (3.12)$$

Completando a análise tem-se a fração da variância total do conjunto presente nos primeiros M' autovalores dado por:

$$q_{M'} = \frac{\sum_{i=1}^{M'} \lambda^i}{\sum_{i=1}^M \lambda^i} \quad (3.13)$$

3.6 Classificação e Identificação de uma Nova Imagem.

Quando uma nova imagem é apresentada ao sistema deve-se encontrar os pesos ω que a representam no sub-espço de faces. Isto é conseguido projetando o vetor Γ_{nova} no sub-espço de faces:

$$\omega_n = \mu_n^T \cdot (\Gamma_{nova} - \Psi) \quad \text{para } n = 1, \dots, M' \quad (3.14)$$

Os pesos formam um vetor $\Omega^T(nova) = [\omega_1, \omega_2, \dots, \omega_n]$, que descreve a contribuição de cada autoface na representação da imagem de entrada. Portanto, com os vetores de pesos associados a cada imagem do conjunto de treinamento, e mais o vetor de pesos da nova imagem representando as faces em um sistema de coordenadas ótimo, pode-se utilizar qualquer algoritmo de reconhecimento de padrões para descrever as classes re-

ferentes às faces. Assim, uma face de entrada pode ser identificada com alguma classe de treinamento (onde cada pessoa é uma classe), ou como uma face desconhecida, ou simplesmente como uma imagem não face.

Adotando-se como classificador a distância Euclidiana pode-se exemplificar a atuação do sistema de forma simples. Para encontrar a classe mais semelhante à imagem de entrada é necessário encontrar a classe que possui a menor distância Euclidiana da imagem de entrada.

$$\varepsilon_k^2 = \|(\Omega - \Omega_k)\|^2 \quad (3.15)$$

Este cálculo é realizado para as k-ésimas classes onde $\Omega_k(medio)$ é o calculo da média realizado entre os vetores pesos de cada classe, ou seja, é um centróide para cada classe. Então, qualquer imagem de entrada do sistema é projetada sobre o sub-espço definido pelos autovetores, e de acordo com a medida de distância entre o vetor de sua projeção e o banco de dados tem-se uma classe definida. Portanto, para qualquer imagem de objeto não face haverá uma classe associada gerando erro no sistema, mas isto não é um problema, pois pode-se calcular a distância entre a imagem de entrada e o espaço de face original através do quadrado da distância entre a imagem de entrada ajustada pela média ($\phi = \Gamma - \Psi$) e a sua imagem recuperada após a projeção ($\phi_f = \sum_{n=1}^{M'} \omega_n \cdot \mu^n$), isto é:

$$\varepsilon_{fespace}^2 = \|(\phi_{nova} - \phi_f)\|^2 \quad (3.16)$$

Assim, de acordo com as análises apresentadas acima, uma imagem de entrada e seu vetor de pesos pode ser classificada em quatro possibilidades (ver figura 3.6):

- Próxima ao sub-espço de faces e próxima a uma classe: a imagem é uma face e o individuo é identificado.
- Próxima ao sub-espço de faces e distante de uma classe conhecida: a imagem é uma face, mas não é um individuo apresentado no treinamento ou conhecido.
- Distante do sub-espço de faces e próximo a uma classe: a imagem não é uma face.
- Distante do sub-espço de faces e distante de uma classe: a imagem não é uma face.

Os *thresholds* que definem a máxima distância permitida de qualquer classe, e o que define a máxima distância permitida do sub-espço de faces devem ser escolhidos cuidadosamente para equilibrar as porcentagens de falso-positivo e falso-negativo.

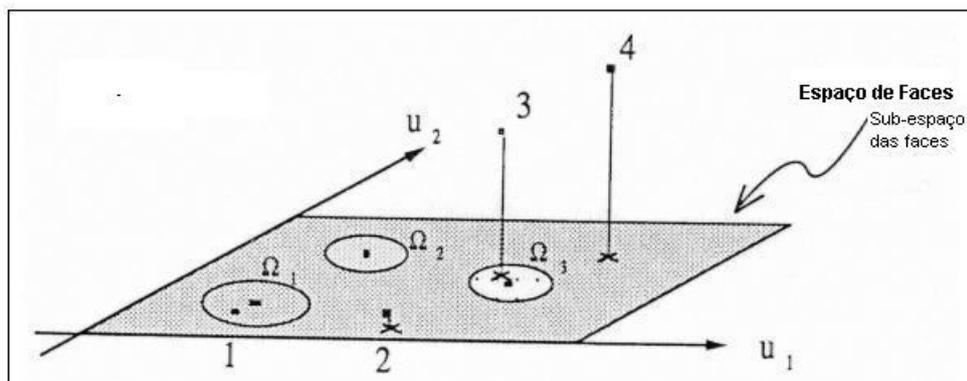


Figura 3.6: Versão simplificada do espaço de faces ilustrando as quatro possibilidades para a projeção de uma imagem no espaço de faces. Neste caso, há dois autovetores (μ_1 e μ_2) e três indivíduos conhecidos (Ω_1 , Ω_2 e Ω_3).

O algoritmo autofaces também mostra-se muito poderoso na reconstrução de imagens de faces, encontrando a face do banco de dados que mais se assemelha à imagem de entrada (veja figura 3.7). Esta aplicação pode ser muito interessante na identificação de suspeitos em um banco de dados policial.



Figura 3.7: Teste com o PCA efetuando recuperação de informação.

3.7 Considerações Finais.

Nesse capítulo foram ressaltadas as características holísticas das imagens faciais, e foi descrito o desenvolvimento matemático do algoritmo PCA aplicado ao reconhecimento de faces. O PCA é uma ótima ferramenta para efetuar a compressão do espaço de imagens diminuindo o processamento computacional na etapa de classificação e identificação.

No próximo capítulo serão descritas as redes neurais PCA, LDA e SOM (Kohonen) utilizadas no desenvolvimento do projeto.

Redes Neurais para o Reconhecimento de Padrões

4.1 Considerações Iniciais.

Neste capítulo é apresentada uma abordagem geral sobre redes neurais artificiais aplicadas em reconhecimento de padrões.

As redes neurais estão se expandindo em aplicações científicas, e três técnicas muito conhecidas para elaborar o reconhecimento e extração de características de padrões, que são PCA, LDA e Kohonen estão sendo muito empregadas, portanto, este capítulo enfoca estas três técnicas relacionando-as com as vantagens e desvantagens quando aplicadas no reconhecimento de faces.

4.2 Redes Neurais Artificiais (RNA).

A natureza sempre foi um elemento de inspiração para a criação humana. As pesquisas em redes neurais foram motivadas pelos estudos físicos e analíticos do comportamento ce-

rebral no reconhecimento e processamento de informações captadas pelos sentidos, pois o cérebro diferentemente do computador digital, processa a informação de uma forma altamente complexa, não-linear e paralela. O cérebro é constituído por neurônios, que transmitem informações eletroquímicas entre si para realizar certos processamentos (análises de imagens e áudio, percepção, reconhecimento de padrões e controle motor) e responder aos estímulos com velocidades superiores aos computadores digitais existentes.

O reconhecimento de padrões em visão computacional é um trabalho árduo, que exige muito processamento de máquina, têm respostas lentas, e muitas vezes específicos na aplicação, portanto não superam as qualidades cerebrais no processamento genérico dos estímulos. Considere, por exemplo, a visão humana, que é uma tarefa de processamento de informação genérica e robusta com relação às aplicações e modificações do ambiente. A função do sistema visual é fornecer uma representação do ambiente à nossa volta e, mais importante que isso, fornecer a informação que necessitamos para interagir com ele. Sendo um pouco mais específico, o cérebro realiza rotineiramente tarefas de reconhecimento perceptivo (reconhecer um rosto familiar presente em uma cena não-familiar, ou armazenar uma nova face) em aproximadamente 200ms, no entanto esta tarefa é complexa para ser executada em um computador convencional, mas o computador ganha na quantidade de informação armazenada (MARR, 1982)(LEVINE, 1985)(CHURCHLAND; SEJNOWSKI, 1992).

Segundo Haykin (1998), uma rede neural é um processador paralelamente distribuído constituído de unidades de processamento simples, que têm a propensão natural para armazenar conhecimento experimental e torná-lo disponível para o uso. Ela se assemelha ao cérebro em dois aspectos:

- O conhecimento é adquirido pela rede a partir de seu ambiente através de um processo de aprendizagem.
- Forças de conexão entre neurônios, conhecidas como pesos sinápticos, são utilizadas para armazenar o conhecimento adquirido.

Portanto, uma rede neural baseada nos sistemas neurais biológicos pode ser definida como generalizações de modelos matemáticos baseados nas seguintes suposições:

- Os neurônios processam as informações de forma simples e individual, mas com ligações entre as saídas e entradas dos neurônios vizinhos.
- As conexões de ligação transmitem as informações (sinais) entre os neurônios.

- Cada conexão possui um peso a ela associado.
- O sinal de saída de cada neurônio é determinado pela soma ponderada das suas entradas, que passa por uma função de ativação (quase sempre não linear) pré-definida.

A forma de um neurônio da rede neural (figura 4.2) é parecida com o neurônio celular, figura 4.1, e existem muitas arquiteturas de redes neurais. Conseqüentemente, uma rede neural é caracterizada por seu padrão de conexão entre os neurônios (arquitetura ou topologia da rede), seu método para determinar os pesos das conexões (algoritmo de treinamento ou aprendizagem) e pela função de ativação de seus neurônios.

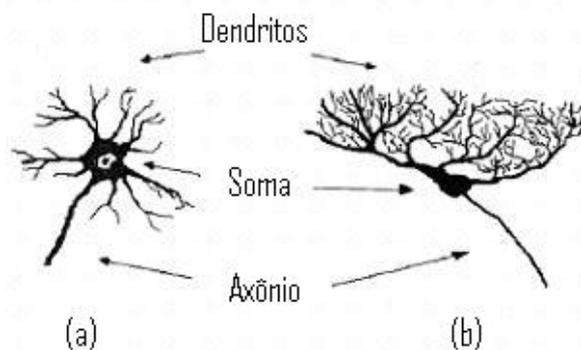


Figura 4.1: Neurônios do sistema nervoso central dos vertebrados: (a) Neurônio motor da célula espinhal; (b) Célula de Purkinje encontrada no cerebelo. Esta célula é notável pela extensa ramificação da sua árvore dendrital, da qual apenas uma pequena parte é mostrada na figura, podendo ultrapassar várias dezenas de vezes as dimensões do corpo celular (HAYKIN, 1998).

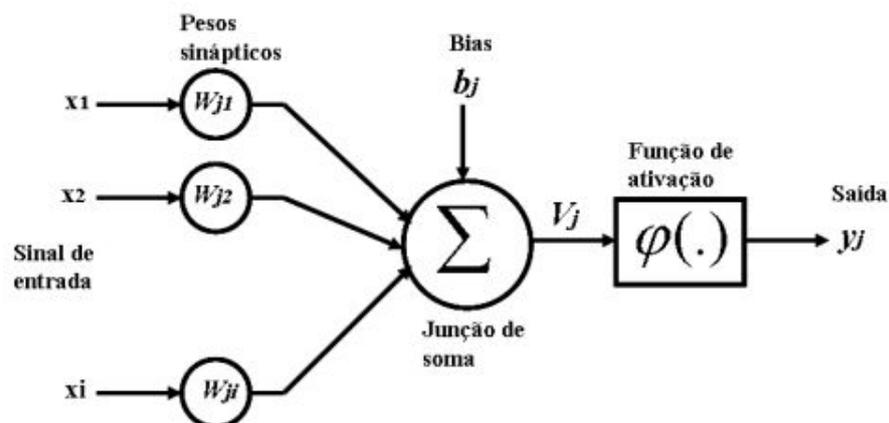


Figura 4.2: Modelo matemático não-linear de um neurônio.

As aplicações em reconhecimento de faces necessitam de sistemas que tenham alta capacidade de processamento com compressão e diminuição das redundâncias presentes

nas informações de imagem digital, além de fácil adaptabilidade às mudanças do ambiente com reconhecimento de padrões genéricos e confiáveis. Portanto, as redes neurais e suas características únicas são um ótimo auxílio para a implementação de sistemas em reconhecimento de faces e visão computacional.

4.3 Características dos Processos de Aprendizagem.

A propriedade mais significativa de uma rede neural é certamente a habilidade de aprender a partir do seu ambiente e melhorar o seu desempenho através da aprendizagem. Portanto, o problema do aprendizado em redes neurais é simplesmente o problema de encontrar através de um processo interativo (relativo a uma aplicação na qual cada entrada provoca uma resposta), e iterativo (reiterado) um conjunto de parâmetros livres que possibilite à rede o desempenho desejado. Visto de uma forma ideal, o reconhecimento de uma rede neural sobre o ambiente em que está inserida deve aumentar a cada iteração do processo de aprendizagem. Mendel e McClaren (1970) definem aprendizagem no contexto de redes neurais como:

"Aprendizagem é um processo pelo qual os parâmetros livres de uma rede neural são adaptados através de um processo de estimulação pelo ambiente no qual a rede está inserida. O tipo de aprendizagem é determinada pela maneira pela qual a modificação dos parâmetros ocorre."

Esta definição implica nos seguintes eventos seqüenciais:

1. A rede neural é estimulada por um ambiente.
2. A rede neural sofre modificações nos seus parâmetros livres como resultado desta estimulação.
3. A rede neural responde de uma maneira nova ao ambiente, devido às modificações ocorridas na sua estrutura interna.

Interpretando os conceitos definidos, e descrevendo a interação entre dois neurônios x_i e x_j conectados por um peso sináptico $w_{ji}(n)$ (ou seja, o peso sináptico $w_{ji}(n)$ é referente a conexão entre o neurônio de saída j e o de entrada i com valor no tempo n) pode-se intuir uma regra básica de atualização dos pesos para efetuar a aprendizagem.

$$w_{ji}(n+1) = w_{ji}(n) + \Delta w_{ji}(n), \quad (4.1)$$

Onde Δw_{ji} é o ajuste aplicado ao peso w_{ji} , e $w_{ji}(n + 1)$ é o novo valor do peso sináptico w_{ji} . Em particular, o ajuste Δw_{ji} é computado como o resultado de um estímulo do ambiente (evento 1), e o valor atualizado $w_{ji}(n + 1)$ define a mudança feita na rede devido a esse estímulo (evento 2). O evento três acontece quando a nova rede estiver operando com os pesos sinápticos atualizados.

4.3.1 Aprendizagem Supervisionada.

Em uma aprendizagem supervisionada, a rede neural é treinada com auxílio de um supervisor. Para tanto, a rede deverá possuir pares de entrada e saída, ou seja, um conjunto de entradas e um conjunto com as saídas desejadas para cada entrada padrão. Toda vez que for apresentada à rede uma entrada, deverá ser verificado se a saída obtida (gerada a partir dos cálculos efetuados a partir dos pesos que a rede possui) confere com a saída desejada para aquela entrada. Sendo diferente, a rede deverá ajustar os pesos de forma que armazene o conhecimento desejado. Esta interatividade do treino deverá ser repetida com todo conjunto de treinamento (entradas e saídas), até que a taxa de acerto esteja dentro de uma faixa considerada satisfatória. Essa forma de aprendizagem é bem conhecida e tem demonstrado excelentes resultados em aplicações reais. A figura 4.3 apresenta o diagrama em blocos representativo deste sistema.

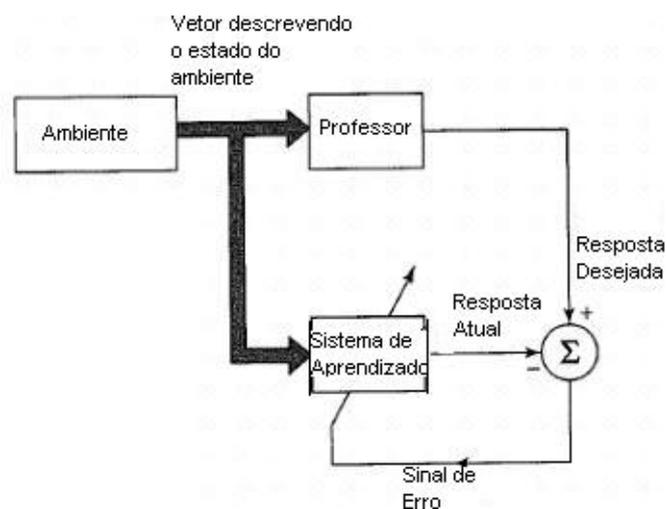


Figura 4.3: Diagrama de blocos representativos da aprendizagem supervisionada (HAYKIN, 1998).

4.3.2 Aprendizagem por Reforço.

Neste tipo de aprendizagem não é utilizado um supervisor, mas sim um crítico (ou juiz). Sendo assim, o aprendizado de um mapeamento entrada/ saída é desempenhado através da interação continuada com o ambiente buscando minimizar um índice escalar de desempenho. Neste caso, o crítico (ou juiz) possui apenas indicações imprecisas (por exemplo: sucesso/insucesso da rede) sobre o comportamento final desejado. Para ser mais exato, neste tipo de aprendizagem tem-se apenas uma avaliação qualitativa do comportamento do sistema sem, no entanto, poder medir quantitativamente o erro (desvio do comportamento em relação ao comportamento de referência desejado). Exemplo: aplicações em robótica autônoma, onde supondo uma situação hipotética, sabe-se que seguir em frente não é possível, pois existe um obstáculo, mas em compensação não se tem uma medida numérica que indique para que lado seguir e exatamente como se deve proceder para desviar do obstáculo.

A figura 4.4 mostra o diagrama em blocos de uma forma de sistema de aprendizagem por reforço, construído em torno de um crítico que converte um sinal de reforço primário recebido do ambiente em um sinal de reforço de maior qualidade chamado "sinal de reforço heurístico", ambos escalares.

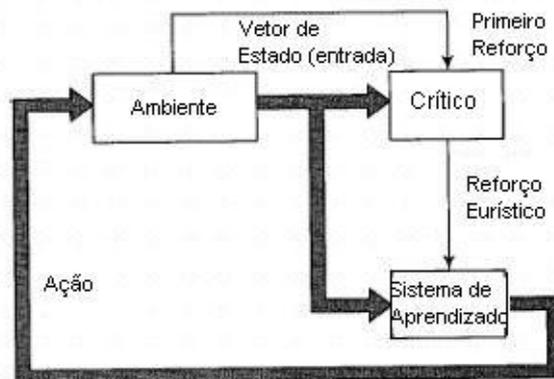


Figura 4.4: Diagrama de blocos da aprendizagem por reforço (HAYKIN, 1998).

4.3.3 Aprendizagem Não-Supervisionada.

Em algoritmos não-supervisionados ou auto-organizados não há um tutor externo, ou crítico para supervisionar o processo de aprendizagem. Esse aprendizado não requer saídas desejadas, e por isso é conhecido pelo fato de não precisar usar supervisores para o seu

treinamento. Para o treinamento da rede são usados apenas os valores de entrada. A rede trabalha essas entradas e se organiza de modo que acabe classificando-as, usando para isso os seus próprios critérios. Esse tipo de rede utiliza os neurônios como classificadores, e os dados de entrada como os elementos para classificação. O processo de classificação fica a encargo da rede neural e o seu algoritmo de aprendizagem. A auto-organização demonstrada em redes neurais não-supervisionadas, envolve, o processo de competição e o processo de cooperação entre os neurônios da rede. Muitos pesquisadores têm utilizado este tipo de rede como detector de características, dada sua capacidade de aprender a discriminar estímulos ocorrendo em partes especialmente diferentes. Em outras palavras pode-se dizer que; a rede pode aprender a distinguir padrões diferentes sem a presença de um supervisor para lhe ensinar.

O treinamento, pode-se assim dizer, inicia quando um exemplo é mostrado à rede e termina quando todas as conexões estão devidamente ajustadas (treinadas) para reconhecer todos os exemplos mostrados à rede (ou pelo menos uma boa parte dos exemplos). A partir desse momento a rede neural está pronta para a fase de testes e, a seguir, para a execução. Quanto ao conjunto de dados que será submetido à rede, pode-se dizer que, quanto mais completo o conjunto, mais tempo computacional (processamento) será utilizado para ajustar as conexões e, certamente, melhor será o seu treinamento. A figura 4.5 apresenta um diagrama desse sistema.

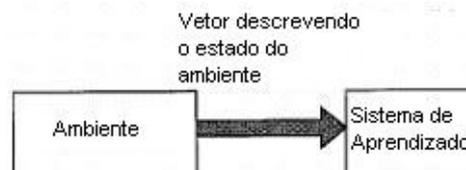


Figura 4.5: Diagrama de blocos da aprendizagem não-supervisionada (HAYKIN, 1998).

4.3.4 Generalização das Redes Neurais.

Nas aplicações utilizando redes neurais é esperado que o conjunto de treinamento apresentado à rede ofereça uma boa simulação dos padrões a serem reconhecidos pela rede no teste. Generalização é a capacidade de um modelo de aprendizagem responder corretamente aos exemplos que lhe são apresentados, sendo que estes exemplos NÃO devem

estar presentes na base de aprendizado (treinamento). Um modelo que tem uma boa generalização é aquele que responde corretamente aos exemplos contidos na base de aprendizado, mas também a outros exemplos diferentes daqueles da base de aprendizado, e que estão contidos em uma base de teste. A capacidade de generalizar é a principal capacidade buscada nas tarefas que envolvem aprendizado. Portanto, a capacidade de generalização é um fator fundamental para as aplicações das redes neurais em reconhecimento de faces, pois em ambientes reais as faces divergem consideravelmente dos padrões utilizados no treinamento devido a rotação, translação, escala, iluminação e mudanças na integridade física do rosto (envelhecimento, mudança no corte de cabelo, uso de óculos, etc).

A figura 4.6 apresenta um gráfico típico da evolução do erro durante a aprendizagem de uma rede neural, comparando a curva do erro referente à base de aprendizado (amostras de treinamento) com a curva do erro da base de teste de generalização.

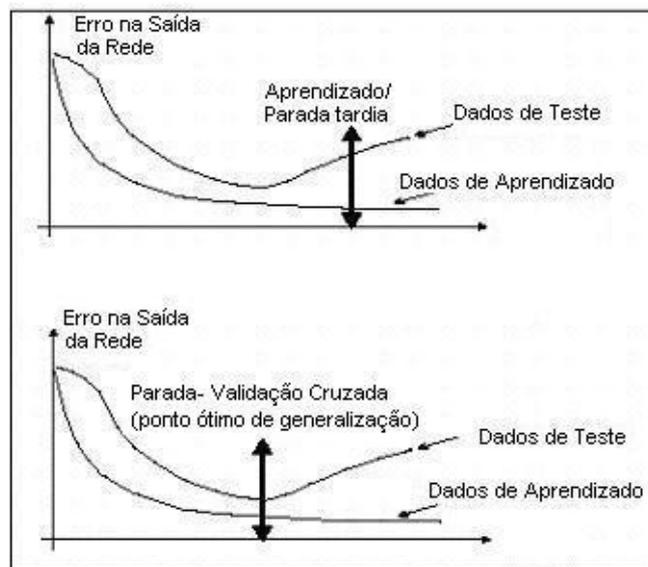


Figura 4.6: Aprendizagem: Erro em relação ao conjunto de aprendizado e de teste.

Uma rede pode se especializar demasiadamente em relação aos exemplos contidos na base de aprendizado. Esse tipo de comportamento leva a um problema de aprendizagem conhecido como super-aprendizagem (*over-training/over-fitting*). Normalmente o *over-fitting* pode ser detectado/evitado através do uso de um teste de generalização por validação cruzada (*cross-validation*).

4.3.5 Aprendizagem Hebbiana.

Hebb postulou, na obra *The Organization of Behavior*(1949), que sinapses biológicas eficazes mudam na proporção da correlação entre os sinais pré-sinápticos x e pós-sinápticos y . O Postulado de Hebb diz que:

"Quando um axônio da célula A está perto o suficiente para excitar a célula B, e participa do seu disparo repetida ou persistentemente, então algum processo de crescimento ou modificação metabólica acontece em uma das células ou em ambas, de tal forma que a eficiência da A como uma das células que dispara B é aumentada."

Hebb propôs esta modificação como uma base da aprendizagem associativa (no nível celular), que resultaria em uma modificação permanente do padrão de atividade de um "agrupamento de células nervosas"espacialmente distribuído. Stent (1973) e Changeux e Danchin (1976) expandiram esse postulado em uma regra de duas partes:

1. Se dois neurônios em ambos os lados de uma sinapse (conexão) são ativados simultaneamente (isto é, sincronamente), então a força daquela sinapse é seletivamente aumentada.
2. Se dois neurônios em ambos os lados de uma sinapse são ativados assincronamente, então aquela sinapse é seletivamente enfraquecida ou eliminada.

De posse das idéias acima, pode-se inferir uma sinapse hebbiana como uma sinapse que usa um mecanismo dependente do tempo, altamente local e fortemente interativo para aumentar a eficiência sináptica como uma função da correlação entre as atividades pré-sináptica e pós-sináptica. Portanto, o ajuste aplicado ao peso sináptico w_{ji} no tempo n é expresso em função da entrada $x_i(n)$ e saída $y_j(n)$ do neurônio, onde η é uma constante positiva que determina a velocidade da aprendizagem.

Logo,

$$\Delta w_{ji}(n) = \eta \cdot y_j(n) \cdot x_i(n) \quad (4.2)$$

Na rede de uma única unidade, mostrada na figura 4.7, os sinais pré-sinápticos e pós-sinápticos, respectivamente x e y , podem ser equacionados por:

$$W_0(n + 1) = W_0 + \eta \cdot y_0(n) \cdot X(n) \quad (4.3)$$

Onde W_0 é o vetor de pesos associado ao neurônio de índice 0, e X é o vetor de entradas.

$$y_0(n) = X(n)^T \cdot W_0 = W_0^T \cdot X(n) \quad (4.4)$$

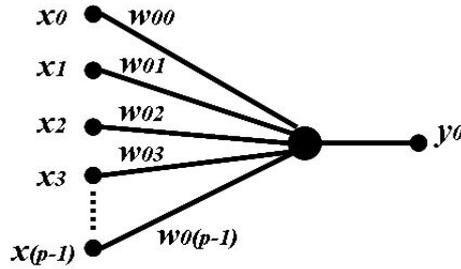


Figura 4.7: RNA de uma única unidade, a ser treinada pela aprendizagem Hebbiana.

A equação 4.3 expressa a mudança no peso sináptico, que ocorre na proporção da correlação entre os sinais pré e pós-sinápticos. A equação 4.4 expressa a saída da RNA.

O conjunto de vetores X de entrada possui distribuição de probabilidade arbitrária, e a cada tempo n um vetor X escolhido aleatoriamente do conjunto deve ser apresentado à rede. Então, segundo os princípios da aprendizagem Hebbiana, quanto mais provável for uma particular entrada X , maior será a correlação da saída y com esta entrada. Assim, quanto mais provável for X , maior será a saída y . Quanto maior for a saída y , mais a variação do peso sináptico que a encorajou aumentará. Dessa forma os pesos sinápticos permaneceriam crescendo sem limite, impedindo a continuação do processo de aprendizagem. Buscando solucionar este problema, Foldiak (1989), denominou como regra Anti-Hebbiana o ajuste do peso sináptico w_{ji} no tempo n expresso na forma:

$$\Delta w_{ji}(n) = -\eta \cdot y_j(n) \cdot x_i(n) \quad (4.5)$$

A regra Anti-Hebbiana impõem um limite ao crescimento dos pesos sinápticos w_{ji} introduzindo um fator de esquecimento não linear no ajuste dos pesos. Então, redefinindo $\Delta w_{ji}(n)$ conforme Kohonen (1988a)(1988b) tem-se:

$$\Delta w_{ji}(n) = \eta \cdot y_j(n) \cdot x_i(n) - \alpha \cdot w_{ji}(n) \cdot y_j(n) \quad (4.6)$$

onde α é uma nova constante.

A regra Hebbiana Modificada (HAYKIN, 1998) definida é fundamental para treinar uma rede neural composta de uma única camada de neurônios lineares com o objetivo de realizar a Análise de Componentes Principais para um determinado conjunto de dados.

4.3.6 Aprendizagem Competitiva.

Nessa forma de aprendizagem os neurônios da camada de saída de uma rede neural competem entre si para serem ativados. Enquanto em uma rede neural baseada em aprendizagem Hebbiana vários neurônios de saída podem estar ativados simultaneamente, em uma rede baseada em aprendizagem competitiva somente um único neurônio é ativo a qualquer tempo. Essa característica torna a aprendizagem competitiva altamente apropriada para descobrir características estatisticamente peculiares, que podem ser usadas para classificar um conjunto de padrões de entrada (HAYKIN, 1998).

Há três elementos básicos para uma regra de aprendizagem competitiva (RUMELHART; ZIPSER, 1985):

1. Um conjunto de neurônios que são todos iguais exceto por alguns pesos sinápticos aleatoriamente distribuídos, e que, portanto, respondem de forma diferente a um dado conjunto de padrões de entrada.
2. Um limite imposto na "força" de cada neurônio.
3. Um mecanismo que permite aos neurônios competirem pelo direito de responder a um dado subconjunto de entradas, tal que apenas um neurônio de saída, ou somente um neurônio por grupo está ativo, de cada vez. O neurônio que vence a competição é chamado de neurônio *winner-takes-all* (ou, vencedor leva tudo).

Os neurônios individuais da rede neural tornam-se detectores de características para diferentes classes de padrões de entrada, pois aprendem a se especializar em conjuntos de padrões similares. Assim, na forma mais simples de aprendizagem competitiva, a rede tem uma única camada de neurônios de saída, cada um deles completamente conectado aos nós de entrada. Sendo assim, para um neurônio j ser o neurônio vencedor, seu potencial de ativação v_j para um padrão específico de entrada X deve ser o maior dentre todos os neurônios na rede. O sinal de saída y_j do neurônio vencedor j é feito igual a um; os sinais de saída de todos os outros neurônios que perdem a competição são feitos iguais a zero conforme descrito por 4.7.

$$y_j = \begin{cases} 1, & \text{se } v_j > v_k \text{ para todo } k, \text{ com } k \neq j \\ 0, & \text{caso contrário} \end{cases} \quad (4.7)$$

Onde o potencial de ativação v_j representa a ação combinada de todas as entradas progressivas e realimentada para o neurônio j .

Sendo w_{ji} o peso sináptico que conecta o nó de entrada i ao neurônio j , e como os neurônios irão competir uns com os outros é justo inicializar os pesos sinápticos excitatórios com valores positivos tais que a soma dos pesos para o neurônio j seja igual a 1. Então:

$$\sum_i w_{ji} = 1 \text{ para todo } j \quad (4.8)$$

De acordo com a regra padrão de aprendizagem competitiva, a mudança Δw_{ji} aplicada ao peso sináptico w_{ji} é definida por:

$$\Delta w_{ji} = \begin{cases} \eta \cdot (x_i - w_{ji}), & \text{se o neurônio } j \text{ vence a competição} \\ 0, & \text{se o neurônio } j \text{ perde a competição} \end{cases} \quad (4.9)$$

Onde η é o parâmetro razão de aprendizagem. Esta regra tem o efeito global de mover o vetor de pesos sinápticos W_j do neurônio vencedor j na direção do padrão de entrada X .

Um exemplo clássico de rede neural que utiliza a aprendizagem competitiva é a rede de Kohonen, cuja as qualidades foram testadas nas aplicações envolvendo reconhecimento de faces para memorizar as faces individuais e classificar uma imagem teste com relação aos grupos de faces memorizadas.

4.4 Redes Neurais para Reconhecimento de Padrões.

O reconhecimento de padrões muitas vezes envolve a compressão e extração de características presentes nos dados visando conseguir as informações mais relevantes para determinar um padrão. As redes neurais podem ser vistas como um sistema que mapeia um espaço de entrada p -dimensional em um espaço de saída m -dimensional definido por uma função critério aplicada sobre as variáveis de entrada. Pode-se definir uma função mapeamento Ω com o espaço entrada e saída definidos por $R^p \Rightarrow R^m$ para $m \leq P$. Portanto, o

paradigma de aprendizagem e as regras de adaptação dos pesos da rede definem a forma de mapeamento entre a entrada e a saída, além de criar uma divisão hierárquica para os modelos de redes neurais existentes (veja figura 4.8).

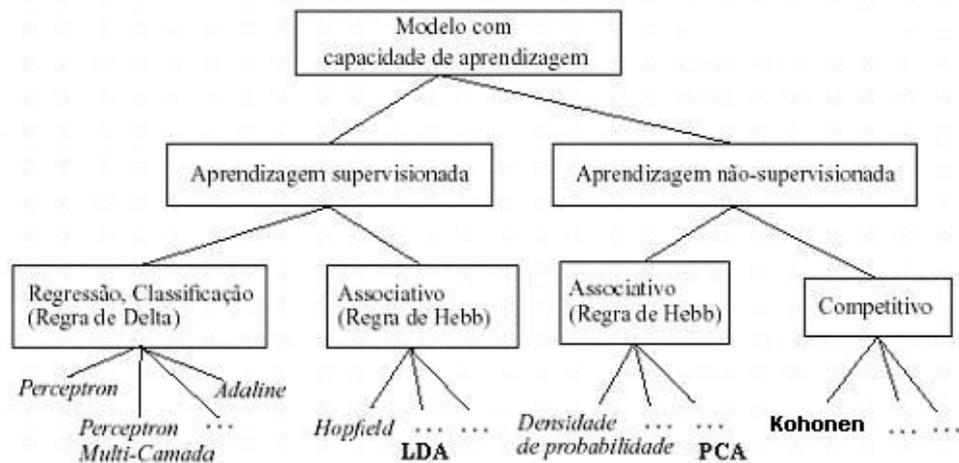


Figura 4.8: Classificação estrutural e funcional de redes neurais artificiais.

Observando a hierarquia descrita acima e relacionando os modelos neurais aplicados em reconhecimento de faces serão apresentadas nas próximas seções, quatro redes neurais que apresentam uma significância relevante para a área, fazendo parte da proposta descrita no capítulo 5 para efetuar a compressão e extração de características dos dados relacionados ao reconhecimento de faces.

4.4.1 Redes Auto-Organizadas.

Os conceitos envolvidos nas Redes PCA e Kohonen têm como fundamento teórico os modelos de aprendizagem auto-organizada ou aprendizagem não-supervisionada. O objetivo de um algoritmo para aprendizagem auto-organizada é descobrir padrões significativos ou características nos dados de entrada, e fazer essa descoberta sem um professor. Para fazer isso, o algoritmo dispõe de um conjunto de regras de natureza local, que o capacita a aprender e calcular um mapeamento de entrada-saída com propriedades desejáveis específicas; o termo "local" significa que a modificação aplicada ao peso sináptico de um neurônio é confinada à vizinhança imediata daquele neurônio. O elo de realimentação entre as mudanças nos pesos sinápticos e as mudanças nos padrões de atividade deve ser positivo para que se obtenha auto-organização da rede. De acordo com esse critério pode

ser extraído o primeiro princípio da auto-organização, conforme mostra Von der Malsburg (1990a)(1990b):

Princípio 1: Modificações dos pesos sinápticos tendem a se auto-amplificar

- Segundo o postulado de aprendizagem de Hebb: "Se dois neurônios em cada lado de uma sinapse são ativados simultaneamente, então a força daquela sinapse é seletivamente aumentada". Portanto para que ocorra o equilíbrio no sistema, o aumento na força de algumas sinapses é reduzido em outras.

Princípio 2: A limitação de recursos conduz à competição entre as sinapses, e conseqüentemente, à seleção das sinapses que crescem de forma mais vigorosa às custas de outras.

- Considerando que uma única sinapse não pode produzir de forma eficiente eventos favoráveis pode-se deduzir o terceiro princípio.

Princípio 3: As modificações em pesos sinápticos tendem a cooperar localmente.

- Todos os três princípios de auto-organização até agora descritos são relacionados somente à própria rede neural, mas para que a aprendizagem seja efetuada é necessário haver redundâncias nos padrões de entrada. Então o quarto princípio estabelecido por Barlow (1989) pode ser descrito como:

Princípio 4: A ordem e estrutura nos padrões de ativação representam a informação redundante que é adquirida pela rede neural na forma de conhecimento, a qual é pré-requisito necessário à aprendizagem auto-organizada.

Uma parte desse conhecimento pode ser obtida através da observação de parâmetros estatísticos, tais como, média, variância e matriz de correlação dos dados de entrada. Além disso, os quatro princípios da aprendizagem auto-organizada constituem a base dos algoritmos adaptativos para análise de componentes principais e mapas auto-organizados de Kohonen.

4.4.2 Rede Neural PCA Auto-Organizante.

A análise de componentes principais (conhecida como a transformação de Karhunen-Loève) maximiza a taxa de redução da variância de um conjunto de dados. Oja (1982) foi o primeiro a demonstrar que um neurônio baseado na regra de aprendizagem Hebbiana pode

convergir para a primeira componente principal de um conjunto de dados depois do treinamento. Esse modelo de neurônio único pode ser expandido em uma rede alimentada adiante (*feedforward*) com uma única camada de neurônios lineares visando encontrar todas as componentes principais dos dados apresentados na entrada da rede (SANGER, 1989a)(SANGER, 1989b).

Decorrentes dos trabalhos de Oja e Sanger, outros algoritmos têm sido propostos, considerando novas regras de aprendizagem e configurações de rede neurais artificiais. Trabalhos relevantes encontrados na literatura incluem os de (RUBNER; TAVAN, 1989), (DIAMANTARAS; KUNG, 1991), (OJA, 1989), (OJA, 1992a), (OJA, 1992b), (XU; YUILLE, 1995), (CHEN; LIU, 1992).

O conhecimento adquirido pela rede neural PCA é obtido através da observação repetida dos parâmetros estatísticos presentes nos dados de entrada, e Oja (1982) demonstrou que um único neurônio linear (figura 4.9) com uma regra adaptativa do tipo hebbiana para seus pesos sinápticos pode evoluir para a primeira componente principal de uma distribuição de dados. O vetor peso $W_0(n)$ em um neurônio operando sob a regra de aprendizagem de Hebb é dado por:

$$w_{0i}(n+1) = w_{0i}(n) + \eta \cdot y_0(n) \cdot x_i(n) \quad \text{com } i = 1, 2, \dots, p-1 \quad (4.10)$$

Onde n é o tempo discreto, η é a taxa de aprendizagem, e $p-1$ é a quantidade de entradas.

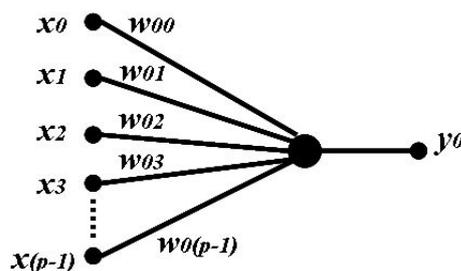


Figura 4.9: RNA de uma única unidade, a ser treinada pela aprendizagem Hebbiana.

Logo,

$$y_0(n) = \sum_{i=0}^{p-1} w_{j_i}(n) \cdot x_i(n) \quad (4.11)$$

Analisando a regra de atualização (4.10) nota-se que ocorre um crescimento ilimitado do peso sináptico, por isso é necessário incorporar uma saturação ou normalização na regra de aprendizagem para impor uma competição entre as sinapses do neurônio na busca de recursos limitados e essenciais para a estabilização. Matematicamente tem-se:

$$w_{0i}(n+1) = \frac{w_{0i}(n) + \eta \cdot y_0(n) \cdot x_i(n)}{\left(\sum_{i=1}^{p-1} [w_{0i}(n) + \eta \cdot y_0(n) \cdot x_i(n)]^2\right)^{\frac{1}{2}}} \quad (4.12)$$

A normalização ocorre com relação a todos os pesos sinápticos, que buscam a saturação para esse neurônio. Assumindo que a taxa de aprendizagem η seja pequena, e expandindo a equação 4.12 em série de potências em η , pode-se escrever:

$$w_{0i}(n+1) = w_{0i}(n) + \eta \cdot y_0(n) \cdot [x_i(n) - y_0(n) \cdot w_{0i}(n)] + O(\eta^2) \quad (4.13)$$

Onde o termo $O(\eta^2)$ representa efeitos de segunda ordem e de ordens mais altas em η , portanto, para η pequeno pode-se escrever:

$$w_{0i}(n+1) = w_{0i}(n) + \eta \cdot y_0(n) \cdot [x_i(n) - y_0(n) \cdot w_{0i}(n)] \quad (4.14)$$

Observando que o termo $-y_0(n) \cdot w_{0i}(n)$ é responsável pela estabilização modificando a entrada $x_i(n)$, escreve-se:

$x'_i(n) = x_i(n) - y_0(n) \cdot w_{0i}(n)$, que pode ser visto como a entrada efetiva da i -ésima sinapse definido como:

$$w_{0i}(n+1) = w_{0i}(n) + \eta \cdot y_0(n) \cdot x'_i(n) \quad (4.15)$$

Haykin (1998) apresenta vários trabalhos que provam a convergência desta regra Hebbiana modificada com probabilidade "1" para um vetor de comprimento unitário, que tem a mesma direção do autovetor máximo da matriz de correlação do vetor de entrada.

A generalização da regra de aprendizagem da equação 4.15 foi proposta por Sanger (1989a)(1989b). O processo de generalização consiste em aplicar as regras propostas por Sanger e Oja em uma rede neural progressiva, composta não mais por um único neurônio linear (figura 4.9), mas sim, por uma camada de neurônios lineares (figura 4.10) capazes de extrair todos os componentes principais de um conjunto de dados, pois cada neurônio converge para uma componente.

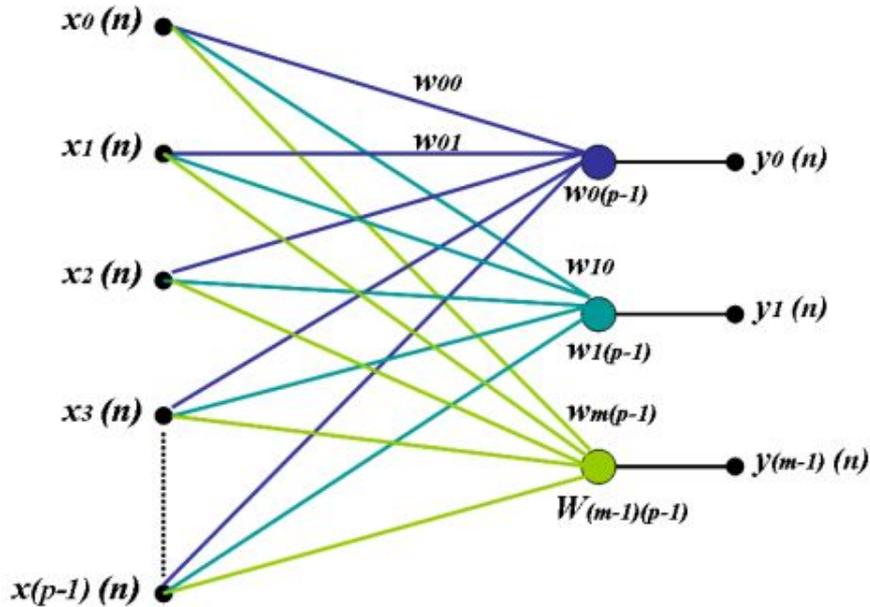


Figura 4.10: RNA progressiva com uma única camada de neurônios lineares, a ser treinada pela aprendizagem Hebbiana Generalizada.

O único objetivo de treinamento da rede é o conjunto de pesos sinápticos w_{ij} conectando os nós de entrada i aos nós de saída j .

A RNA da figura 4.10 tem p nós na camada de entrada e m neurônios na camada de saída, com $m < p$.

Os pesos sinápticos w_{ji} conectam os nós de entrada i aos neurônios da camada de saída j , com $i = 0, 1, \dots, p - 1$ e $j = 0, 1, \dots, m - 1$ (considerando novamente a notação expandida).

A saída $y_j(n)$ produzida pelo neurônio j no tempo n em resposta ao conjunto de entradas $x_i(n)$ é dada por:

$$y_j(n) = \sum_{i=0}^{p-1} w_{ji}(n) \cdot x_i(n), \quad j = 0, 1, \dots, m - 1 \quad (4.16)$$

A adaptação do peso sináptico, devida ao algoritmo Hebbiano Generalizado é expressa pela equação 4.17.

$$\Delta w_{ji}(n) = \eta \cdot [y_j(n) \cdot x_i(n) - y_j(n) \cdot \sum_{k=0}^j w_{ki}(n) \cdot y_k(n)], \quad i = 0, 1, \dots, p-1 \quad j = 0, 1, \dots, m-1 \quad (4.17)$$

onde $\Delta w_{ji}(n)$ é a mudança aplicada ao peso sináptico $w_{ji}(n)$ no tempo n e η é o parâmetro razão de aprendizagem.

Ao observar a equação 4.17 considerando $j = 0$ percebe-se que o algoritmo para um simples neurônio, apresentado na secção anterior, é um caso particular do algoritmo Hebbiano Generalizado ora apresentado.

A equação que expressa a variação no peso sináptico imposta pelo algoritmo Hebbiano Generalizado pode também ser escrita como:

$$\Delta w_{ji}(n) = \eta \cdot y_j(n) \cdot x'_i(n) - \eta \cdot y_j^2(n) \cdot w_{ji}(n), \quad i = 0, 1, \dots, p-1 \quad j = 0, 1, \dots, m-1 \quad (4.18)$$

onde o vetor $x'_i(n)$ representa a forma deflacionada do vetor de entrada $x_i(n)$, conforme equação 4.19.

$$x'_i(n) = x_i(n) - \sum_{k=0}^{j-1} w_{ki}(n) \cdot y_k(n) \quad (4.19)$$

Cabe aqui examinar a operação do algoritmo passo a passo.

1. Para o primeiro neurônio ($j=0$)

- A equação 4.18 reduz-se ao caso de um único neurônio (examinado anteriormente).
- O primeiro neurônio a convergir é aquele associado ao maior auto-valor.
- A rede extrai o primeiro componente principal do vetor de dados de entrada $X(n)$.

2. Para o segundo neurônio ($j=1$)

- A equação 4.19 torna-se $x'_i(n) = x_i(n) - w_{0i}(n) \cdot y_0(n)$.

- Desde que o primeiro neurônio já tenha convergido para o primeiro componente principal, o segundo neurônio vê vetores de entrada $X'(n)$ dos quais o primeiro auto-vetor da matriz Cx já foi extraído (deflacionado), conforme pode ser visto na equação 4.17.
- O segundo neurônio extrairá, portanto, o primeiro componente principal de $X'(n)$, que é na verdade, o segundo componente principal do espaço de entrada original dos vetores $X(n)$ (segundo auto-valor e correspondente auto-vetor da matriz Cx).

3. Para os demais neurônios da rede:

- Cada conjunto de pesos sinápticos convergido representa um auto-vetor da matriz de correlação do conjunto de dados de entrada (espaço vetorial de entrada).
- Os auto-vetores obtidos dessa forma encontram-se ordenados em ordem decrescente de valor dos auto-valores associados.

Quanto à convergência dessa rede, Sang (1989b) provou, que se a matriz de pesos sinápticos $W(n)$ representando os pesos de todos os neurônios for associada a valores aleatórios no tempo $n = 0$, então, com probabilidade "1", o algoritmo hebbiano generalizado irá convergir, e $W^T(n)$, no limite, irá se aproximar de uma matriz cujas colunas serão os primeiros m autovetores da matriz Cx de correlação dos vetores de entrada $X(n)$, ordenados por ordem decrescente de autovalor. Portanto, no limite:

$$\Delta W_j(n) \Rightarrow 0 \quad \text{e} \quad W_j(n) \Rightarrow a_j, \quad j = 0, 1, \dots, m - 1$$

$$\text{com } \|W_j(n)\| = 1$$

$$a_j = \text{autovetores da matriz } Cx$$

4.4.3 Rede Neural PCA Adaptativa.

A rede neural APEX (*Adaptive Principal Components Extraction*) ou mais conhecida como Rede PCA Adaptativa também é uma rede neural, que extrai os autovetores e autovalores relacionados a um conjunto de dados representativos de " m " dimensões correlacionados. Essa rede foi proposta por Rubner (1989) e analisada esquematicamente por Diamantaras e Kung (1996). Assim como as redes PCA auto-organizantes, esta consiste

de m neurônios de entrada e L neurônios de saída com pesos associados entre cada ligação da camada de entrada com a saída. Além disso, o algoritmo usa conexões laterais entre as saídas possibilitando uma natureza iterativa na medida em que, se for fornecida as primeiras $(j - 1)$ componentes principais, a j -ésima componente é facilmente calculada, veja figura 4.11.

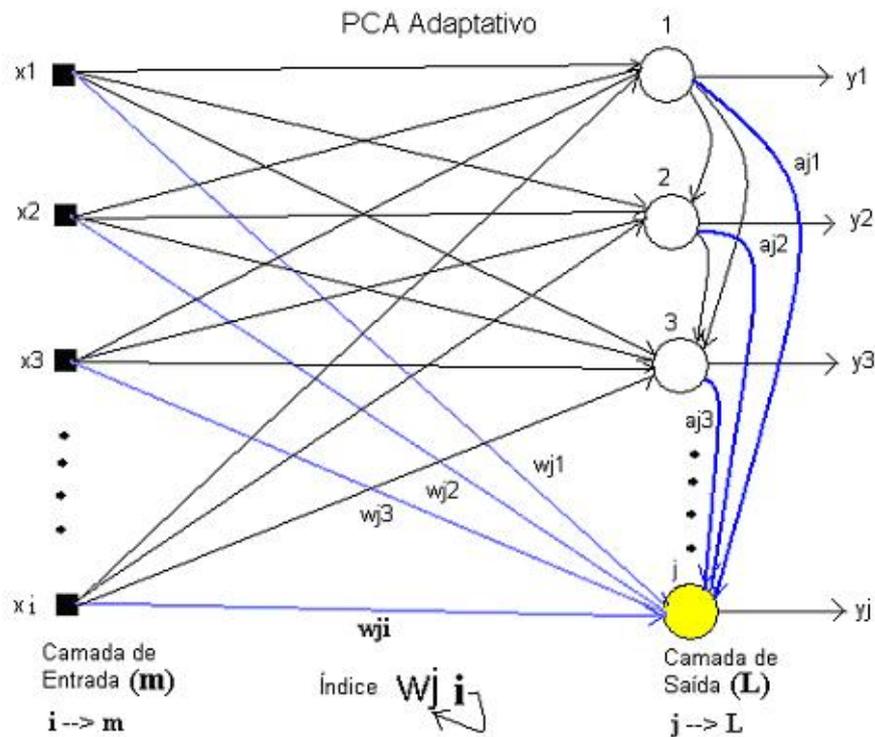


Figura 4.11: Rede neural PCA Adaptativa - APEX.

O algoritmo APEX usa tanto conexões para frente como para trás, interpretadas da seguinte forma:

- As conexões para frente entre as entradas e cada neurônio de saída $1, 2, 3, \dots, j$, com $L < m$ são conexões que operam de acordo com a regra de aprendizagem hebbiana, pois são excitatórias e responsáveis pela auto-amplificação.

Vetor peso para frente: $W_j = [w_{j1}(n), w_{j2}(n), \dots, w_{jm}(n)]^T$

- As conexões laterais das saídas individuais do neurônio $1, 2, 3, \dots, (j - 1)$ para o neurônio j aplicam uma realimentação à rede operando de acordo com uma regra de aprendizagem anti-hebbiana com efeito inibitório sobre o neurônio.

Vetor peso de realimentação: $A_j(n) = [a_{j1}(n), a_{j2}(n), \dots, a_{j(j-1)}(n)]^T$

A saída do neurônio j no tempo n produzida em resposta ao conjunto de entradas $x_i(n) \mid i = 1, 2, 3, \dots, m$ é dada por:

$$y_j(n) = \sum_{i=0}^m w_{ji}(n) \cdot x_i(n) + \sum_{l=1}^{j-1} a_{jl}(n) \cdot y_l(n) \quad (4.20)$$

ou em termos matriciais:

$$y_j(n) = W_j^T(n) \cdot X(n) + A_j^T(n) \cdot YR(n) \quad , \text{ onde } YR(n) = [y_1(n), y_2(n), \dots, y_{(j-1)}(n)]^T \quad (4.21)$$

São condições para encontrar os autovetores e autovalores associados aos dados de entrada $X(n)$:

- O vetor de entrada $X(n)$ deve representar um processo estacionário cuja matriz de correlação R (ou Cx) tenha autovalores distintos arranjados em ordem decrescente: $\lambda_1, \lambda_2, \dots, \lambda_{(j-1)}, \lambda$.

- Assume-se também que os neurônios $1, 2, 3, \dots, (j-1)$ da rede já tenham convergido para suas respectivas condições estáveis:

$$W_k(0) = q_k$$

$$a_k(0) = 0$$

onde q_k é o autovetor associado com o k -ésimo autovalor da matriz de correlação R , e o passo de tempo $n = 0$ refere-se ao início dos cálculos do neurônio j da rede.

A equação de atualização para o vetor peso para frente $w_{ji}(n)$ é efetuada de acordo com a regra hebbiana, sendo adicionado um termo *momentum* para acelerar a convergência:

$$\Delta w_{ji}(n+1) = \eta(n) \cdot x_i \cdot y_j + \beta(n) \cdot \Delta w_{ji}(n), \quad \text{com } j = 1, 2, \dots, L \text{ e } i = 1, 2, \dots, m \quad (4.22)$$

ou

$$\Delta W_j(n+1) = \eta(n) \cdot y_j(n) \cdot X(n) + \beta(n) \cdot \Delta W_j(n) \quad (4.23)$$

Já a equação de atualização para o vetor peso de realimentação $a_{jl}(n)$ é realizada de acordo com a regra anti-hebbiana, sendo adicionado um termo *momentum* também para acelerar a convergência. Dessa forma, tem-se:

$$\Delta a_{jl}(n+1) = -\mu(n).y_l.y_j + \beta(n).\Delta a_{jl}(n) \quad \text{com } j = 1, 2, \dots, L \text{ e } l = 1, 2, \dots, (L-1) \quad (4.24)$$

ou

$$\Delta A_j(n+1) = -\mu(n).y_j(n).YR(n) + \beta(n).\Delta A_j(n) \quad (4.25)$$

Onde $\eta(n)$ e $\mu(n)$ são parâmetros de aprendizagem, e $\beta(n)$ um termo *momentum* para melhorar a aproximação dos resultados em um local de mínimo.

Visualizando as equações de atualização dos pesos pode-se concluir, que se as taxas de aprendizagem $\eta(n)$ e $\mu(n)$ forem um valor suficientemente pequeno para assegurar que os ajustes dos vetores de peso ocorram lentamente, então, no limite, o vetor de peso para a frente e a potência média da saída (variância) do neurônio j se aproximam, respectivamente, do autovetor normalizado q_j e do autovalor correspondente λ_j da matriz de correlação R , como mostrado a seguir:

$$\lim_{n \rightarrow \infty} W_j(n) = q_j$$

$$\lim_{n \rightarrow \infty} \sigma_j^2(n) = \lambda_j$$

$$\text{onde } \sigma_j^2(n) = E[y_j^2(N)],$$

e a taxa de aprendizagem definida conforme segue:

$$\eta_j(n) \text{ ou } \mu_j(n) = 1/\lambda_{(j-1)}$$

Obs: o algoritmo PCA Adaptativo está descrito no Apêndice A.

4.4.4 Rede Neural LDA.

A extração de características e a projeção multivariada dos dados são temas importantes no reconhecimento de padrões e na análise de dados. A extração de características pode evitar a "maldição da alta dimensionalidade", implementando a habilidade de generalização dos classificadores, e reduzindo as exigências computacionais na classificação

dos padrões. Além disso, a projeção dos dados permite reduzir a dimensionalidade oferecendo uma melhor forma de visualização da estrutura/informação intrínseca aos dados mostrando as tendências de aglomeração. Segundo esse contexto, a Análise de Componentes Principais (PCA) é um método não supervisionado de extração de características, mas se as informações de categoria dos padrões de treinamento são conhecidas, então é mais apropriado utilizar um método de aprendizagem supervisionada para representar os dados, pois as primeiras componentes principais com maior variância podem não apresentar o melhor resultado para a classificação. A Análise Discriminante Linear (LDA) a priori, incorpora as informações relacionadas às categorias dos padrões para efetuar a projeção/transformação maximizando a dispersão entre as classes (*between-class*), enquanto mantém a dispersão dentro das classes (*within-class*) constantes ou mínimas.

Observando a importância do método LDA, e visando uma implementação paralela com redes neurais, Mao e Jain (1993)(1995) propuseram uma arquitetura de rede neural e um algoritmo de aprendizagem envolvendo etapas supervisionadas e auto-organizantes para realizar a análise discriminante linear (LDA). Matematicamente, eles estudaram o problema da seguinte forma:

Seja $\zeta_i^{(l)} = (\xi_{i1}^{(l)}, \xi_{i2}^{(l)}, \dots, \xi_{id}^{(l)})^T$ denotado como o i -ésimo padrão na classe l , $i = 1, 2, \dots, nl$, $l = 1, 2, \dots, c$, onde c é o número de categorias ou classes. Então $n = \sum_{l=1}^c nl$ denota o número total de padrões. A matriz de covariância intra-classe, $\sum W$, é

$$\sum W = \frac{1}{n} \cdot \sum_{l=1}^c \sum_{i=1}^{nl} (\xi_i^{(l)} - m^{(l)}) \cdot (\xi_i^{(l)} - m^{(l)})^T \quad (4.26)$$

Onde $m^{(l)}$ é o vetor médio da classe l , $l = 1, 2, \dots, c$. Similarmente é definida a matriz de covariância entre-classe, $\sum B$, como

$$\sum B = \frac{1}{n} \cdot \sum_{l=1}^c (m^{(l)} - m) \cdot (m^{(l)} - m)^T \quad (4.27)$$

Onde m é o vetor médio do conjunto total dos dados. A matriz de dispersão total é, portanto,

$$\sum T = \frac{1}{n} \cdot \sum_{l=1}^c \sum_{i=1}^{nl} (\xi_i^{(l)} - m) \cdot (\xi_i^{(l)} - m)^T = \sum W + \sum B \quad (4.28)$$

A meta do LDA é, então, encontrar uma transformação Φ , $d.m.$, onde o determinante $\frac{\|\Phi^T \cdot \sum B \cdot \Phi\|}{\|\Phi^T \cdot \sum W \cdot \Phi\|}$ é maximizado. Assim pode ser provado que a transformação, Φ , é composta pelos m autovetores correspondente aos m maiores autovalores não-zeros de $(\sum W)^{-1} \cdot \sum B$. Devido ao fato que a matriz $\sum B$ tem um *rank* máximo de $c - 1$, o valor de m deve ser menor que c . Portanto, a dimensionalidade do espaço projetado é limitado pelo número de classes. Nesse trabalho é usada a matriz de covariância total $\sum T$ no lugar da matriz de covariância entre-classes $\sum B$ na função critério, tornando a saída não limitada pelo número de classes.

A rede neural proposta por Mao e Jain para efetuar a análise discriminante linear é apresentada na figura 4.12b, e possui duas camadas, onde cada uma é igual à rede neural (APEX) mostrada anteriormente.

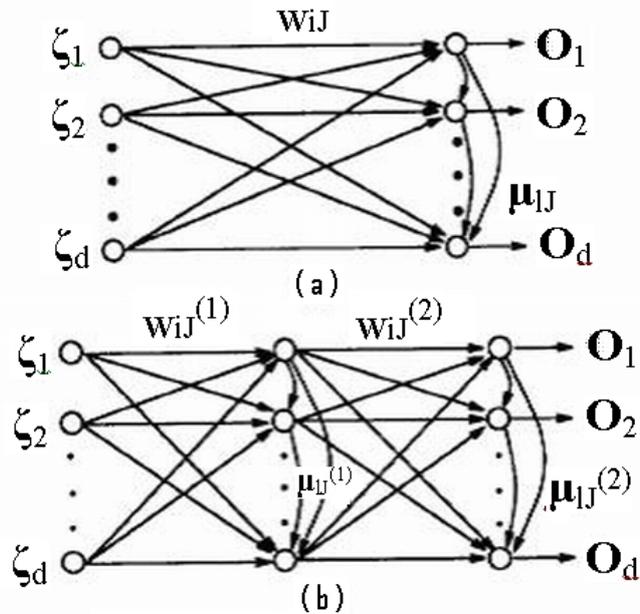


Figura 4.12: (a) Rede neural PCA proposta por Rubner e Tavan (1989). (b) Rede neural LDA proposta por Mao e Jain (1995)(1993).

Nos neurônios da rede são usadas funções de ativação linear. Os pesos $W_{ij}^{(1)}$ e $W_{ij}^{(2)}$ são conexões intercamadas do i -ésimo neurônio na camada de entrada (ou intermediária) para o j -ésimo neurônio na camada intermediária (ou de saída). Conseqüentemente $\mu_{ij}^{(1)}$ e $\mu_{ij}^{(2)}$ denotam os pesos laterais do i -ésimo neurônio para o j -ésimo neurônio dentro da camada intermediária (ou de saída). Além disso, $\xi = (\xi_1, \xi_2, \dots, \xi_d)^T$ é um padrão de entrada d -dimensional, e $\rho = (\rho_1, \rho_2, \dots, \rho_d)^T$ e $o = (o_1, o_2, \dots, o_d)^T$ são vetores de saída da camada intermediária e final, respectivamente. Pode-se então, escrever as saídas como:

$$\rho_j = \sum_{i=1}^d W_{ij}^{(1)} \cdot \zeta_i + \sum_{l < j} \mu_{lj}^{(1)} \cdot \rho_l \quad (4.29)$$

$$o_j = \sum_{i=1}^d W_{ij}^{(2)} \cdot \rho_i + \sum_{l < j} \mu_{lj}^{(2)} \cdot o_l \quad (4.30)$$

Obs: o algoritmo LDA proposto por Mao e Jain (1995)(1993) está descrito no Apêndice A.

A partir do método proposto, vários outros surgiram na literatura. Chatterjee et Roychowdhury (1997) questionaram o método de Mao e Jain alegando, que haviam alguns problemas: a primeira camada da rede tem que ser treinada inteiramente, até que a convergência seja obtida, antes que o treinamento da segunda camada possa começar. Assim, na formulação atual, o treinamento das duas camadas não pode prosseguir simultaneamente, sendo pouco prático quando a entrada do treinamento é um fluxo ou uma seqüência dos dados. O método requer todos os dados juntos para o treinamento da rede, e isso pode ser impossível em muitas aplicações em tempo real. Além disso, as classes e suas médias devem ser conhecidas antes do treinamento. O novo método proposto por eles supera essas limitações. Devido à natureza adaptativa do algoritmo, no treinamento, ambas as camadas são calculadas simultaneamente, e não requerem o conhecimento a priori das médias das classes. Ao contrario, as médias das classes são estimadas durante o processo de treinamento.

Outras arquiteturas de redes neurais com outros algoritmos de aprendizagem podem ser encontradas na literatura: (Chatterjee et al., 1997), (Príncipe; Xu, 1997), e um método proposto por Moghaddam e Zadeh (2002) com maior velocidade de convergência, que os anteriores. Mas, o método proposto por Mao e Jain é mais simples e prático em aplicações que não exigem respostas em tempo real.

4.4.5 Rede Neural de Kohonen.

As redes de Kohonen são mapas auto-organizáveis constituídos de neurônios dispostos na forma de grades baseados na aprendizagem competitiva, cuja finalidade é descobrir padrões nos dados de entrada. Um sinal incidente de dimensão arbitrária é levado em um mapa discreto uni ou bi-dimensional topologicamente ordenado de acordo com a proximidade dos padrões dispostos na dimensão original. Esta rede foi proposta por Teuvo

Kohonen da Universidade de Helsinki, na Finlândia, e surgiu devido ao interesse do grupo em extrair características de um conjunto de fonemas observados usando elementos da Teoria de Probabilidade e da Teoria das Decisões Clássicas.

Os neurônios da grade de saída, que estão diretamente conectados às unidades de entrada competem entre si para responder a um dado padrão de entrada, e apenas um será o vencedor, que poderá atualizar os seus pesos em direção ao dado. Exemplificando, a rede neural de Kohonen (1988c) baseia-se no método de classificação do vizinho mais próximo, que interpreta um conjunto de dados como padrões e usa estatística tradicional para formar classes C_i de padrões, onde cada classe é representada por um pequeno conjunto de padrões P_i . Cada novo padrão p pode ser classificado através da identificação da classe C_p para a qual p é mais próximo com respeito a alguma métrica já definida. Um mapa auto-organizável é, portanto, caracterizado pela formação de um mapa topográfico dos padrões de entrada no qual as localizações espaciais (isto é, coordenadas) dos neurônios na grade são indicativas das características estatísticas intrínsecas contidas nos padrões de entrada.

A estrutura proposta por Kohonen tem uma estreita relação com a organização e funcionamento do córtex cerebral, que apresenta características de auto-organização e memória associativa no processamento das informações sensoriais.

O mapa de características de Kohonen é uma rede de duas camadas. A primeira camada é a camada de entrada. Tipicamente, a segunda é a camada competitiva e é organizada como uma grade bi-dimensional (veja figura 4.13). Todas as interconexões vão da primeira camada para a segunda; as duas camadas são completamente interconectadas, pois cada unidade de entrada é conectada a todas as unidades da camada competitiva. A figura 4.14 apresenta várias estruturas de conexões entre um neurônio e seus vizinhos.

Quando um padrão de entrada é apresentado, cada unidade na primeira camada assume o valor da entrada e as unidades da segunda camada somam suas entradas e competem para encontrar uma única unidade vencedora. Cada interconexão no mapa de características de Kohonen tem um peso associado, que são iniciados aleatoriamente. Os valores de peso são atualizados durante o treino da rede.

Assim, supondo que os vetores de entrada são de dimensão m (pertencem ao espaço R^m) e a rede é de dimensão 2 com r linhas e c colunas, um padrão de entrada apresentado ao mapa de características de Kohonen é denotado como:

$$X = [x_1, x_2, \dots, x_k, \dots, x_m]$$

e os pesos são dados por $W_j = [w_{j1}, w_{j2}, \dots, w_{jm}]$, para $j = 1, 2, \dots, l$

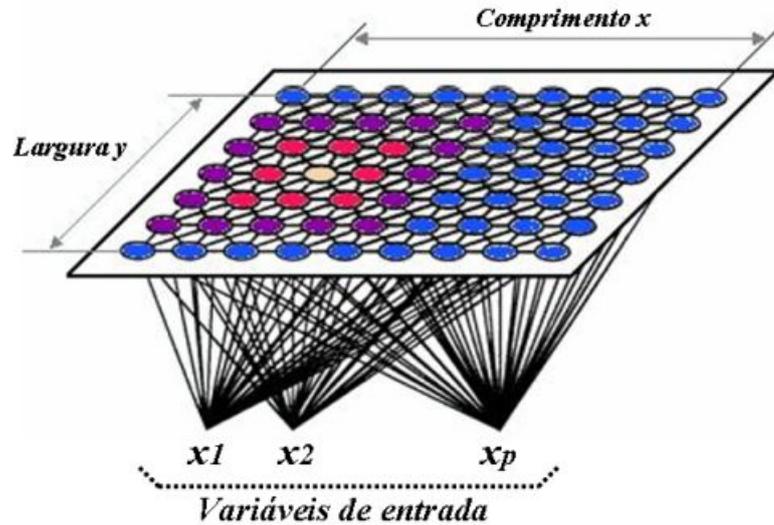


Figura 4.13: Exemplo de estrutura da Rede de Kohonen com neurônios de entrada e saída.

com $\|X\| = \|W\| = 1$, onde j identifica a unidade na camada competitiva.

No treinamento, o primeiro passo na operação de uma rede de Kohonen é calcular o grau de similaridade para cada unidade na camada competitiva. Este valor quantifica o quanto os pesos de cada unidade se igualam aos valores correspondentes do padrão de entrada. O grau de similaridade para a j -ésima unidade é:

$$\|X - W\|$$

que é a distância Euclidiana entre os vetores X e W_j , e é calculada por:

$$\sqrt{\sum_j (x_k - w_{jk})^2} \quad (4.31)$$

Com $k = 1, 2, \dots, m$ e $j = 1, 2, \dots, l$.

A unidade com o maior grau de similaridade vence a competição. Isto é equivalente a maximizar o produto interno $W_j \cdot X^T$ (HAYKIN, 1998).

Se for usado o índice i para identificar o neurônio com maior similaridade com o vetor X que está entrando na rede no momento, i é tal que:

$$\|X - W_i\| = \min_j (\|X - W_j\|) \quad (4.32)$$

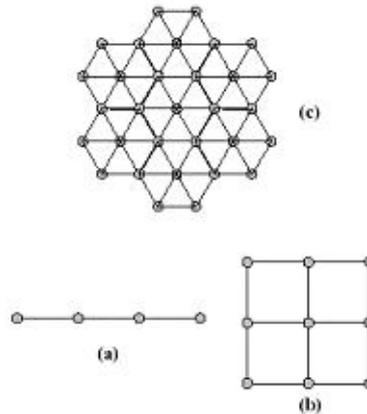


Figura 4.14: Exemplos de arranjos. (a) arranjo linear; (b) arranjo retangular; (c) arranjo hexagonal.

Se for usado o índice $i(X)$ para identificar o neurônio com maior similaridade com o vetor X , pode-se então determinar $i(X)$ aplicando a condição:

$$i(X) = \arg \min_j (\|X - W_j\|) \quad (4.33)$$

Depois que a unidade vencedora é identificada, o próximo passo é identificar a vizinhança em torno dela. A vizinhança, ilustrada nas figuras 4.15 e 4.16 de atualização dos pesos, consiste daquelas unidades processadoras que estão próximas à vencedora na grade da camada competitiva.

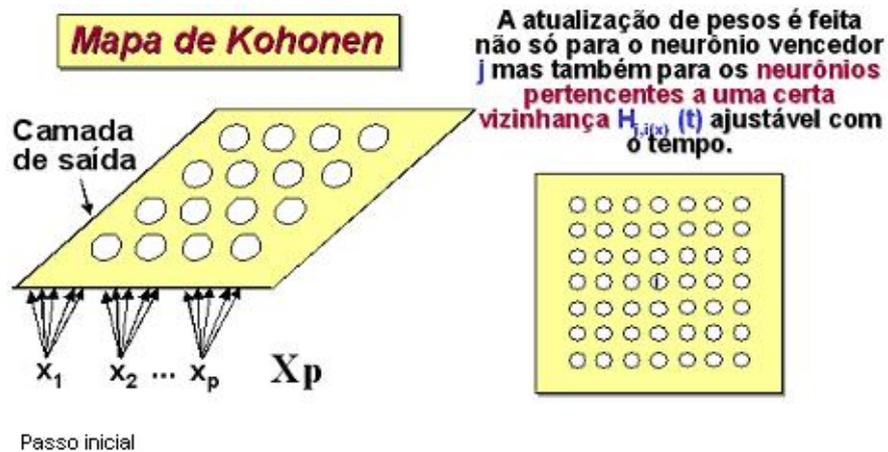


Figura 4.15: Neurônio vencedor no processo competitivo.

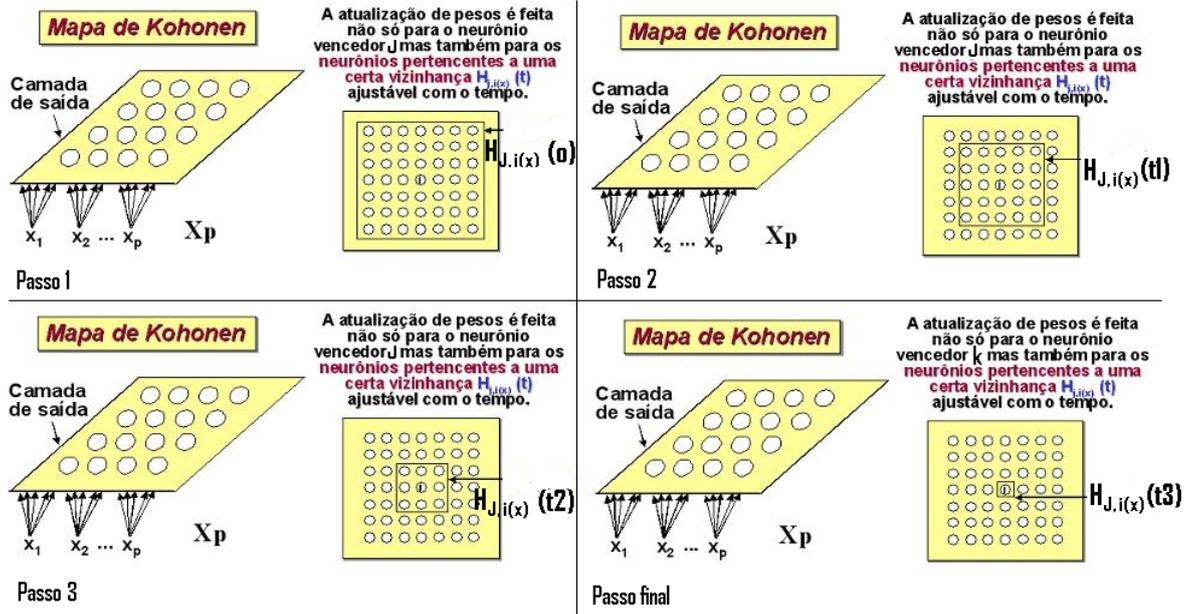


Figura 4.16: Atualização dos pesos na vizinhança do neurônio vencedor.

A vizinhança neste caso consiste das unidades que estão dentro de um quadrado que é centrado na unidade vencedora i . O tamanho da vizinhança muda, como mostrado pelos quadrados de diferentes tamanhos na figura. A vizinhança é denotada pelo conjunto de unidades $H_i(x)$. Os pesos são atualizados para todos os neurônios que estão na vizinhança da unidade vencedora, usando a equação:

$$W_j(n+1) = \begin{cases} W_j(n) + \eta(n) \cdot [x(n) - W_j(n)], & \text{se } j \in H_{i(x)}(n) \\ W_j(n), & \text{caso contrário} \end{cases} \quad (4.34)$$

onde $\eta(n)$ é o parâmetro da taxa de aprendizagem, e $H_{i(x)}(n)$ é a função de vizinhança centrada em torno do neurônio vencedor $i(x)$; ambos $\eta(n)$ e $H_{i(x)}(n)$ variam dinamicamente durante a aprendizagem para obter melhores resultados.

Desse ajustamento resulta a mudança dos pesos da unidade vencedora e de sua vizinhança, tornando-os mais parecidos com o padrão de entrada. A unidade vencedora então torna-se mais provável de (ou mais apta a) vencer a competição seja ao se apresentar o mesmo padrão na entrada, ou um semelhante. A figura 4.17 mostra a aprendizagem de forma geométrica.

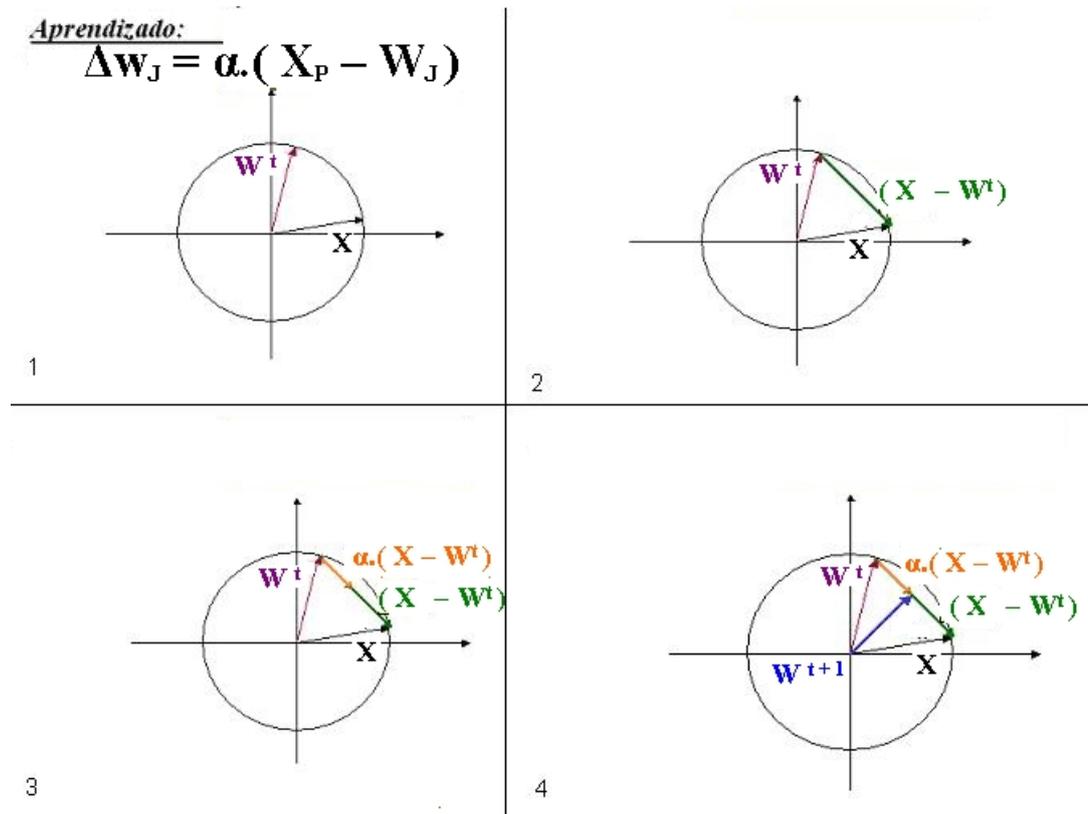


Figura 4.17: Visualização geométrica da atualização dos pesos na rede de Kohonen.

Note que existem dois parâmetros que devem ser especificados: o valor de $\eta(n)$ que é a taxa de aprendizagem na equação de ajustamento 4.34, e o tamanho da vizinhança $H_{j,i(x)}$. A taxa $\eta(n)$ deve começar em um valor inicial em $n = 0$ e então decrescer gradualmente com o aumento do tempo n . Esta exigência pode ser satisfeita escolhendo-se um decremento exponencial para $\eta(n)$, como:

$$\eta(n) = \eta_0 \cdot \exp\left(\frac{-n}{\tau_1}\right), \quad n = 1, 2, \dots \quad (4.35)$$

E para o processo cooperativo entre o neurônio vencedor e sua vizinhança pode-se escolher uma função Gaussiana para $H_{j,i(x)}$ da seguinte forma:

$$H_{j,i(x)}(n) = \exp\left(\frac{-d_{j,i}^2}{2 \cdot \sigma^2(n)}\right), \quad n = 0, 1, 2, \dots, \quad (4.36)$$

$$\sigma(n) = \sigma_0 \cdot \exp\left(\frac{-n}{\tau_2}\right), \quad n = 1, 2, \dots, \quad (4.37)$$

Obs: Os parâmetros τ_1 e τ_2 são constantes de tempo. Para a grade bi-dimensional a distância entre o neurônio vencedor r_i e o atualizado r_j é:

$$d_{j,i}^2 = \|r_j - r_i\|^2 \quad (4.38)$$

4.4.5.1 Critério de Convergência

Para assegurar convergência dos vetores-peso, dois requisitos devem ser satisfeitos:

1. η deve satisfazer as seguintes condições:

$$0 < \eta(n) < 1 \quad \text{para todo } n$$

$\eta(n)$ deve ser uma função decrescente

Quando n atinge um certo limite, diz-se que a aprendizagem ocorreu.

2. O tamanho da vizinhança Nz deve satisfazer duas condições:

$$1 < Nz(n) < \text{Max (linha, coluna)} \quad \text{para todo } n$$

$Nz(n)$ deve ser uma função decrescente onde $\text{Max (linha,coluna)}$ é considerado o diâmetro da rede.

4.4.5.2 Critério de Parada

O parâmetro $\eta(n)$ deve ser decrescente e atualizado a cada nova entrada. Após todas as entradas serem apresentadas, verifica-se um limite próximo de zero estabelecido para a convergência. Em caso positivo, o treinamento está finalizado e o resultado pode ser visualizado através dos neurônios da rede. Portanto, para cada nova entrada após o treinamento, é associado o neurônio que melhor a representa de acordo com o critério da menor distância euclidiana.

4.5 Considerações Finais.

Os principais conceitos envolvendo as redes neurais artificiais, utilizadas na compressão e extração de características relevantes em um banco de dados, foram descritas nesse capítulo. Os conceitos de aprendizagem supervisionada e não-supervisionado também foram apresentados associando as características estruturais das redes PCA, PCA Adaptativo, LDA e SOM (Kohonen) nas aplicações de reconhecimento de padrões.

No próximo capítulo será descrito detalhadamente o algoritmo implementado para o reconhecimento de faces, e apresentado os testes relacionados aos subespaços PCA e LDA juntamente com as avaliações gerais do algoritmo.

Implementação do Algoritmo e Resultados

5.1 Considerações Iniciais.

O reconhecimento de faces humanas busca desenvolver algoritmos matemáticos para fins de autenticação ou identificação, extraindo características importantes para o reconhecimento facial e direcionando a busca no banco de dados de treinamento. Na autenticação o usuário precisa digitar algum código ou passar, por exemplo, um crachá ou *smart card* para que na seqüencia o sistema capture a face e compare com a face indexada pelo código no banco de faces . É a chamada verificação um para um (*one-to-one*) - ou seja, o sistema não faz uma busca em todo o banco de faces, apenas comprova que a pessoa é ela mesma. A identificação exige um sistema mais sofisticado, pois a comparação é feita de 1 para "M"(*one-to-many*) indivíduos dentro do banco de faces.

Cada algoritmo biométrico possui vantagens e deficiências, e miram aplicações específicas. Em locais com alta concentração de pessoas, como shoppings e aeroportos, o reconhecimento facial é a tecnologia de excelência para separar uma pessoa da multidão para averiguação posterior (aplicação *one-to-many*). Mas dificilmente será usada

para autorizar um saque no banco (aplicação *one-to-one*), pois apresenta deficiências com relação às taxas de falsa aceitação (aceita-se uma pessoa não autorizada) e falsa rejeição (rejeita-se alguém que realmente está cadastrado). A intenção deste trabalho é classificar/identificar uma face com o seu respectivo representante na base de dados associando o indivíduo correspondente (aplicação *one-to-many*).

5.2 Modelo Proposto.

Embora a etapa de detecção tem como objetivo segmentar a face na imagem e normalizá-la, muitas vezes essa etapa não consegue normalizar as faces com relação à iluminação, rotação e escala transferindo esses problemas para a etapa de reconhecimento. A proposta deste trabalho de pesquisa é baseada no reconhecimento de faces utilizando as características holísticas da imagem, pois são eficazes quando aplicados ao reconhecimento de faces em bancos de dados grandes com variações na iluminação, rotação e escala das faces na cena, considerando que estes fatores estão presentes nas aplicações do mundo real.

O diagrama em blocos mostrado na figura 5.1 representa um modelo automático completo para o reconhecimento de faces baseado em métodos holísticos com estágio de detecção e de reconhecimento para imagens de vídeo ou fotografia, garantindo uma representação compacta para a face, que pode ser usada tanto para o reconhecimento como para aplicações de compressão/transmissão de imagens de face. Neste trabalho não serão enfocados os elementos presentes no estágio de detecção automática da face, pois toda dedicação será aplicada no estágio de reconhecimento/identificação da mesma.

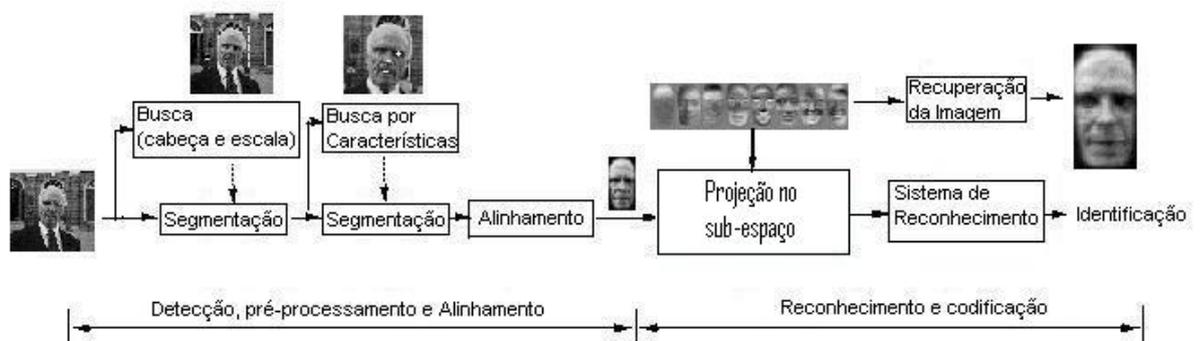


Figura 5.1: Diagrama em blocos de um sistema de reconhecimento de faces.

Os métodos holísticos em reconhecimento de faces tomam como vetor de características os pixels relacionados à face na imagem, e para evitar o problema da alta dimensionalidade, vários algoritmos estatísticos de redução de dimensionalidade podem ser utilizados, como por exemplo: Análise de Componentes Principais (PCA), Discriminantes Lineares e Redes Neurais (LAWRENCE et al., 1996)(ROMDHANI, 1996). O método de reconhecimento com PCA é o mais popular, devido a sua facilidade na implementação, tendo sido freqüentemente utilizado em associação com pré-processamentos de normalização de imagens para melhorar o desempenho. A classificação pode ser feita de diversas maneiras, geralmente através de redes neurais ou sistemas estatísticos.

A técnica PCA foi exaustivamente estudada nas aplicações envolvendo reconhecimento de faces, e a sua implementação complementada por uma etapa utilizando LDA tem apresentado bons resultados. O método Sub-espço LDA (PCA + LDA) tornou-se muito comum para reconhecimento de faces a partir de 1997, com a publicação do artigo (BELHUMEUR et al., 1997). Nesse artigo, os autores comparam PCA com o Sub-espço LDA e mostram que o espaço de características criado pela transformação LDA proporciona resultados de classificação muito melhores que o espaço criado pela transformada PCA para o reconhecimento de pessoas em imagens com grandes variações de iluminação e expressões faciais, pois o LDA utiliza informações das classes associadas a cada padrão para extrair linearmente as características mais discriminantes.

Na literatura encontram-se vários trabalhos explorando as aplicações do Sub-espço LDA em reconhecimento de faces. (MARTINEZ; KAK, 2001), (ZHAO et al., entre 1997 e 2002) e (BELHUMEUR et al., 1997) compararam os métodos PCA e Sub-espço LDA utilizando vários bancos de Faces com inúmeras situações que poderiam causar erro no reconhecimento (translação, escala, rotação, iluminação, etc...). (WANG; TANG, 2002), (WANG; TANG, 2004), (LU et al., 2003) e (YANG et al., entre 1995 e 2000) analisaram os efeitos relacionados a quantidade de amostras de treinamento com relação ao desempenho do algoritmo. (ZHAO et al., 2002?), (ETEMAD; CHELLAPPA, 1997) e (ZHAO; CHELLAPPA, entre 1995 e 2000) empregaram o Sub-espço LDA em vários bancos de faces observando o desempenho e os efeitos sofridos pelos algoritmos com relação a quantidade de componentes discriminantes. A maioria dos trabalhos existentes na área utiliza os conceitos de distância Euclidiana, distância Angular e redes neurais *perceptron* com algoritmo de aprendizado *Back-propagation* para efetuar a classificação/identificação das classes.

Em nosso trabalho, será efetuada uma análise do algoritmo Sub-espço LDA utilizando um conjunto de redes neurais SOM na etapa de classificação e uma rede neural

PCA adaptativa nos cálculos dos componentes principais (PCA) necessários ao algoritmo (SANTOS; GONZAGA, 2005b). Espera-se que as redes SOM apresentem bons resultados na memorização das faces com leves variações de rotação, pois existem artigos, como por exemplo (CHAOYANG et al., 2003) e (TAKAHASHI; KURITA, 2001), que descrevem o desempenho dessas redes em seqüências de imagens com variações rotacionais.

Os mapas de características auto-organizantes (SOM) foram utilizados na classificação de faces, porque este algoritmo tem sido aplicado com sucesso em vários problemas no domínio de *cluster analysis*, reconhecimento da fala, quantização de vetores, etc, devido a sua característica de formar mapas topologicamente ordenados. A principal meta da SOM é transformar/memorizar padrões de dimensionalidade arbitrária em respostas de neurônios arranjados em 1-D ou 2-D preservando o mapeamento com relação à vizinhança dos padrões de um espaço para outro (CHAOYANG et al., 2003).

A arquitetura utilizada foi o modelo Mapa Concorrente Auto-Organizante (*Concurrent Self-Organizing Maps* - CSOM) implementado por Neagoe e Ropot (2002) para efetuar a classificação de faces e imagens de satélites. A etapa de classificação foi implementada com uma coleção de pequenas redes SOM utilizando a regra de aprendizado *winner-takes-all* (o ganhador leva tudo). Cada SOM do modelo é treinada individualmente para prover o melhor resultado com relação a uma única classe, sendo o número de redes SOM igual à quantidade de classes.

Os vetores de faces de pessoas diferentes se agrupam no espaço n-dimensional de acordo com características preponderantes; por exemplo, os vetores de faces representando as pessoas carecas tendem a formar um *cluster* orientado por essa característica. O ideal seria que faces da mesma pessoa se agrupassem próximas no espaço n-dimensional, mas as mudanças de iluminação e rotação interferem na distribuição dos vetores formando *clusters* guiados por essas características e aproximando faces de pessoas diferentes em uma mesma região (SANTOS; GONZAGA, 2005a). Baseado nesses conceitos a rede CSOM foi escolhida ao invés de escolher a rede SOM, pois a arquitetura SOM apresenta problemas para classificar faces de várias pessoas com diferentes patamares de iluminação e rotação.

O diagrama em blocos do modelo proposto para a fase de treinamento é mostrado na figura 5.2. O diagrama mostra as etapas relacionadas à construção do sub-espaço PCA e posteriormente o sub-espaço LDA baseados nas amostras de treinamento, sendo apresentado na seqüência o estágio de classificação implementado pelas redes neurais individuais SOM.

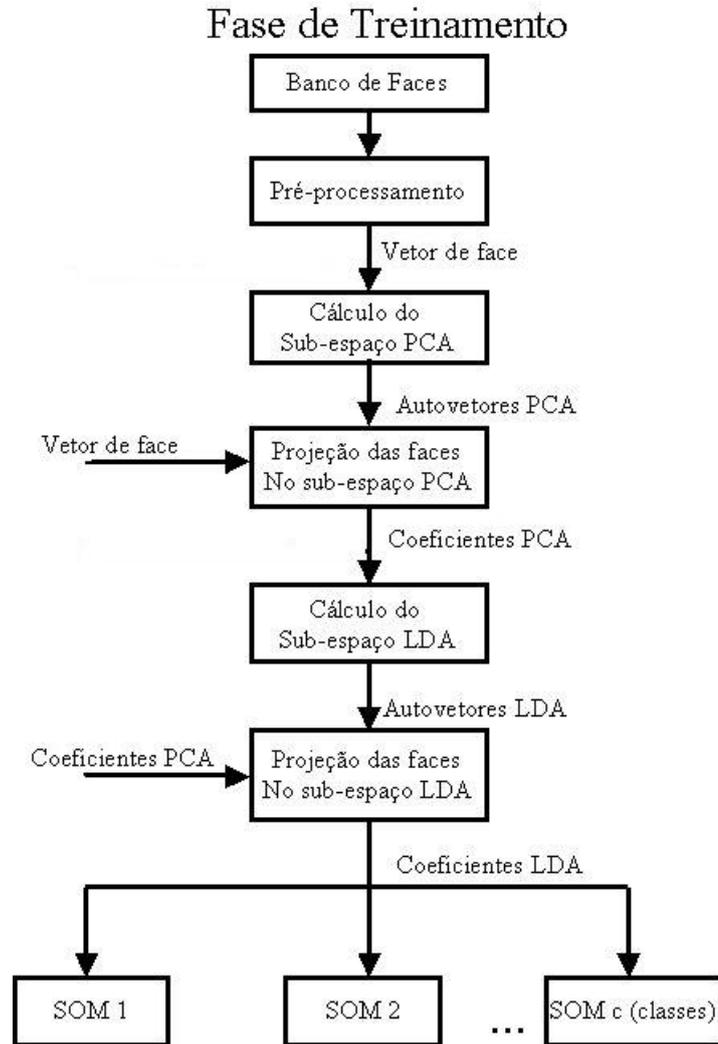


Figura 5.2: Diagrama em blocos do modelo proposto - fase de treinamento.

Os cálculos relacionados à construção do sub-espço PCA serão implementados através da rede neural APEX, pois matematicamente, o PCA é um método que calcula os autovetores e autovalores da matriz de covariância dos dados de entrada, e este cálculo através dos métodos tradicionais (JAMES et al., 1993) tem seu custo computacional alto. Como o nosso interesse está relacionado com a finalidade de encontrar os primeiros autovetores relacionados aos primeiros autovalores sem necessariamente encontrar todos os autovetores, a rede neural APEX é um método que oferece a possibilidade de se calcular apenas os primeiros autovetores de interesse. Além disso, as redes neurais podem ser implementadas por *hardware* com tecnologia VLSI ocupando pouco espaço e oferecendo alto desempenho no processamento paralelo efetuado pelos neurônios da rede (SPENCER; SINENCIO, 1999). A implementação em *hardware* das redes neurais envolvidas

no projeto não é o objetivo deste trabalho ficando apenas como sugestão para trabalhos futuros.

O diagrama em blocos do modelo para a fase de teste/reconhecimento é mostrado na figura 5.3. Na fase de teste o padrão é aplicado na entrada do algoritmo e projetado no sub-espço PCA, onde decide-se através de um *threshold* pré-definido se a imagem apresentada é ou não é face; se a imagem for uma face, os seus coeficientes PCA são projetados no sub-espço LDA para posteriormente serem aplicados em paralelo nas entradas das redes SOM. Após os cálculos de similaridades efetuados pelas redes SOM tem-se como resposta a classe com maior similaridade com relação à imagem de entrada. A similaridade é definida através da norma Euclidiana, e cada classe possui um valor de *threshold* para defini-la como representante ou não da face de teste.

Nessa etapa, as redes SOM também oferecem a possibilidade de serem implementadas por *hardware* aumentando a dinâmica da rede e aproveitando o processamento paralelo.

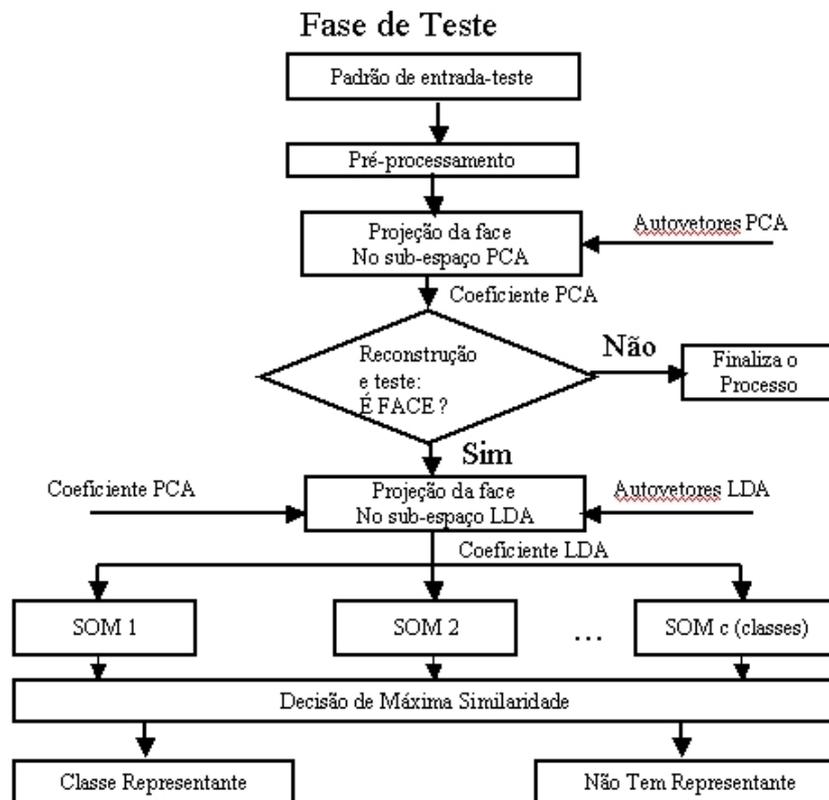


Figura 5.3: Diagrama em blocos do modelo proposto - fase de teste.

5.3 Banco de Imagens.

Para avaliar o desempenho do modelo proposto, e ressaltar os possíveis desafios tratados pelo algoritmo no reconhecimento de faces foi utilizado o banco de faces da ORL, que foi produzido pela *Olivetti Research Laboratory in Cambridge, UK*. Este banco de dados é gratuito/ público e contém 10 imagens diferentes de 40 indivíduos distintos (total de 400 imagens). Há variações na expressão facial e nos detalhes faciais, pois para o mesmo indivíduo as imagens foram adquiridas em horas diferentes, com condições luminosas não padronizadas, e os indivíduos podem se apresentar com ou sem óculos. Todas as imagens foram adquiridas sob uma cena de fundo (*background*) homogênea e escura com os indivíduos em posição frontal e ereta apresentando uma tolerância de mais ou menos 20 graus com relação à rotação da cabeça. Estas imagens estão na escala de cinzas (256 níveis) com uma resolução de 92x112 pixels. A figura 5.4 mostra alguns indivíduos presentes no banco de faces ORL.



Figura 5.4: Exemplo de faces presentes no banco de faces ORL.

A maioria dos experimentos em reconhecimento de faces trabalha com algumas dezenas de pessoas totalizando algumas centenas de faces. Para aumentar o tamanho do banco de faces, as imagens foram espelhadas com relação ao eixo *y* gerando mais 400 imagens virtuais, pois elas não refletem a realidade com relação à simetria da face, porque o lado direito e esquerdo da face não são simétricos, e sim, ligeiramente semelhantes. Mas as imagens espelhadas são um artifício para aumentar o conjunto de faces no banco de dados que agora possui 20 imagens diferentes de 40 indivíduos.

Esse banco de dados foi escolhido porque é público, possui faces segmentadas com plano de fundo homogêneo, possui várias faces de uma mesma pessoa (importante para a aplicação LDA), e as faces apresentam várias divergências (iluminação, rotação, escala, feição, etc) interessantes para análises.

Vários trabalhos (fazer as citações) utilizam esse banco de dados para avaliar os seus algoritmos com 5 faces de treinamento e 5 faces de teste para cada pessoa. A nossa proposta é realizar os testes de forma mais abrangente testando várias possibilidades e combinações de agrupamentos.

Para os vários experimentos apresentados a seguir, o conjunto de imagens foi dividido em um conjunto de treinamento e em um conjunto de teste com 5, 6, 7, 8, 9, 10 faces de treinamento aleatoriamente escolhidas para cada indivíduo e o restante definido como faces de teste. Além disso, para cada conjunto de treinamento (5, 6, 7, 8, 9, 10) foram montados 10 conjuntos diferentes (treinamento e teste) para avaliar o comportamento dos resultados do algoritmo embasado nos conceitos estatísticos de distribuição normal das amostras, portanto os resultados apresentados são uma estimativa para as diferentes possibilidades na formação aleatória dos conjuntos de treinamento e teste.

O algoritmo foi implementado usando o software Matlab 6.5 em um computador pessoal Pentium IV - CPU 2.4GHz com 512MB de RAM e sistema operacional Microsoft Windows 2000.

5.4 Pré-Processamento das Imagens.

As imagens de face passaram por um filtro da mediana com vizinhança 5×5 e foram reduzidas para o tamanho de 46×56 . Buscando melhorar o contraste entre os níveis de cinza das imagens ressaltando o relevo da face foi realizado um pré-processamento para normalizar a intensidade dos níveis de cinza através de uma equalização do histograma. A figura 5.5 mostra a equalização do histograma efetuada aplicando as técnicas presentes em (GONZALEZ, 1992).

5.5 Descrição da Fase de Treinamento.

O processo pelos quais os resultados foram obtidos deve ser bem compreendido antes de considerar os resultados das experiências. Para visualizar este processo são apresentados

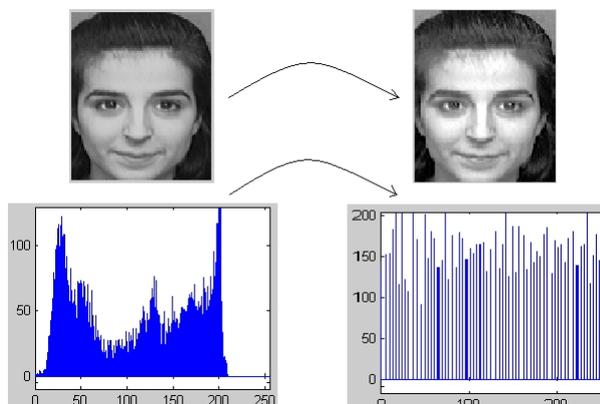


Figura 5.5: Equalização do Histograma de uma face.

as etapas e alguns resultados conseguidos utilizando o banco de dados ORL com 8 faces de treino para cada pessoa e 12 faces de teste.

Para a geração das autofaces, deve-se calcular as bases do espaço PCA através dos vetores de face de treinamento. Os vetores da base computados pelo PCA estão no sentido da maior variação dos vetores de treinamento. Esses vetores da base são computados pela solução de um problema de autovalor, onde os vetores da base PCA são autovetores da matriz de covariância dos pixels das faces de treinamento. Os autovetores são computados através de uma rede neural PCA Adaptativa. Os parâmetros da rede são:

Eta (η)= 0.02 - valor do parâmetro de aprendizagem hebbiana.

Micro (μ)= 0.05 - valor do parâmetro de aprendizagem anti-hebbiana.

Beta (β)= 0.01 - valor do parâmetro momentum.

Número de épocas= 1000

A figura 5.6 mostra um gráfico relacionado a convergência da soma dos pesos laterais da rede tendendo a zero. A rede é estável e os pesos relacionados às conexões entre neurônios tendem aos autovetores que representam as principais componentes relacionadas aos dados de entrada, de acordo com a teoria apresentada no capítulo 4. A rede PCA Adaptativa facilita a busca das primeiras principais componentes, mas os parâmetros da rede devem ser definidos adequadamente para ocorrer a convergência da rede com um número baixo de épocas para o treinamento.

Esses autovetores são definidos no espaço da imagem, e podem ser vistos como imagens e certamente parecem faces, por isso são chamados autofaces (ver figura 5.8).

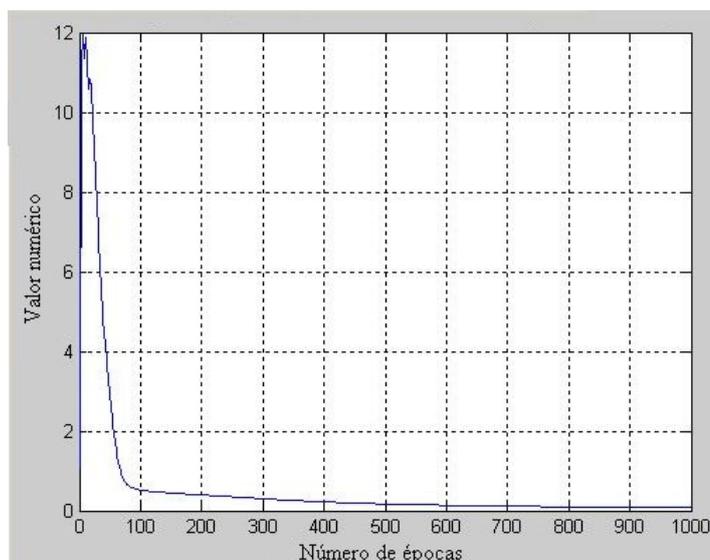


Figura 5.6: Convergência - gráfica da soma dos pesos laterais da rede (8 faces de treino).



Figura 5.7: Face média.

A figura 5.7 apresenta a face média, e a figura 5.8 mostra as primeiras autofaces, que representam as variações das faces de treinamento com relação a média. A autoface apresentada pela face média é um ótimo filtro para faces, pois a multiplicação deste vetor por um vetor de face (produto interno entre dois vetores) resulta um número perto de um, enquanto que vetores de imagens não face apresentam como resultado do produto interno um número distante de um. A primeira autoface representa a maior variação com relação à média (dos vetores de treinamento), a segunda autoface representa a segunda maior variação com relação à média, e assim por diante.

Uma característica interessante inerente ao método PCA pode ser observada na figura 5.9, onde é factual notar que os últimos autovetores não carregam informações importantes para o reconhecimento, pois seus valores referem-se a detalhes específicos de certas imagens do banco de treinamento. Se esses autovetores forem incluídos no modelo ocorrerá perda de generalidade (faces de um mesmo indivíduo com características diferentes podem não ser associadas ao mesmo), mas esses detalhes específicos podem ser impor-

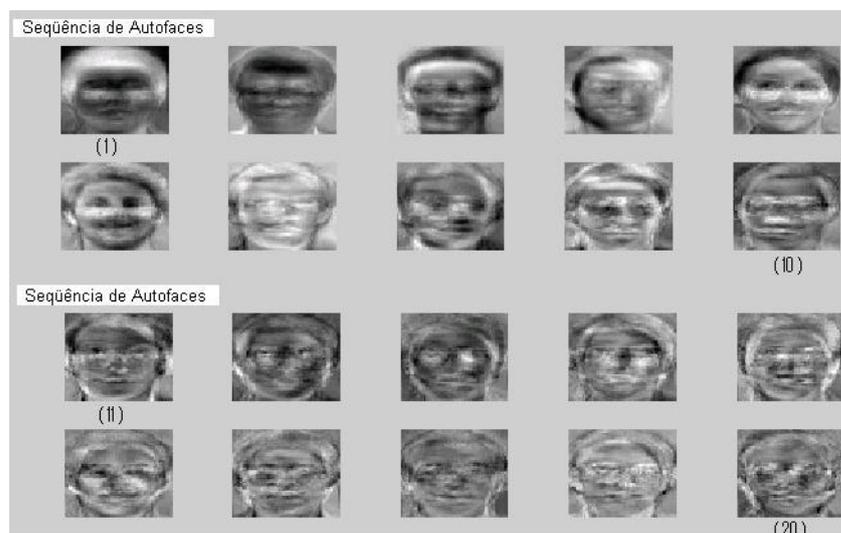


Figura 5.8: Conjunto de autofaces (autovetores) colocados em ordem de 1 a 20.

tantes para aumentar a similaridade entre as faces de teste e as faces de treinamento possibilitando um *threshold* mais estreito e preciso para cada classe (melhorando a precisão do algoritmo, mas rejeitando muitos indivíduos que estavam presentes no treinamento).

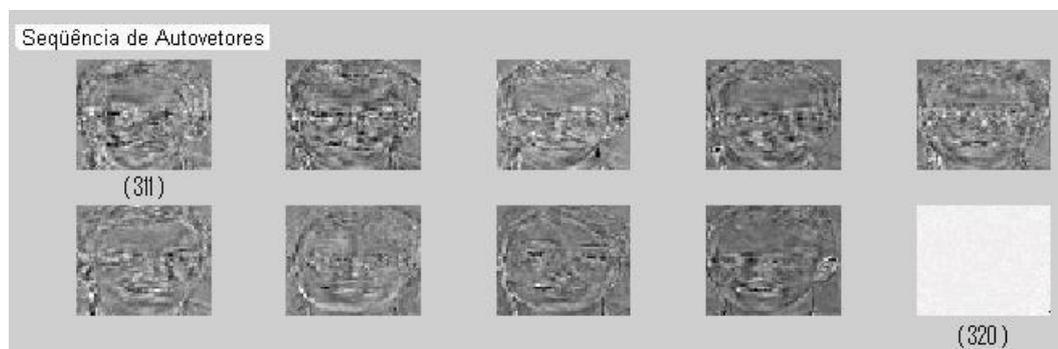


Figura 5.9: Autovetores de 311 a 320.

Se analisarmos o valor numérico dos autovalores (figura 5.10) é fácil concluir que realmente os primeiros autovetores relacionados com os primeiros autovalores possuem as maiores concentrações de informação relacionada ao conjunto de treinamento. Portanto, o grande problema para aplicações reais é definir o ponto de maior eficiência do algoritmo sem conhecer o conjunto de teste.

No espaço dimensional PCA, cada autoface pode ser vista como uma característica. Quando uma face particular é projetada no sub-espaço PCA, seu novo vetor representativo é composto de seus valores de peso com relação às bases do sub-espaço, portanto, os

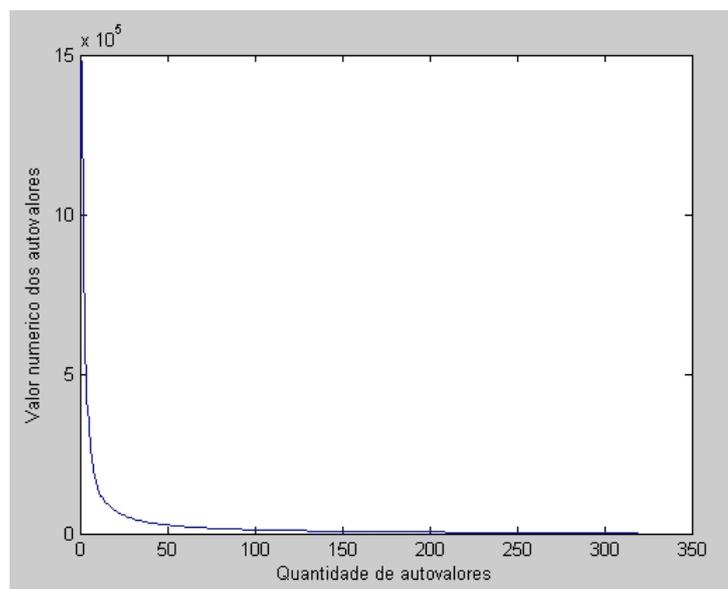


Figura 5.10: Visualização dos Autovalores.

pesos descrevem a importância de cada autovetor na caracterização da face. A figura 5.11 descreve este processo detalhadamente.

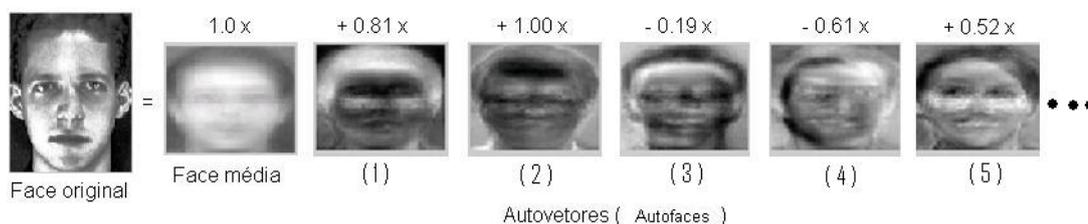


Figura 5.11: Composição de uma face através de seus pesos e respectivas autofaces.

Neste exemplo, uma face é projetada no sub-espço PCA, sendo descrita no sub-espço por seus coeficientes (ou pesos) relacionados aos autovetores, e pode facilmente ser obtida novamente no espço original de faces através de uma combinação linear dos autovetores somados com a face média.

Após o cálculo dos coeficientes relacionados a cada face projetada no sub-espço PCA é realizada a etapa de treinamento do módulo LDA. A etapa LDA busca encontrar um eixo de coordenadas ótimo no sub-espço para representar as faces com a maior distância entre as médias das classes e com a menor variância dentro das classes. Após criado o sub-espço LDA, as faces são projetadas no sub-espço e seus coeficientes são utilizados para realizar a etapa de treinamento do classificador (rede SOM - Kohonen). Os pesos

de uma mesma classe (ou os coeficientes LDA das faces de um mesmo indivíduo no treinamento) alimentam uma única rede neural SOM representante da classe.

A identificação da classe similar à imagem de entrada é efetuada pelo estágio final composto pelas redes SOM, mas deve-se estabelecer a quantidade ótima de autovetores presentes para compor os sub-espacos PCA e LDA, pois se estes estágios não apresentarem um bom resultado para o reconhecimento eles podem acarretar erros, ou inserirem ruídos para o estágio seguinte de classificação.

5.6 Testes Realizados.

Os testes realizados têm por objetivo avaliar as interações entre as variáveis inerentes ao PCA, LDA e redes SOM na obtenção das taxas de reconhecimento globais para o algoritmo proposto. Todas as etapas envolvidas no processo foram avaliadas individualmente, mas sempre utilizando o ponto ótimo das etapas anteriores. Para analisar o desempenho do algoritmo global foram definidas as variáveis ótimas para cada etapa, e avaliada as taxas de reconhecimento juntamente com a atualização automática constante dos neurônios das redes SOM e a busca por similaridade no banco de dados.

5.6.1 Avaliação de Desempenho (PCA e LDA).

As etapas PCA e LDA são importantes para construir o algoritmo com um alto desempenho no reconhecimento das faces, pois as duas técnicas representam o núcleo do algoritmo fazendo a compressão e representação ótima dos dados (PCA), e ressaltando as diferenças entre as classes para aumentar o desempenho na identificação (LDA). A escolha dos parâmetros (quantidade de faces de treinamento e quantidade de autovetores para representar os subespaços) adequados é fundamental para incrementar a correlação entre as etapas PCA e LDA na busca de taxas de reconhecimento de faces expressivas com alto desempenho e imunidades às pequenas mudanças entre as faces de teste e de treinamento armazenadas no banco de dados. Portanto se faz necessário avaliar o desempenho do algoritmo com relação às diferentes possibilidades para os parâmetros PCA e LDA avaliando-se qual é a interação entre eles.

Para avaliar a interação entre os parâmetros PCA e LDA foi proposto analisar o comportamento de desempenho do algoritmo PCA e LDA escolhendo-se o intervalo de autovetores, que definam 60 a 99% de energia na construção do subespaço PCA e 70 a 99%

para o LDA variando-se o número de autovetores de 2 em 2 para o PCA e todos os autovetores contidos no intervalo de energia LDA. Deve-se notar que cada subespaço PCA apresentado no intervalo de energia gera mais uma quantidade enorme de subespaços relacionados ao intervalo de energia LDA (veja figura 5.12).

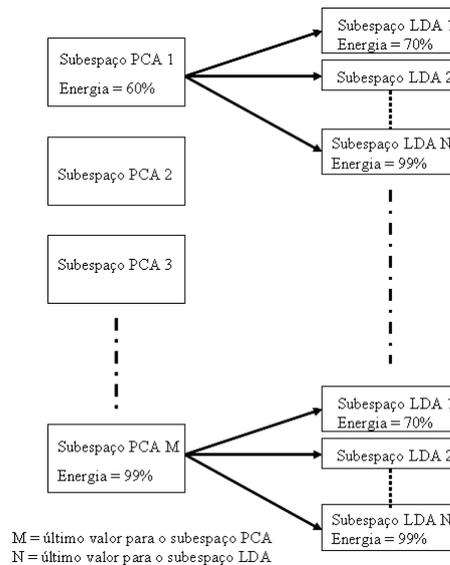


Figura 5.12: Diagrama das possíveis combinações PCA e LDA.

Para a análise dos resultados foi elaborada uma seleção aleatória com possibilidades iguais entre as 20 faces para cada indivíduo perfazendo diferentes grupos de teste e treinamento submetidos ao algoritmo para levantar as curvas de rendimento e interação entre o PCA e o LDA. Foram avaliadas 6 diferentes combinações de faces selecionadas para o treinamento formando grupos com 5, 6, 7, 8, 9 e 10 faces de cada indivíduo com o restante sendo utilizado para montar o grupo de teste. Buscando aumentar a confiança dos testes, para cada combinação foi realizado o mesmo teste 10 vezes possibilitando construir um resultado médio com um intervalo de confiança de 95% supondo uma distribuição normal (*t-student*) para a seleção aleatória das faces com relação às taxas de reconhecimento.

Nas figuras 5.13 a 5.18 são apresentadas às taxas de reconhecimento obtidas para o PCA + LDA variando-se a quantidade de autovetores PCA para formar o subespaço com os limites de energia pré-determinados anteriormente, e utilizando a distância euclidiana como classificador. Cada subespaço PCA gera uma variedade de combinações de subespaços LDA representados pelos autovetores LDA contidos no intervalo de energia pré-determinado, mas os gráficos apresentam apenas as combinações que resultaram em uma máxima taxa de reconhecimento para cada subespaço PCA.

As linhas cheias apresentam as taxas médias de reconhecimento, enquanto que as tracejadas indicam o intervalo de confiança para os resultados médios apresentados nas figuras.

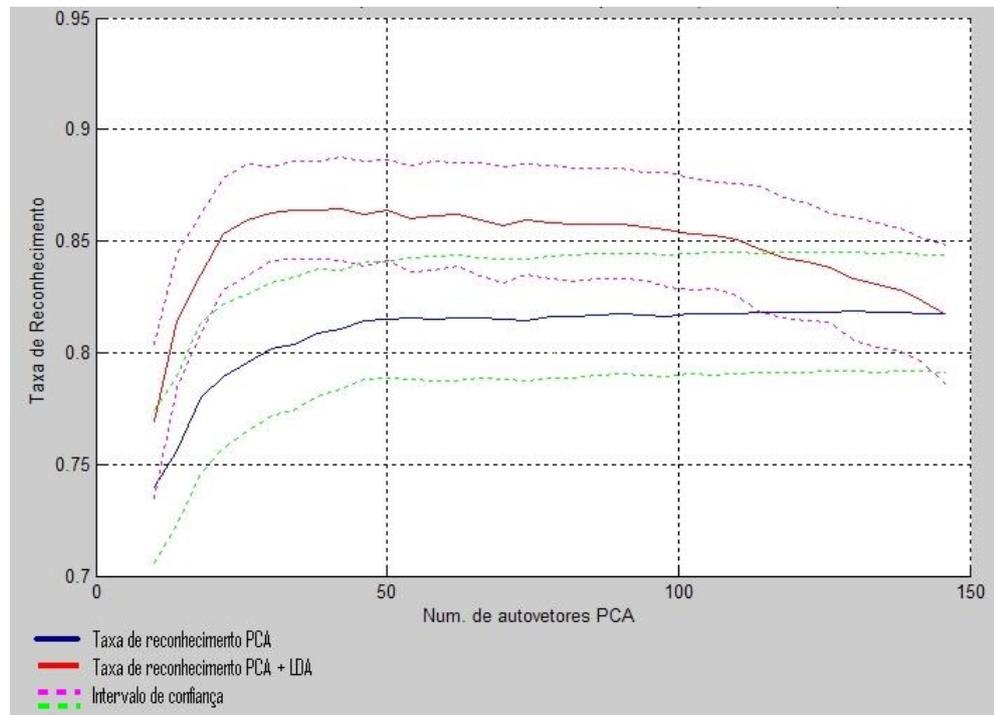


Figura 5.13: Desempenho PCA e máximo desempenho LDA (5 faces de treino).

Analisando os gráficos das figuras 5.13 a 5.18 pode-se obter algumas características relacionadas ao algoritmo aplicado ao banco de dados proposto seguindo o detalhamento:

- Para todos os grupos de teste, os resultados relacionados à etapa PCA apresentam um fase transitória seguida por uma fase com taxas de reconhecimento estacionárias podendo decrescer com o aumento da quantidade de autovetores no subespaço.

Os últimos autovetores não carregam informações significativas para o grupo de treinamento global, mas sim detalhes específicos de cada face podendo ser considerados como ruído que afeta a generalização e a capacidade de reconhecimento do algoritmo.

- Relativamente, a etapa PCA alcança a convergência com poucos autovetores provando a alta capacidade de compressão do método; um fator muito importante para a elaboração de projetos que envolvem grandes quantidades de faces.

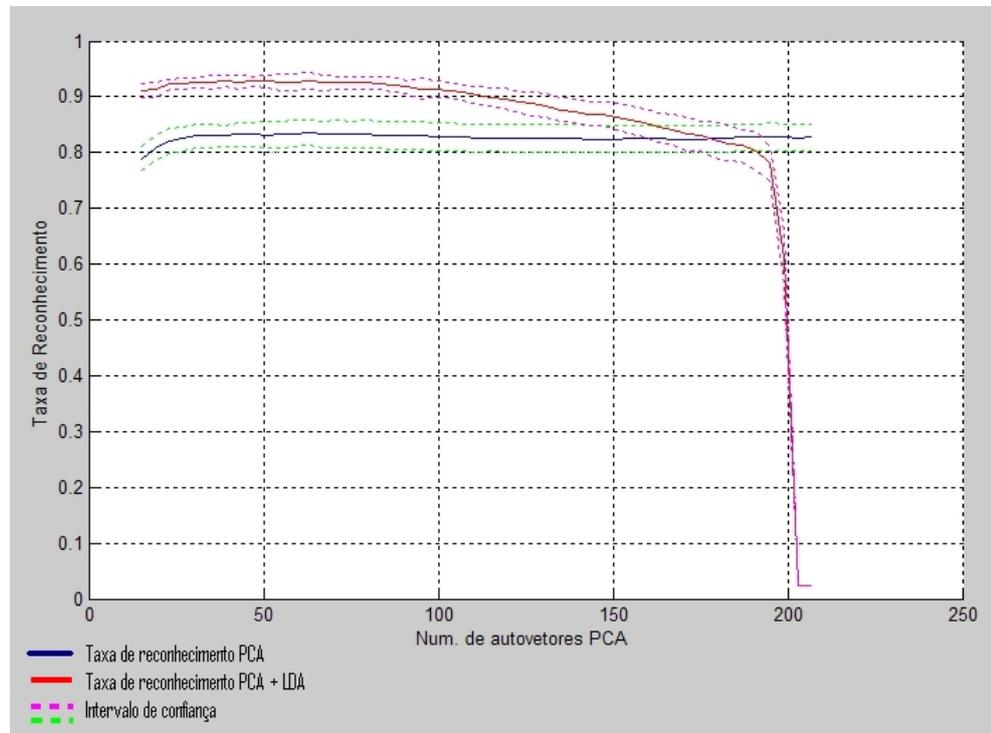


Figura 5.14: Desempenho PCA e máximo desempenho LDA (6 faces de treino).

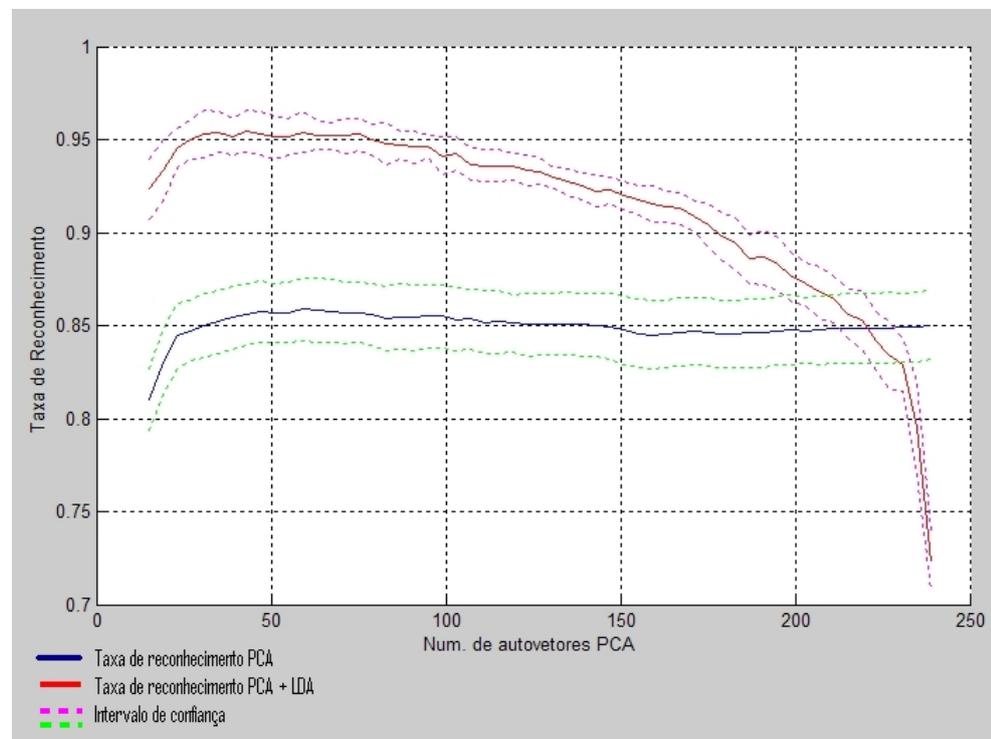


Figura 5.15: Desempenho PCA e máximo desempenho LDA (7 faces de treino).

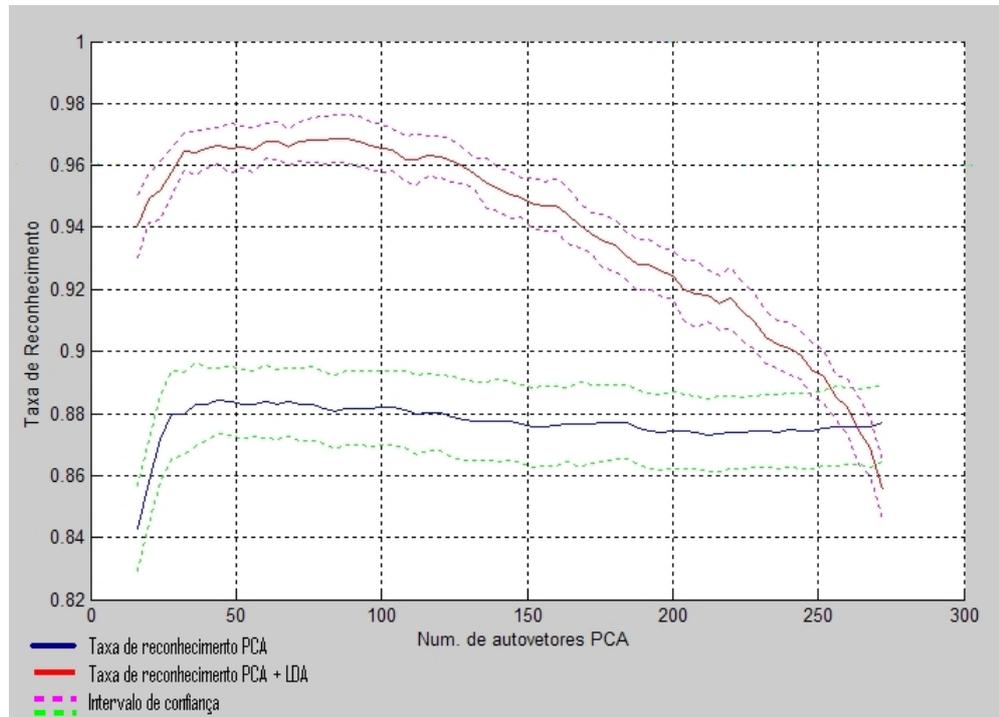


Figura 5.16: Desempenho PCA e máximo desempenho LDA (8 faces de treino).

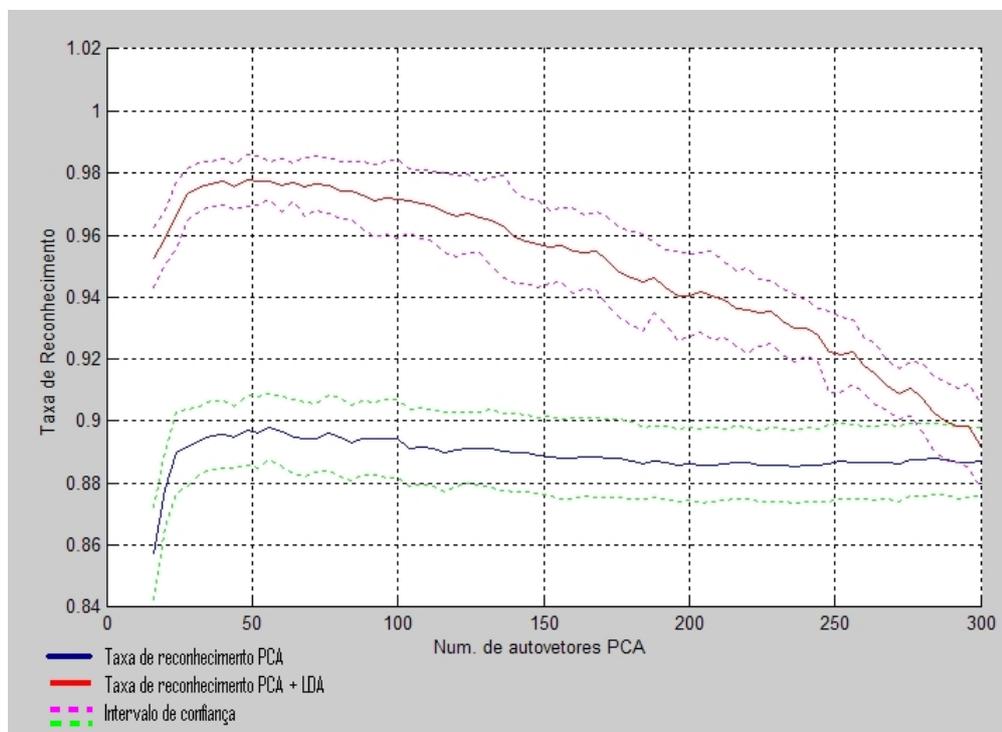


Figura 5.17: Desempenho PCA e máximo desempenho LDA (9 faces de treino).

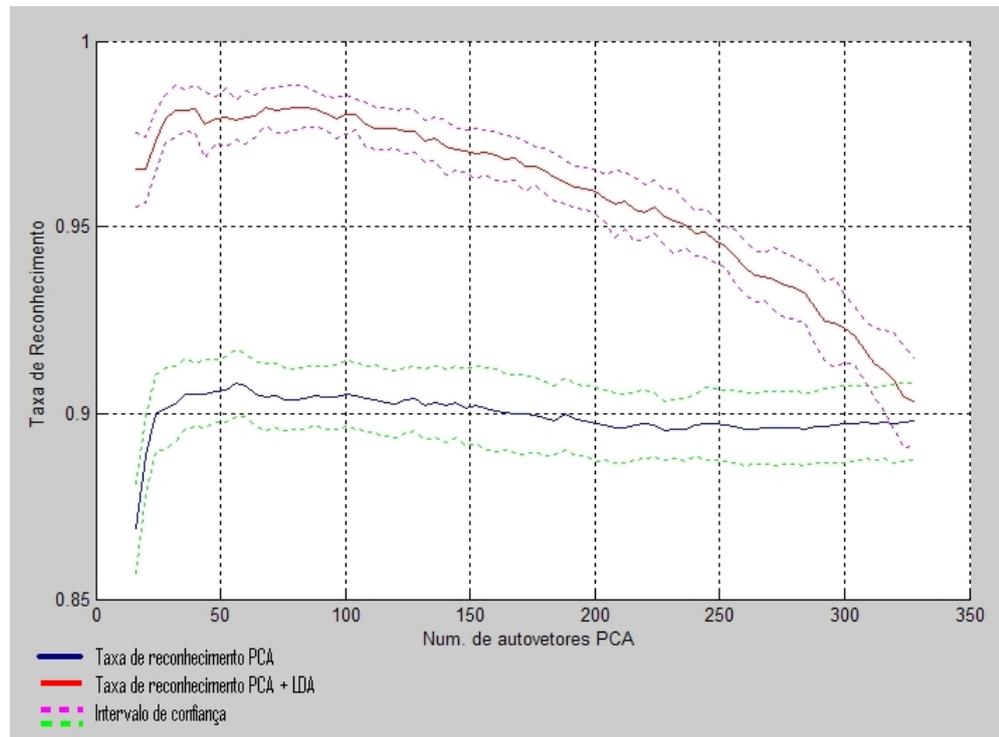


Figura 5.18: Desempenho PCA e máximo desempenho LDA (10 faces de treino).

- A etapa LDA mostrou-se capaz de aumentar as taxas de reconhecimento com relação ao PCA, lembrando que: quanto mais altas (próximas de um) são as taxas de reconhecimento de uma etapa, maiores serão as dificuldades enfrentadas pelas próximas etapas para incrementar em percentuais significativos os resultados anteriores.
- Uma deficiência intrínseca relacionada ao método LDA pode ser observada na prática; quando se tem poucas amostras de treinamento para cada classe de indivíduos, a matriz S_W (ou variância dentro das classes) pode se tornar singular dependendo do conjunto de treinamento, e gerar uma instabilidade nos resultados do método. As figuras 5.14 e 5.15 mostram uma queda abrupta nas taxas de reconhecimento quando a quantidade de autovetores PCA, que formam os coeficientes de projeção para as faces, aumenta gerando um espaço dimensional maior na entrada LDA com uma quantidade de amostras pequena de treinamento.
- Para o banco de dados utilizado, um aumento na quantidade de autovetores (ou energia) no subespaço PCA não significa um aumento de rendimento nos resultados da etapa LDA, mas sim um decréscimo, pois o aumento nos autovetores PCA

causa uma perda de generalidade e leva S_W próximo à singularidade afetando o desempenho global do algoritmo.

- É evidente notar nos gráficos que um aumento no grupo de treinamento leva o algoritmo a obter taxas de reconhecimento próximas de 1 (ou 100%). Nos gráficos das figuras 5.16, 5.17 e 5.18 tem-se pontos com taxas de reconhecimento próximas ou superiores a 98%.
- Uma característica nos gráficos que chama a atenção é notar que os picos das taxas de reconhecimento para o método PCA não são os mesmos para o LDA, mas estão próximos, e logo após a região transitória observada.

5.6.2 Avaliação do Desempenho LDA (ou sub-espaço LDA).

A quantidade de autovetores LDA para a construção do subespaço é um fator importante para se analisar, pois os autovetores PCA e LDA deveriam ser armazenados pelo modelo para possibilitar efetuar a projeção e classificação das faces. Deve haver um equilíbrio entre a taxa de reconhecimento e a quantidade de autovetores, para que o algoritmo não ocupe muita memória de armazenamento e apresente robustez em aplicações que exijam tomada de decisão em tempo real.

Para analisar o comportamento das taxas de reconhecimento com relação aos autovetores LDA (de acordo com os limites de concentração da energia no subespaço), foram escolhidos alguns subespaços PCA próximos à região de maiores taxas de reconhecimento do algoritmo. As figuras 5.19 a 5.24 apresentam os resultados para os diferentes grupos de testes e as quantidades de autovetores PCA que constituem os respectivos subespaços.

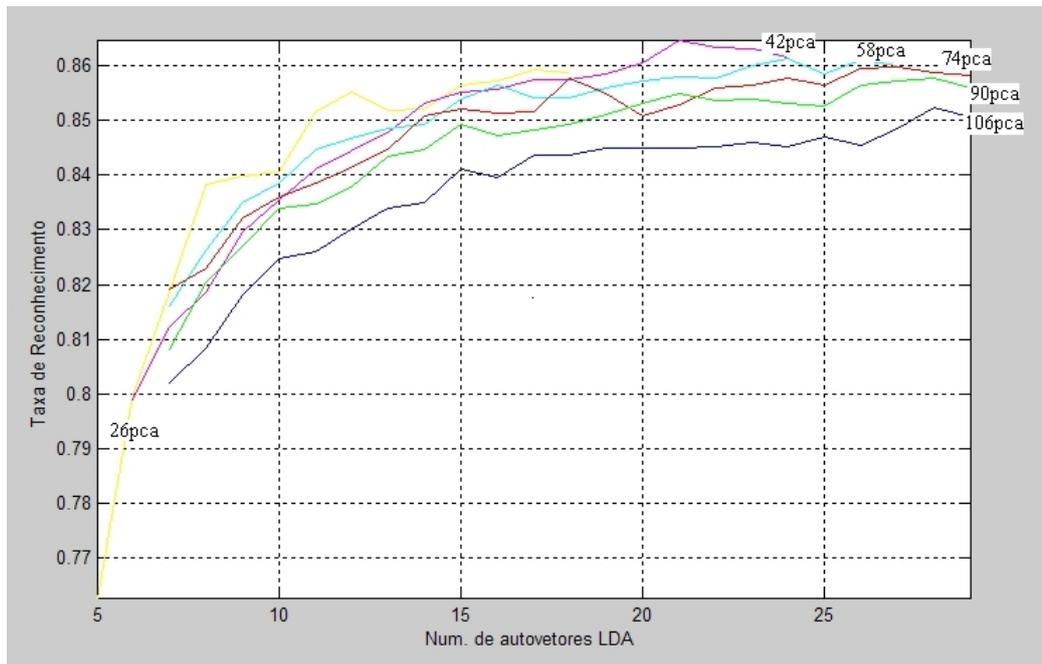


Figura 5.19: Desempenho LDA (5 faces de treino).

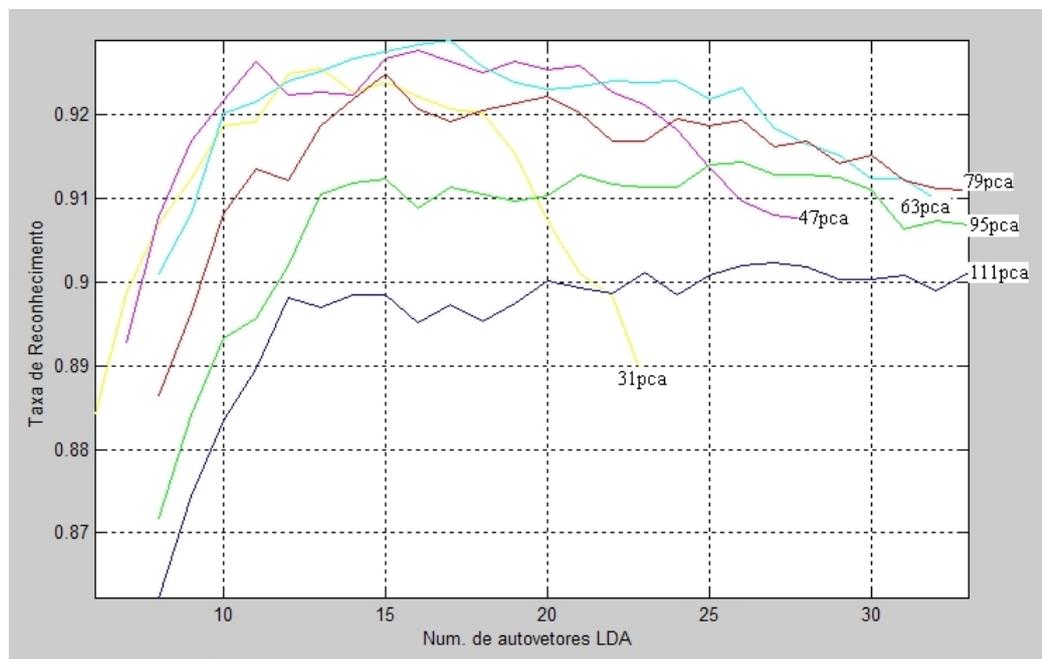


Figura 5.20: Desempenho LDA (6 faces de treino).

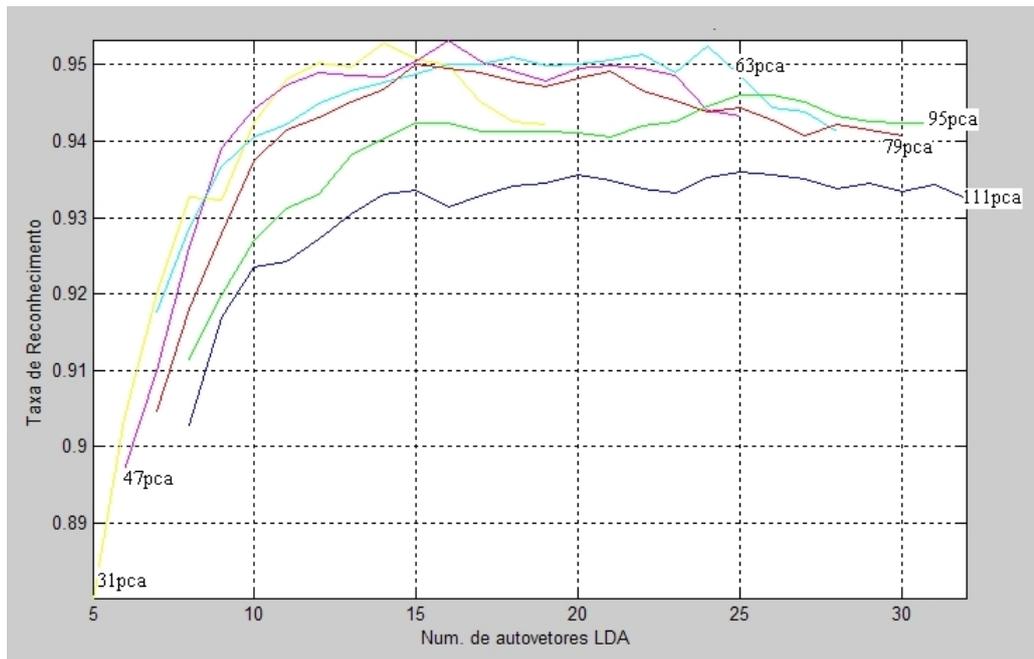


Figura 5.21: Desempenho LDA (7 faces de treino).

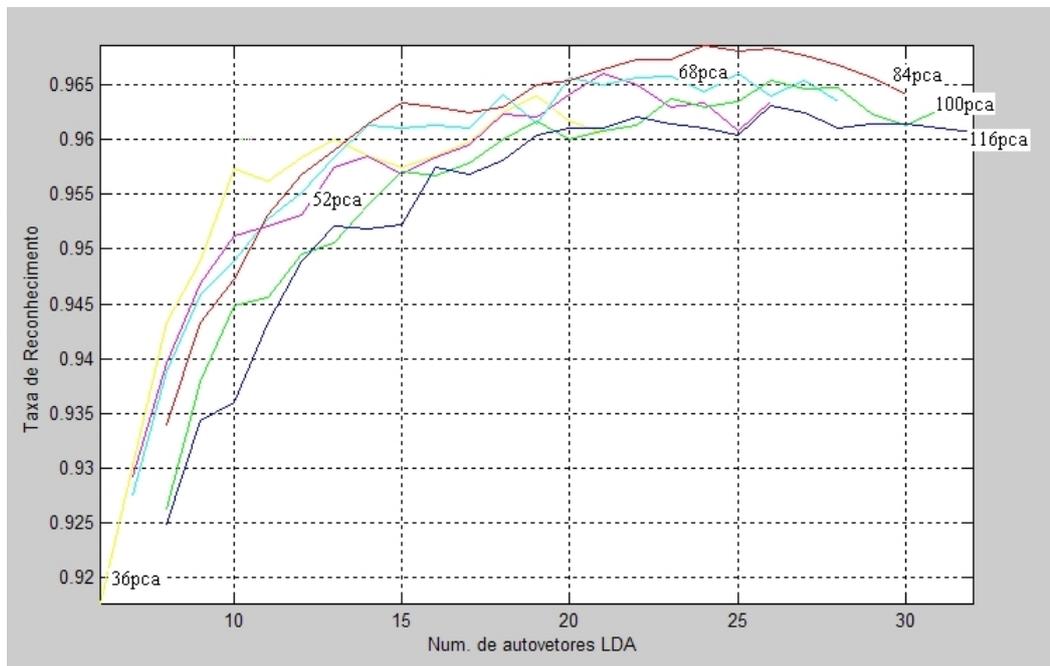


Figura 5.22: Desempenho LDA (8 faces de treino).

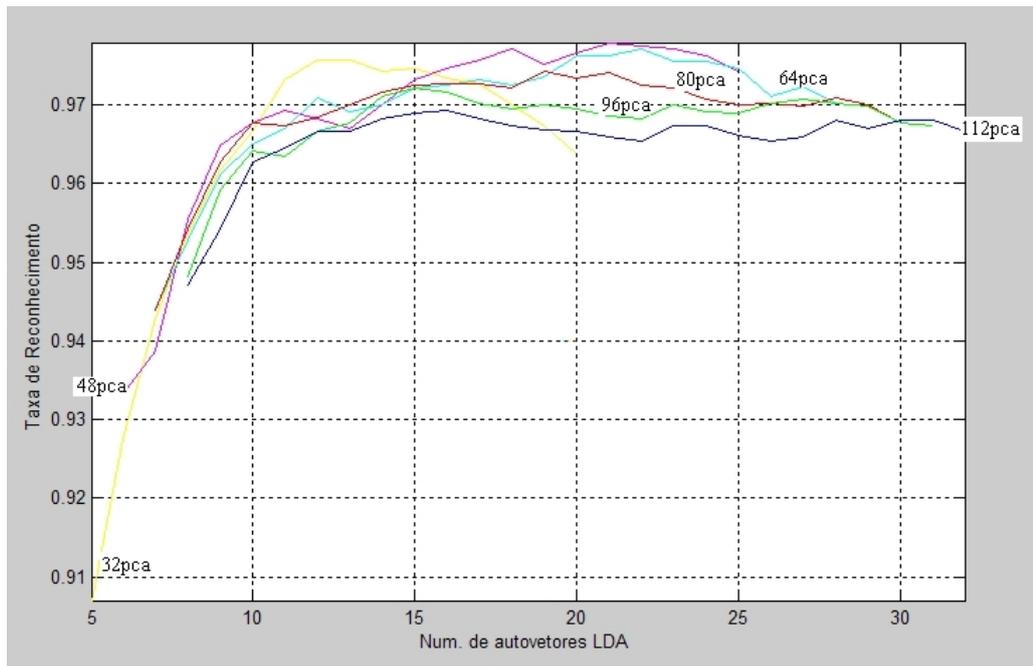


Figura 5.23: Desempenho LDA (9 faces de treino).

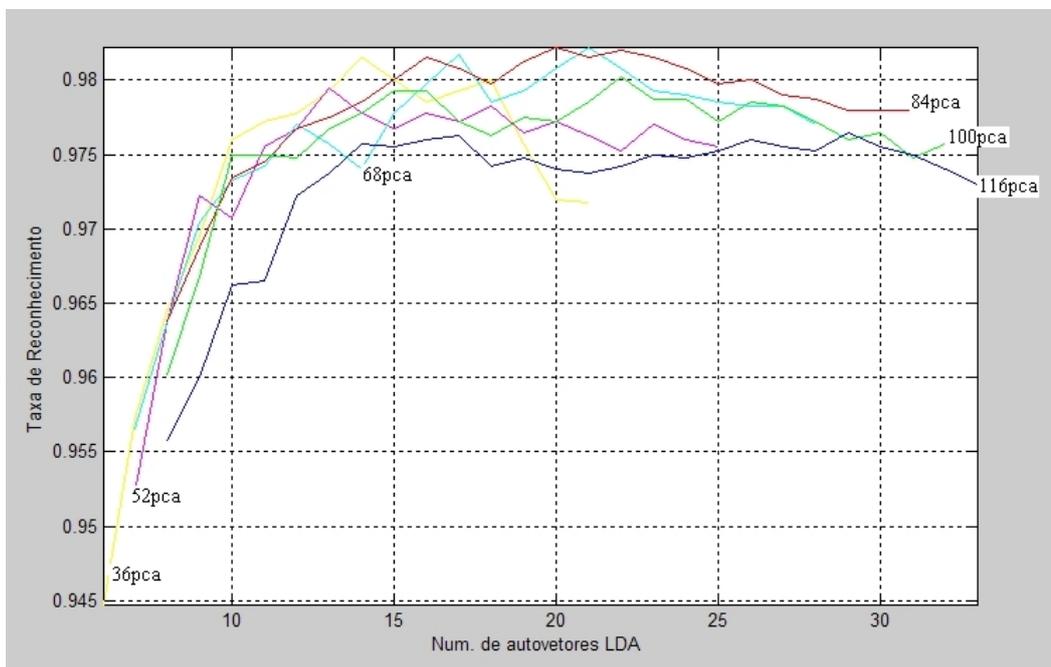


Figura 5.24: Desempenho LDA (10 faces de treino).

Aparentemente, o LDA apresenta taxas de reconhecimento mais estáveis quando a quantidade de autovetores PCA é maior, mas as taxas são significativamente mais baixas e o gasto com armazenamento é maior. Para todos os testes, boas taxas de reconhecimento

são conseguidas com a quantidade de autovetores LDA entre 15 e 25, ou seja, ocorre uma maior compressão dos dados e paralelamente um aumento na generalidade e desempenho do algoritmo.

5.6.3 Avaliação do Desempenho da Rede Neural SOM (Kohonen).

A última etapa do algoritmo é o classificador, que neste trabalho foi utilizado um conjunto de redes SOM (Kohonen) treinadas com as saídas da etapa LDA representativas de cada classe. As redes SOM podem possibilitar uma atualização ordenada dos neurônios representantes de cada classe no banco de dados de acordo com um *threshold* estabelecido para a similaridade entre o teste e o neurônio vencedor. Em aplicações reais as faces de teste podem variar por algum motivo (intensidade de luz do local, envelhecimento da pessoa, ...etc) das faces de treinamento, e se apresentarem intermitentemente com essas características ao modelo, portanto há a necessidade de se modificar os pesos dos neurônios na base de dados para se aproximarem ordenadamente da face teste.

Para avaliar o desempenho das redes SOM e observar se há algum decréscimo nas taxas de reconhecimento foi escolhido o grupo com 8 faces de teste, e a combinação de subespaços PCA e LDA com maiores taxas de reconhecimento. O subespaço PCA foi definido com 84 autovetores e o LDA foi variável, e os resultados comparativos entre os algoritmos PCA+LDA e PCA+LDA+SOM são mostrados nas figuras 5.25 a 5.29 com diferentes configurações para a estrutura das redes SOM.

Na figura 5.25, tem-se os seguintes parâmetros para a rede:

- Estrutura topológica da rede: rede de Kohonen linear
- Atualização da vizinhança: gaussiana
- Quantidade de neurônios: 4
- Quantidade de épocas: 450

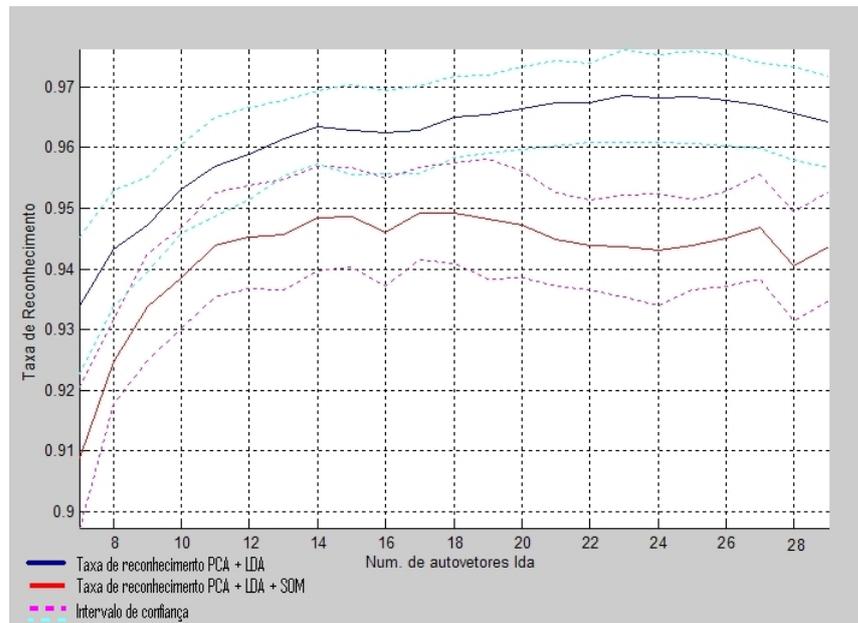


Figura 5.25: Desempenho das redes SOM (8 faces de treino e 4 neurônios).

Na figura 5.26, tem-se os seguintes parâmetros para a rede:

- Estrutura topológica da rede: rede de Kohonen linear
- Atualização da vizinhança: gaussiana
- Quantidade de neurônios: 8
- Quantidade de épocas: 450

Na figura 5.27, tem-se os seguintes parâmetros para a rede:

- Estrutura topológica da rede: rede de Kohonen linear
- Atualização da vizinhança: gaussiana
- Quantidade de neurônios: 12
- Quantidade de épocas: 450

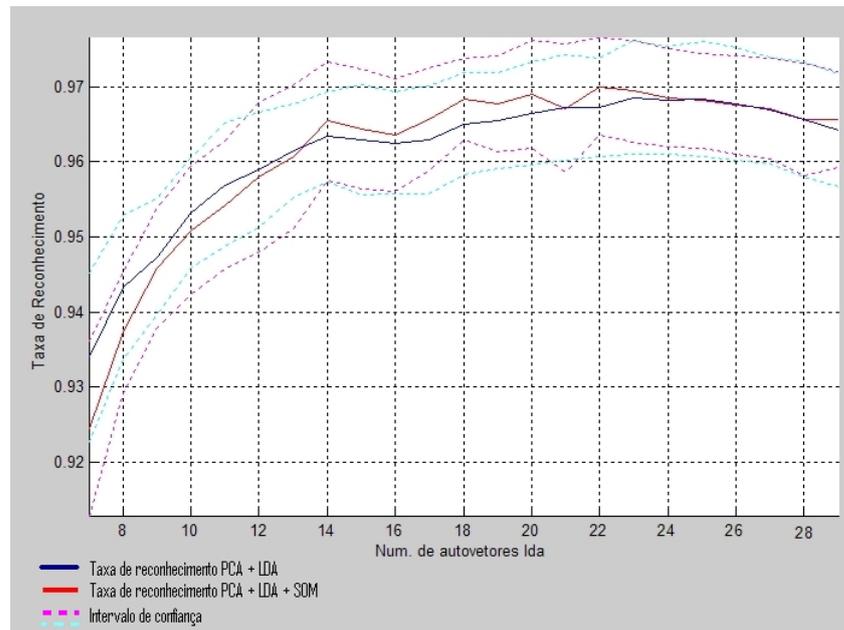


Figura 5.26: Desempenho das redes SOM (8 faces de treino e 8 neurônios).

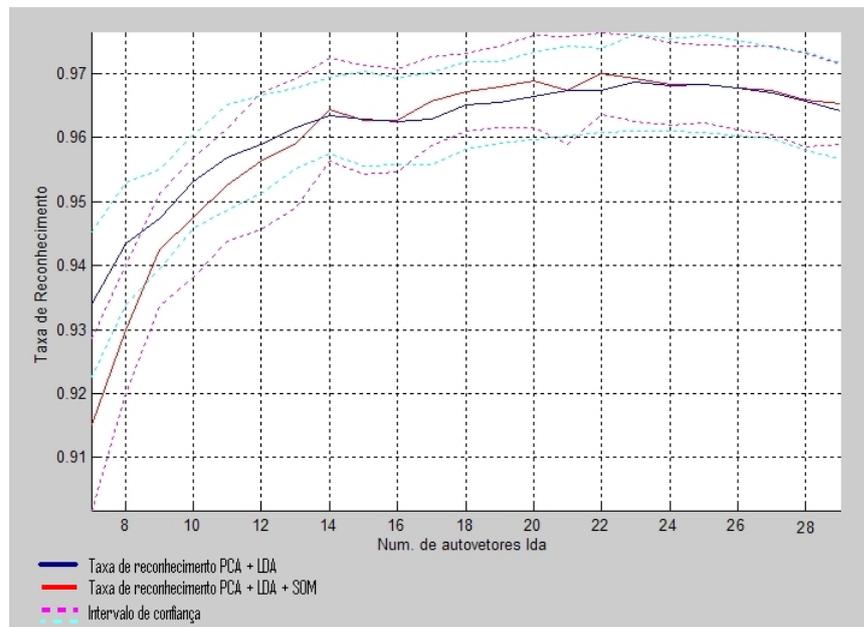


Figura 5.27: Desempenho das redes SOM (8 faces de treino e 12 neurônios).

Na figura 5.28, tem-se os seguintes parâmetros para a rede:

- Estrutura topológica da rede: rede de Kohonen quadrática
- Atualização da vizinhança: gaussiana

- Quantidade de neurônios: 3x3
- Quantidade de épocas: 450

Na figura 5.29, tem-se os seguintes parâmetros para a rede:

- Estrutura topológica da rede: rede de Kohonen quadrática
- Atualização da vizinhança: gaussiana
- Quantidade de neurônios: 4x4
- Quantidade de épocas: 450

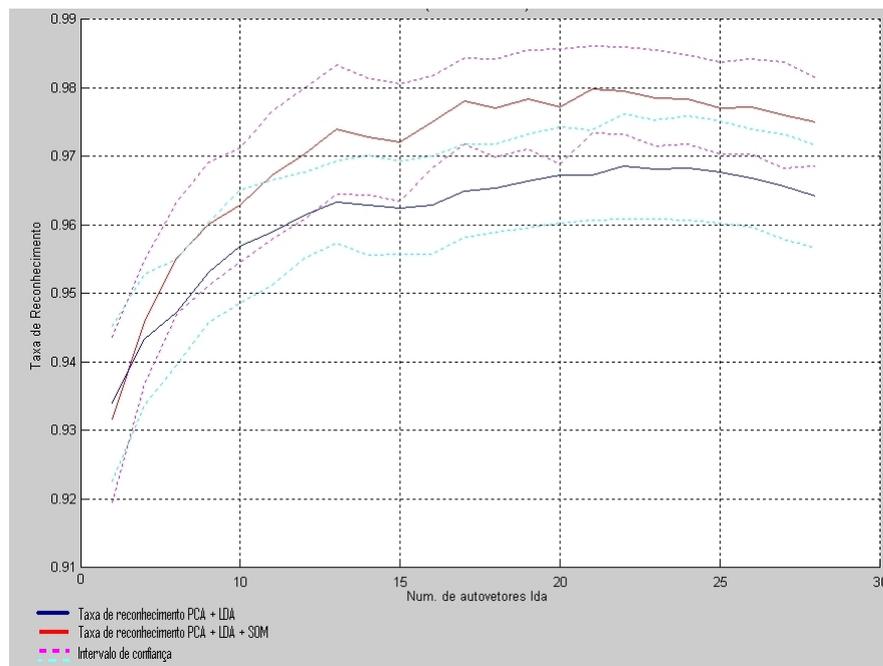


Figura 5.28: Desempenho das redes SOM (8 faces de treino e 3x3 neurônios).

As redes SOM abrem a possibilidade de se diminuir a quantidade de coeficientes na representação de cada face de treinamento, mas ocorre uma pequena perda na taxa de reconhecimento como foi mostrado na figura 5.25. Para os casos onde a quantidade de neurônios é superior à quantidade de coeficientes LDA tem-se um pequeno acréscimo nas taxas de reconhecimento, sendo importante ressaltar que ligeiros acréscimos nas taxas quando se está próximo do reconhecimento total (igual a 1 ou 100%) é difícil. Quanto à topologia da rede, a rede quadrática mostrou-se um pouco melhor que a linear.

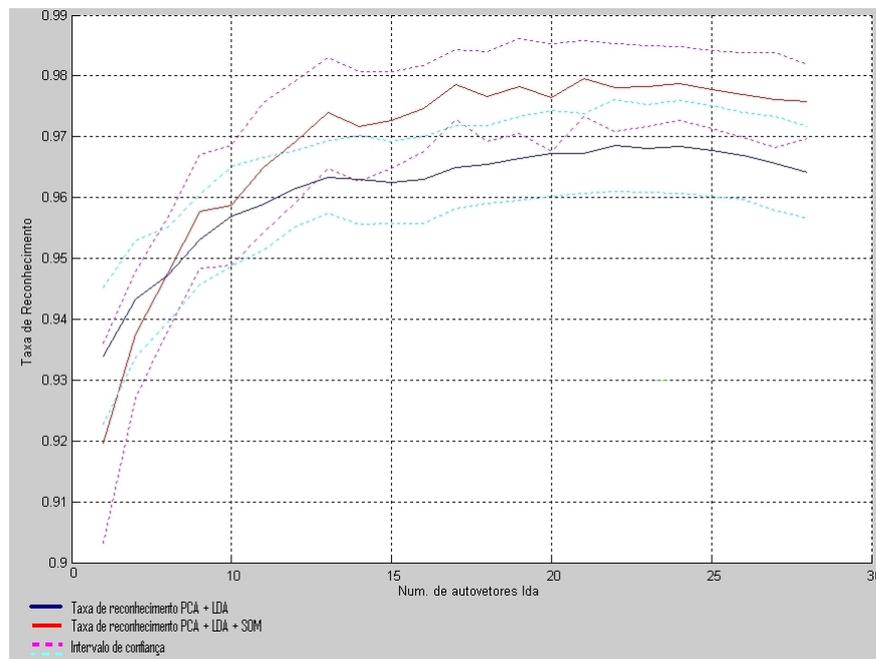


Figura 5.29: Desempenho das redes SOM (8 faces de treino e 4x4 neurônios).

5.6.4 Definição dos Subespaços PCA e LDA.

Qual a quantidade de autovetores ótima para a representação dos subespaços PCA e LDA? Analisando-se os testes anteriores é possível identificar combinações entre as quantidades de autovetores PCA e LDA, que maximizam as taxas de reconhecimento para o conjunto de faces em questão. A dificuldade na maioria das aplicações reais está definir o ponto ótimo dos subespaços sem conhecer o conjunto de teste tendo-se apenas as informações sobre o conjunto de treinamento. Mas se o conjunto de treinamento contém muitas faces de cada pessoa (> 10) pode-se dividi-lo formalizando-se um grupo de teste para a calibração do algoritmo.

Considerando-se apenas as informações obtidas através do banco de faces de treinamento, uma opção para definir os subespaços seria observar a evolução da quantidade de energia (ou informação) acumulada nos subespaços com o acréscimo dos autovetores associados aos seus autovalores que representam a energia do subespaço. As figuras 5.30 a 5.35 mostram a evolução da energia acumulada nos subespaços com diferentes quantidades de faces no agrupamento de treinamento.

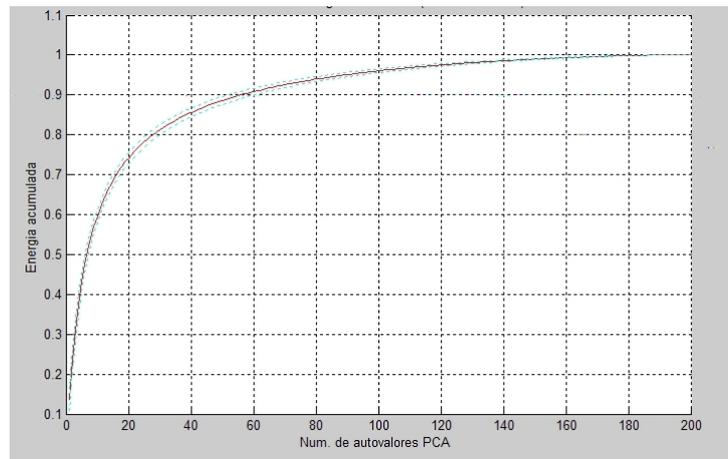


Figura 5.30: Evolução da energia acumulada (5 faces de treino).

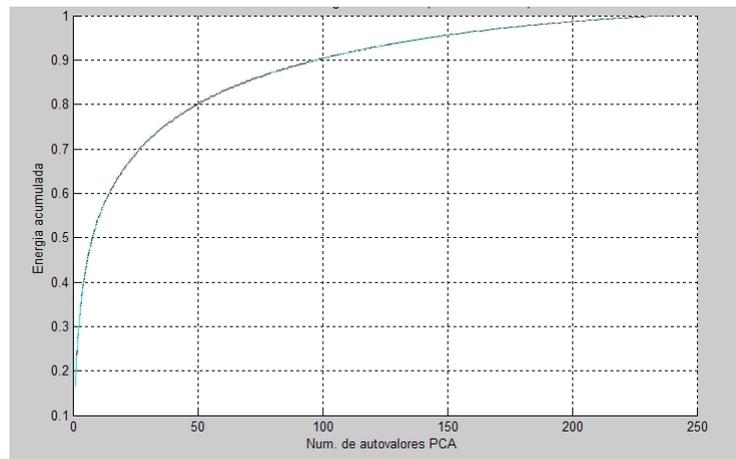


Figura 5.31: Evolução da energia acumulada (6 faces de treino).

Avaliando-se o espalhamento da energia com relação à quantidade de autovetores no subespaço PCA, e observando-se os gráficos das taxas de reconhecimento mostrados anteriormente pode-se chegar a algumas conclusões. Os valores máximos nas taxas de reconhecimento estão entre os 50 e 100 primeiros autovetores e nesse intervalo a energia acumulada do subespaço está próxima de 80 a 85% com o intervalo de confiança para a média sobreposto à mesma, além disso, nesse intervalo as taxas de reconhecimento estão em um regime estacionário com leves flutuações.

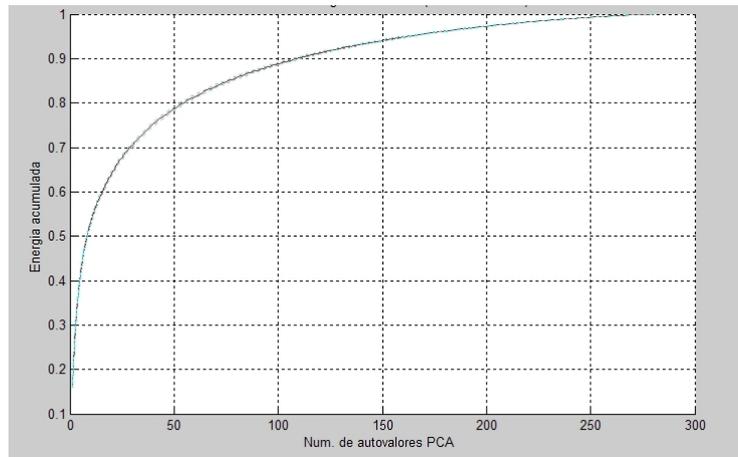


Figura 5.32: Evolução da energia acumulada (7 faces de treino).

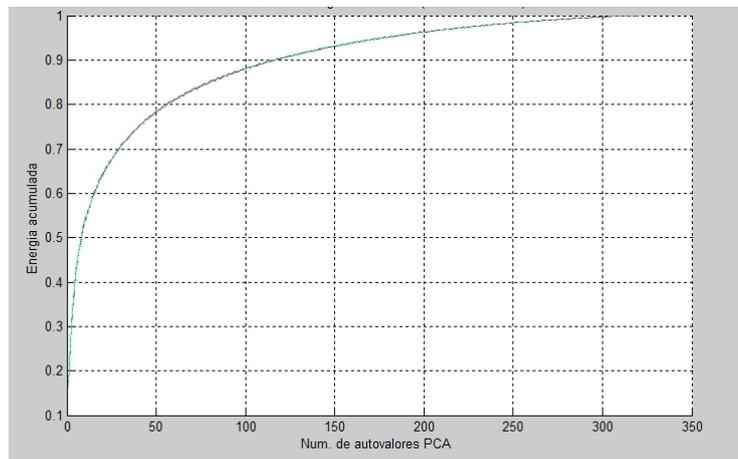


Figura 5.33: Evolução da energia acumulada (8 faces de treino).

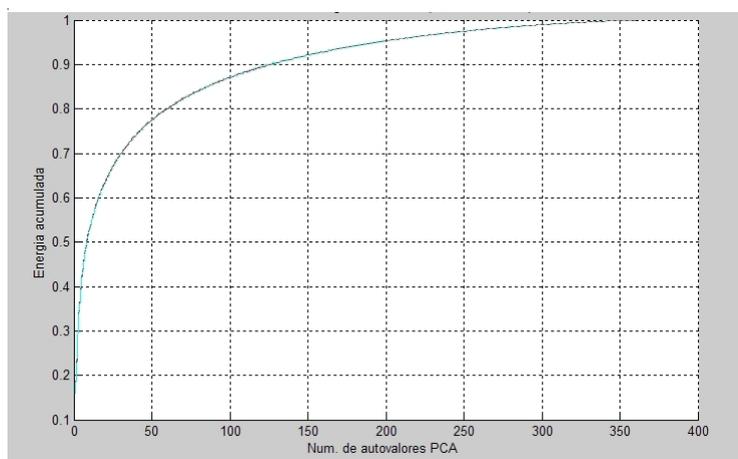


Figura 5.34: Evolução da energia acumulada (9 faces de treino).

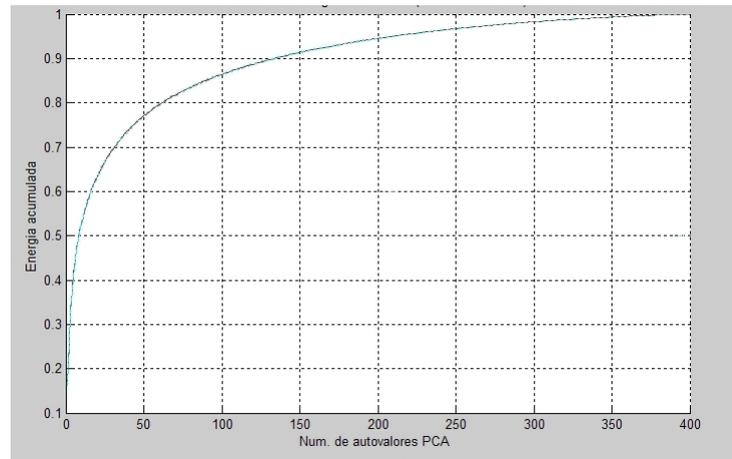


Figura 5.35: Evolução da energia acumulada (10 faces de treino).

Para modelar o subespaço PCA foi estabelecido empiricamente um limite máximo de energia igual a 80% do subespaço total, pois as taxas de reconhecimento estão estáveis e a quantidade de autovalores que define o grau de compressão do método está entre 50 a 75, o que é pequeno quando comparado com a dimensão original do espaço de faces ($48 \times 56 = 2576$). A figura 5.36 apresenta um gráfico com a evolução da energia acumulada no subespaço e as taxas de reconhecimento associadas à quantidade de autovetores para 8 faces de treino, onde é evidente a presença de altas taxas de reconhecimento com a energia próxima a 80%.

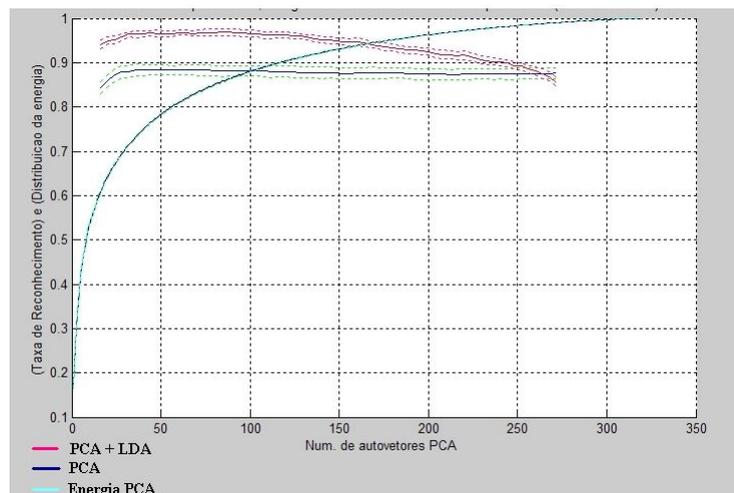


Figura 5.36: Desempenho PCA, energia acumulada e desempenho LDA (8 faces de treino).

5.6.5 Aplicação do Algoritmo.

Os sub-capítulos anteriores oferecem informações experimentais para estabelecer todas as variáveis relacionadas ao algoritmo proposto. Na escolha das variáveis ótimas deve-se atentar aos valores relacionados à quantidade de autovetores na formação dos subespaços e as taxas de reconhecimento, pois um algoritmo ótimo deve apresentar uma alta taxa de compressão com uma boa generalização, ou seja:

- Uma alta taxa de compressão possibilita a redução de dados armazenados, aumenta a velocidade de treinamento e diminui o tempo de resposta do algoritmo para aplicações em tempo real.
- Uma boa generalização aumenta as taxas de reconhecimento quando as faces de teste são ligeiramente diferentes das faces armazenadas no banco de treinamento com relação à iluminação, rotação, escala, uso de óculos, maquiagem, envelhecimento e outros fatores que podem dificultar o reconhecimento.

Para analisar o desempenho do algoritmo com relação à identificação das faces e possíveis erros na classificação foi proposta a seguinte estrutura para os testes:

- 8 faces de treino escolhidas aleatoriamente para cada pessoa, pois com 8 faces a matriz LDA não é singular.
- 25 pessoas foram escolhidas aleatoriamente para calcular o subespaço PCA no treinamento.
- 15 pessoas foram inseridas no conjunto para a etapa de treinamento após o cálculo do subespaço PCA.
- O subespaço LDA sempre deve ser calculado com todas as faces de treinamento das classes (pessoas) envolvidas, pois ele utiliza as informações das classes para estabelecer o melhor subespaço. A cada nova pessoa inserida é necessário fazer uma avaliação do subespaço LDA e das taxas de reconhecimento para essa pessoa, pois pode ser necessário fazer os cálculos para o subespaço novamente.
- O teste é repetido 10 vezes para possibilitar o cálculo de valores médios com um intervalo de confiança de 95%.
- Subespaço PCA com 80% de energia.

- Subespaço LDA com 95% de energia.
- Redes SOM linear com 8 neurônios, pois se deseja manter a quantidade de dados armazenados e com mais neurônios tem-se mais dados em função de um ligeiro aumento na taxa de reconhecimento.

A tabela 5.1 apresenta os resultados para o desempenho do algoritmo em cada etapa. Um resultado interessante é o relacionado com a taxa de reconhecimento para as pessoas que não foram inseridas para o cálculo do subespaço PCA, pois a taxa de reconhecimento para esse grupo foi maior que a taxa referente às próprias classes que foram utilizadas no cálculo do subespaço PCA. Esse resultado ressalta a característica de generalização relacionada ao método PCA, e oferece argumentos para incentivar aplicações com banco de dados grandes (>1000 faces) fazendo a construção do subespaço PCA apenas com as faces mais significantes e relevantes de um grupo de pessoas em uma determinada região ou raça.

Tabela 5.1: Taxas de reconhecimento

	PCA (grupo treinado)	PCA (grupo adicionado)	PCA + LDA (grupo total)	PCA + LDA + SOM (grupo total)
Média	88,3%	90,5%	94,3%	95,6%
Limite Inferior	86,6%	88,5%	93,5%	94,2%
Limite Superior	89,7%	92,4%	95,8%	96,9%

Na tabela 5.2 é mostrado a quantidade de identificações incorretas apresentadas pelas 40 classes de pessoas somando os erros encontrados em todos os testes relacionados às variações aleatórias distribuídas nas 10 execuções do teste.

Associando os resultados da tabela 5.2 com uma análise visual das imagens de face pode-se deduzir que as maiores deficiências na identificação foram notadas com faces em escalas diferentes. As faces com divergências na iluminação e rotação não apresentaram taxas de erros significantes.

A pessoa de classe identificada como 1 (figura 5.37) apresenta o pior resultado, mas analisando as imagens dessa classe nota-se uma grande variação na escala e rotação das faces podendo provocar erros dependendo das faces escolhidas para treino e teste.

Nos resultados existem 9 pessoas que apresentaram erro zero, ou seja, foram identificadas corretamente em todos os testes. A figura 5.38 da pessoa de classe identificada como 13 apresenta variações de rotação, feição e uso de óculos, mas possui erro zero. É

Tabela 5.2: Erros na identificação

peessoas/ classes	Identificações Incorretas
1	25
2	6
3	5
4	7
5	6
6	0
7	1
8	1
9	0
10	6
11	8
12	0
13	0
14	0
15	2
16	20
17	7
18	4
19	12
20	6
21	0
22	10
23	0
24	2
25	4
26	5
27	1
28	19
29	6
30	0
31	0
32	11
33	2
34	2
35	23
36	9
37	14
38	6
39	6
40	20



Figura 5.37: Pessoa com a maior quantidade de erros no reconhecimento.

importante ressaltar que o erro na identificação das faces não depende somente das características da classe teste, mas também das classes utilizadas no banco de dados. Duas faces de pessoas diferentes próximas no espaço de faces podem ser um treino e o outro teste gerando erro na identificação, ou as duas testes não gerando erro na identificação. Também é interessante notar que classes de pessoas com características divergentes da média são mais facilmente identificadas.



Figura 5.38: Pessoa com erro zero no reconhecimento.

5.6.6 Atualização Constante dos Neurônios.

Os mapas topológicos são uma parte importante nos sistemas de processamento de informação neural natural (por exemplo, sistema nervoso no córtex visual) e artificial. A rede SOM é um processo de aprendizado não supervisionado que "aprende" ou "memoriza" a distribuição de um conjunto de padrões sem qualquer informação a priori dos dados. Um padrão é projetado do espaço de entrada para uma posição no mapa - a informação é codificada como a localização do nó ativado com o padrão de entrada.

Nas aplicações de reconhecimento de face, onde várias pessoas se apresentam ao sistema intermitentemente, seria ideal que os padrões de faces armazenados recebessem um reforço positivo quando a identificação é correta. A aplicação das redes SOM para armazenar os coeficientes relacionados a cada classe no banco de dados possibilita a mobilidade dos padrões armazenados no banco de acordo com o estímulo de entrada (faces de teste). As redes SOM podem ser definidas para estarem em aprendizado constante de acordo com um limiar definido para o sinal do neurônio vencedor em resposta a uma face de entrada.

O uso da função vizinhança para atualização dos neurônios possibilita um aprendizado ordenado, onde os neurônios topologicamente próximos ao neurônio vencedor na estrutura da rede SOM são atualizados em direção ao padrão de entrada. Ou seja, os neurônios representantes de cada classe no banco de dados se atualizam em direção às faces de teste, quando a identificação é positiva de acordo com um limiar pré-estabelecido. A figura 5.39 exemplifica o efeito de atualização.

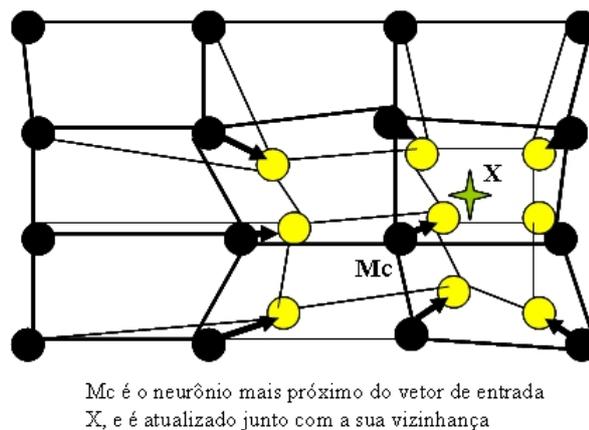


Figura 5.39: Atualização dos neurônios na rede SOM.

Os testes realizados para esse sub-capítulo estão no apêndice B.

5.6.7 Busca por Similaridade.

Muitas aplicações não exigem uma indicação exata da classe correspondente à face de entrada, mas sim, a recuperação de imagens similares à face de entrada em um banco de dados. Uma aplicação típica seria a recuperação de imagens similares no banco de dados da polícia, tendo como entrada uma face virtual semelhante ao indivíduo que se deseja encontrar.

Para analisar a recuperação de imagens foi mantida a estrutura definida para o algoritmo no sub-item 5.6.5, e selecionado aleatoriamente um indivíduo do grupo de teste para efetuar a recuperação da imagem. A face do indivíduo foi modificada com o acréscimo de ruído ou omissão de alguma característica (A figura 5.40 apresenta as faces após a equalização do histograma e modificadas).



Figura 5.40: Faces modificadas.

A recuperação das imagens no banco de dados foi realizada utilizando o método de consulta aos K-Vizinhos mais Próximos (K-Nearest-Neighbor Query - KNNQ). As figuras 5.41 a 5.45 apresentam respectivamente as 10 imagens no banco de dados mais similares às imagens (a), (b), (c), (d) e (e) mostradas na figura 5.40.



Figura 5.41: Faces similares recuperadas para a imagem (a) .



Figura 5.42: Faces similares recuperadas para a imagem (b).



Figura 5.43: Faces similares recuperadas para a imagem (c).



Figura 5.44: Faces similares recuperadas para a imagem (d).



Figura 5.45: Faces similares recuperadas para a imagem (e).

Para as imagens modificadas (a), (b) e (c) o algoritmo conseguiu recuperar com exatidão as 8 faces do respectivo indivíduo utilizadas para o treino. Os resultados para as imagens (d) e (e) não apresentam nenhuma recuperação correta, pois as imagens foram completamente modificadas alterando a distribuição dos tons de cinza da imagem. É importante notar que a face modificada com uma tarja preta na região inferior da face recupera imagens com tons escuros na região inferior das imagens, enquanto que a face modificada com uma tarja preta na região superior da face recupera imagens com tons escuros na região superior das imagens.

5.7 Considerações Finais.

Neste capítulo foram descritos os testes realizados juntamente com discussões sobre os resultados. Uma abordagem geral para a elaboração de um algoritmo utilizado no reconhecimento de faces foi descrita juntamente com a proposta do trabalho, e a apresentação de trabalhos científicos encontrados na literatura que dão suporte ao projeto desenvolvido. O banco de imagens utilizado e o pré-processamento também foram descritos neste capítulo.

Um dos testes avaliou a interação entre a quantidade de autovetores PCA e LDA para a formação dos subespaços com diferentes grupos de treinamento. Os resultados mostraram que existe um ponto ótimo na quantidade de autovetores PCA para maximizar a taxa de reconhecimento apresentada pelo LDA, e esse ponto ótimo é alcançado sem utilizar todos os autovetores PCA. Analisando esses resultados foi definida uma estratégia para estabelecer a quantidade ótima de autovetores PCA na formação do subespaço.

As variáveis intrínsecas aos métodos PCA, LDA e redes neurais SOM foram avaliadas para definir o máximo desempenho do algoritmo. Uma estratégia de atualização dos neurônios nas redes neurais SOM foi avaliada com as variáveis definidas para o máximo desempenho do algoritmo. Isto visando modificar as informações das faces contidas no banco de imagens de treinamento, e considerando as divergências apresentadas pelas imagens de teste com o decorrer de várias apresentações das faces de teste ao algoritmo.

Os últimos testes foram elaborados para avaliar a busca de faces por similaridade no banco de dados , teste importante para as aplicações de recuperação de imagens.

No próximo capítulo são apresentadas as considerações finais do trabalho, as conclusões, contribuições e sugestões para trabalhos futuros.

Conclusões

6.1 Considerações Finais.

Os primeiros sistemas comerciais para o reconhecimento de faces estão começando a ultrapassar a fase de projeto, sendo implementados e testados diversos algoritmos, mas não há até o momento nenhum que se destaque dos outros. Existem muitos algoritmos para o reconhecimento de faces, mas não há um padrão relacionado aos testes e avaliações comparando o desempenho dos algoritmos. Cada algoritmo é restrito a limitações e projetos específicos de aplicação (por exemplo, reconhecimento de faces em documentos, em áreas públicas e ambientes sociais, controle de acesso, etc).

O objetivo de nosso trabalho foi classificar/ identificar uma face com o seu respectivo representante na base de dados associando o indivíduo correspondente (aplicação *one-to-many*). Para isso foi proposto um modelo para o reconhecimento de faces baseando-se nos resultados apresentados por técnicas recentes utilizando o algoritmo do Sub-espço LDA (método holístico) para a codificação das imagens e várias redes neurais SOM em paralelo para a classificação do conjunto de faces.

Os métodos holísticos em reconhecimento de faces tomam como vetor de características os pixels relacionados à face na imagem e para evitar o problema da alta dimensio-

nalidade, vários algoritmos estatísticos de redução de dimensionalidade podem ser utilizados. Neste trabalho, a redução de dimensionalidade foi realizada através do algoritmo Sub-espço LDA utilizando uma rede neural PCA adaptativa nos cálculos dos componentes principais (PCA) necessários.

As etapas PCA e LDA, que formalizam o método Sub-espço LDA, são muito importantes para construir o algoritmo com um alto desempenho no reconhecimento das faces, pois as duas técnicas representam o núcleo do algoritmo fazendo a compressão e representação ótima dos dados (PCA) e ressaltando as diferenças entre as classes para aumentar a performance na identificação (LDA). A escolha dos parâmetros (quantidade de faces de treinamento e quantidade de autovetores para representar os subespaços) adequados é fundamental para incrementar a correlação entre as etapas PCA e LDA na busca de taxas de reconhecimento de faces expressivas com alto desempenho e imunidades às pequenas mudanças entre as faces de teste e de treinamento armazenadas no banco de dados. Os testes foram realizados com vários grupos de treinamento, e os resultados mostraram que não é necessário utilizar todos os autovetores PCA para obter altas taxas de reconhecimento (aproximadamente 98% para o banco de faces ORL).

Nas aplicações de reconhecimento de face, onde várias pessoas se apresentam ao sistema intermitentemente, seria ideal que os padrões de faces armazenados recebessem um reforço positivo quando a identificação fosse correta, e isso foi conseguido com a aplicação das redes SOM. As redes SOM apresentaram resultados satisfatórios na classificação das faces, e possibilitaram a atualização automática do banco de dados, quando as faces de teste apresentavam ligeiras mudanças de iluminação com relação às faces armazenadas no treino.

O algoritmo possui diversas características que lhe conferem robustez e despertam interesses para pesquisa. A velocidade de classificação e busca dentro do conjunto de faces é rápida (menos de 1 segundo)¹. Mas a técnica é lenta no treinamento, que chega a atingir horas dependendo do tamanho da base de faces, além disso, os limiares (*threshold*) devem ser definidos independentemente para cada classe. Outro fator relevante a ser ressaltado é a dependência com relação ao valor de intensidade e posicionamento dos pixels na imagem que formam a face, portanto a rotação, translação, escala, fundo da cena e iluminação ocasionaram os erros encontrados nos testes. Para evitar esses erros, é essencial que a etapa de detecção e pré-processamento seja eficiente para evitar distorções

¹Considerando a plataforma utilizada - software Matlab 6.5 em um computador pessoal Pentium IV - CPU 2.4Ghz com 512MB de RAM e sistema operacional Microsoft Windows 2000.

com relação às características físicas das faces armazenadas, pois essa etapa influencia diretamente nas taxas de reconhecimento e autenticação.

Uma boa aplicação para a técnica apresentada é ordenar a base de faces por similaridade com relação à face de entrada. Sendo essencial controlar as variáveis, que interferem na aquisição das faces para aplicações, onde o necessário é determinar com exatidão se a face apresentada está ou não na base de faces.

6.2 Perspectivas Gerais do Trabalho.

Esse trabalho abordou a área de reconhecimento de faces com relação aos métodos holísticos, e as principais contribuições obtidas foram:

- A proposta de um modelo para o reconhecimento de faces visando contornar os problemas relacionados às pequenas mudanças na rotação, translação, escala, feição, e iluminação das faces com relação ao banco de dados de treinamento.
- Uma avaliação dos parâmetros da interação dos subespaços PCA e LDA na obtenção das taxas de reconhecimento.
- Apresentação de uma metodologia para atualizar o banco de faces de treinamento com relação às faces de teste.

6.3 Linhas de Pesquisas Futuras.

Este trabalho apresentou o resultado de pesquisa envolvendo técnicas holísticas de reconhecimento de faces. A partir de sua idéia e discussão central, vários outros estudos podem ser realizados visando sua complementação e o desenvolvimento de novas técnicas. A seguir são apresentadas algumas sugestões de tópicos que podem ser abordados em estudos futuros:

- Desenvolver um trabalho cooperativo com as delegacias para aplicar o algoritmo em um banco de dados padronizado da polícia visando fazer buscas por similaridade.
- Integrar o algoritmo de reconhecimento com algum método de detecção de faces para implementar um sistema de reconhecimento em ambientes sociais ou para o controle de acesso de pessoas.

- Integrar o algoritmo proposto com outros métodos biométricos para implementar sistemas de autenticação e reconhecimento mais confiáveis.
- Formalizar um modelamento matemático para definir os parâmetros ótimos de interação entre o PCA e LDA.
- Aplicar o PCA 2-D na redução do espaço de faces.
- Utilizar outros métodos de classificação como: LVQ e lógica Fuzzy.
- Associar o algoritmo proposto com alguma técnica de reconhecimento de faces, que visa avaliar as características locais da face para incrementar e tornar o algoritmo mais robusto no reconhecimento.
- Testar o algoritmo proposto em outros bancos de faces.
- Extração do conteúdo das imagens para indexação em bases muito grandes.

Enfim, um algoritmo de reconhecimento de faces deve apresentar bons resultados, quando a face de teste diverge por algum motivo das faces presentes no banco de dados, ou quando o banco de dados é grande. Portanto um estudo criterioso do espalhamento dos vetores de faces no espaço altamente dimensional pode ajudar a elaborar algoritmos mais eficientes.

Referências Bibliográficas

BARLOW, H. B. (1989). Unsupervised learning. *Neural Computation*, 1:295–311.

BARRETT, W. A. (1998). A survey of face recognition algorithms and testing results. *Systems and Computers*, (1):301–305.

BARTLETT, M. S., LADES, H. M., and SEJNOWSKI, T. (1998). Independent component representation for face recognition. *In Proceedings, SPIE Symposium on Electronic Imaging: Science and Technology*, pages 528–539.

BELHUMEUR, P. N., HESPANHA, J. P., and KRIEGMAN, D. J. (1997). Eigenfaces vs. fisherfaces: recognition using class specific linear projection. *IEEE Transactions On Pattern Analysis And Machine Intelligence*, 19(7).

BIEDERMAN, I. (1987). Recognition by components: A theory of human image understanding. *Psych. Rev.*, 94:115–147.

BOURLARD, H. and KAMP, Y. (1988). Auto-association by multilayer perceptrons and singular value decomposition. *Biological Cybern.*, 59:291–294.

BRUCE, V. (1988). Recognizing faces. *Lawrence Erlbaum Associates, London, U.K.*

BRUCE, V. (1999). Identification of human faces. *IEEE - Image Processing and Its Applications, Conference Publication*, (465):615–619.

BRUCE, V., BURTON, M., and DENCH, N. (1994). What's distinctive about a distinctive face? *Quart. J. Exp. Psych.*, (47A):119–141.

- BRUCE, V., HANCOCK, P. J. B., and BURTON, A. M. (1998). *Human face perception and identification*. In *Face Recognition: from theory to applications*. Springer-Verlag, Berlin, Germany.
- BRUNELLI, R. and POGGIO, T. (1992). Face recognition through geometrical features. *Proceedings of the Second European Conference on Computer Vision*, pages 792–800.
- BRUNELLI, R. and POGGIO, T. (1993). Face recognition: features versus templates. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 15(10):1042–1052.
- CAMPOS, T. E. (2001). Técnicas de seleção de características com aplicações em reconhecimento de faces. Master's thesis, IME - Instituto De Matemática E Estatística - USP, São Paulo.
- CAMPOS, T. E., FERIS, R. S., and JUNIOR, R. M. C. (2000). Eigenfaces versus eigeneyes: first steps toward performance assessment of representations for face recognition. *Lecture Notes in Artificial Intelligence*, 1793:197–206.
- CELLI, P. L. F. (1999). Um sistema de apoio à identificação de suspeitos com reconhecimento automático de faces. Master's thesis, ICMC - Instituto De Ciências Matemáticas E De Computação - USP, São Carlos.
- CHANGEUX, J. P. and DANCHIN, A. (1976). Selective stabilization of developing synapses as a mechanism for the specification of neural networks. *Nature*, 264:705–712.
- CHAOYANG, L., FANG, L., and YINXIANG, X. (2003). Face recognition using self-organizing feature maps and support vector machines. *Proceedings of the Fifth International Conference on Computational Intelligence and Multimedia Applications*.
- CHATTERJEE, C. and ROYCHOWDHURY, V. P. (1997). Artificial neural networks for feature extraction and data projection. *IEEE Transactions On Neural Networks*, 8(3).
- CHELLAPPA, R. and ETEMAD, K. (1996). Face recognition using discriminant eigenvectors. in *Proceedings of International Conference on Acoustics, Speech and signal Processing*, pages 2148–2151.
- CHELLAPPA, R., WILSON, C. L., and SIROHEY, S. (1995). Human and machine recognition of faces: a survey. *Proceedings of the IEEE*, 83:705–740.
- CHEN, H. and LIU, R.-W. (1992). Adaptive distributed orthogonalization processing for principal components analysis. *Im. Conf. Acoust. Speech Signal Processing*, 2.

- CHURCHLAND, P. S. and SEJNOWSKI, T. J. (1992). *The computational brain*. Mit Press, Cambridge, MA.
- COTTRELL, G. W. and FLEMING, M. (1990). Face recognition using unsupervised feature extraction. *In Proc. Int. Neural Network Conf.*, 1:322–325.
- DIAMANTARAS, K. I. and KUNG, S. Y. (1991). An unsupervised neural model for oriented principal component extraction. *in Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing (ICASSP-91)*, pages 1049–1052.
- DIAMANTARAS, K. I. and KUNG, S. Y. (1996). *Principal component neural networks: theory and applications*. Wiley, New York.
- DUDA, R. O. and HART, P. E. (1973). *Pattern classification and scene analysis*. Wiley-Interscience, USA.
- DUNTEMAN, G. H. (1989). *Principal components analysis*. Sage Publications, Newbury Park.
- ELLIS, H. D. (1986). *Introduction to aspects of face processing: Ten questions in need of answers*. Aspects of Face Processing.
- ETEMAD, K. and CHELLAPPA, R. (1997). Discriminant analysis for recognition of human face images. *J. Opt. Soc. Am.*, 14:1724–1733.
- FISHER, R. (1938). The statistical utilization of multiple measurements. *Annals of Eugenics*, 8:376–386.
- FOLDIAK, P. (1989). Adaptive network for optimal linear feature extractions. *IEEE International Joint Conference on Neural Networks*, 1:401–4054.
- GINSBURG, A. G. (1978). Visual information processing based on spatial filters constrained by biological data. *AMRL Tech. rep.*, pages 78–129.
- GONG, S., MCKENNA, S., and PSARROU, A. (2000). *Dynamic vision: from images to face recognition*. Imperial College Press, UK.
- GONZALEZ, R. C. and WOODS, R. E. (1992). *Digital Image Processing*. Addison-Wesley Publishing Company, USA.
- GROB, M. (1994). *Visual computing : the integration of computer graphics, visual perception and imaging*. Springer-Verlag, Berlin.

- HANCOCK, P., BRUCE, V., and BURTON, M. (1998). A comparison of two computer-based face recognition systems with human perceptions of faces. *Vis. Res.*, 38:2277–2288.
- HARMON, L. D. (1973). The recognition of faces. *Sci. Am.*, 229:71–82.
- HAYKIN, S. (1998). *Neural networks: a comprehensive foundation*. Prentice Hall, New York.
- HEBB, D. O. (1949). *The organization of behavior: a neuropsychological theory*. Wiley, New York.
- JAIN, A. K., DUIN, R. P. W., and MAO, J. (2000). Statistical pattern recognition: A review. *IEEE Transactions On Pattern Analysis and Machine Intelligence*.
- JAMES, M. L., SMITH, G. M., and WOLFORD, J. C. (1993). *Applied numerical methods for digital computation*. Harper Collings College Publishers.
- JOHNSTON, A., HILL, H., and CARMAN, N. (1992). Recognizing faces: Effects of lighting direction, inversion and brightness reversal. *Cognition* 40, pages 1–19.
- JOLLIFFE, I. T. (1986). *Principal component analysis*. Springer-Verlag, New York.
- KALOCSAI, P. K., ZHAO, W., and ELAGIN, E. (1998). Face similarity space as perceived by humans and artificial systems. *In Proceedings, International Conference on Automatic Face and Gesture Recognition*, pages 177–180.
- KANADE, T. (1973). *Picture processing system by computer complex and recognition of human faces*. PhD thesis, Dept. of Information Science, Kyoto University, Japan.
- KATOH, A. and FUKUI, Y. (1998). Classification of facial expressions using self-organizing maps. *IEEE Conference of Engineering in Medicine and Biology Society*.
- KIRBY, M. and SIROVICH, L. (1990). Application of the karhunen-loeve procedure for the characterization of human faces. *IEEE Trans. Patt. Anal. Mach. Intell.*, 12.
- KOHONEN, T. (1988a). An introduction to neural computing. *Neural Networks*, 1:3–16.
- KOHONEN, T. (1988b). *Self-organization and associative memory*. Springer-Verlag, New York.

- KOHONEN, T. (1988c). The self-organizing map. *Vision*.
- LAWRENCE, S., GILES, C. L., TSOI, A. C., and BACK, A. D. (1996). Face recognition: A hybrid neural network approach. *Technical report, I. A. C. S., U. of Maryland*.
- LAWRENCE, S., GILES, C. L., TSOI, A. C., and BACK, A. D. (1997). Face recognition: a convolutional neural-network approach. *IEEE Trans. Neural Netw.*, 8:98–113.
- LEE, S. J., JUNG, S. B., KWON, J. W., and HONG, S. H. (1999). Face detection and recognition using pca. *Proceedings of the IEEE Region 10 Conference*, 1:84–87.
- LEVINE, M. (1985). *Man and machine vision*. McGraw-Hill, New York.
- LIN, S. H. (2000). An introduction to face recognition technology. *Informing Science Special Issue on Multimedia Informing Technologies*, 3(2).
- LIPOSCAK, Z. and LONCARIC, S. (1999). Face recognition from profiles using morphological operations. *RATFG99*, pages 47–52.
- LIPSCHUTZ, L. (1994). *Álgebra linear : teoria e problemas*. Makron Books, São Paulo.
- LU, J., PLATANIOTIS, K. N., and VENETSANOPOULOS, A. N. (2003). Face recognition using lda-based algorithms. *IEEE Transactions On Neural Networks*, 14(1).
- LUO, M. and YUWAPAT (2003). Eigenfaces for face recognition - final project report. Technical report, Department of Electrical and Computer Engineering - CS.
- MALSBURG, V. D. (1990a). *Network self-organization, in an introduction to neural and electronic networks*. Academic Press, San Diego.
- MALSBURG, V. D. (1990b). Considerations for a visual architecture. *Advanced Neural Computers*.
- MAO, J. and JAIN, A. K. (1993). Discriminant analysis neural networks. *Proc. IEEE International Conference on Neural Networks*, 1:300–305.
- MAO, J. and JAIN, A. K. (1995). Artificial neural networks for feature extraction and data projection. *IEEE Transactions on Neural Networks*, 6:296–317.
- MARR, D. (1982). *Vision*. W. H. Freeman and Company, New York.

- MARTÍNEZ, A. M. and C.KAK, A. (2001). Pca versus lda. *IEEE Transactions On Pattern Analysis And Machine Intelligence*, 23(2):228–233.
- MENDEL, J. M. and MCLAREN, R. W. (1970). *Reinforcement-learning control and pattern recognition systems, in adaptative learning and pattern recognition systems: theory and applications*. Academic Press, New York.
- MOGHADDAM, H. A. and ZADEH, K. A. (2002). Fast linear discriminant analysis for on-line pattern recognition applications. *IEEE*.
- NEAGOE, V. E. and ROPOT, A. D. (2002). Concurrent self-organizing maps for pattern classification. *IEEE International Conference on Cognitive Informatics*.
- OJA, E. (1982). A simplified neuron model as a principal component analyzer. *Journal of Mathematical Biology*, 15.
- OJA, E. (1989). Neural networks, principal components, and subspaces. *International Journal of Neural Systems*, 1.
- OJA, E. (1992a). Principal components, minor components, and linear neural networks. *Neural Networks*, 5(6):927–935.
- OJA, E. (1992b). Self-organizing maps and computer vision. *Neural Networks for Perception*, 1.
- OLIVEIRA, P. R. (1997). Redes neurais artificiais para extração de características. Master's thesis, ICMC - Instituto De Ciências Matemáticas E De Computação - USP, São Carlos.
- OLIVETTI and ORACLE. Database of faces - orl (www.camorl.co.uk). Olivetti and Oracle Research Laboratory.
- O'TOOLE, A., ABDI, H., KENNETH, DEFFENBACHER, A., and VALENTIN, D. (1993). Low-dimensional representation of faces in higher dimensions of the face space. *J. Opt. Soc. Am. A*.
- PENTLAND, A. and CHOUDHURY, T. (2000). Face recognition for smart environments. *IEEE Computer Society Press*, 33(2):50–55.
- PENTLAND, A., MOGHADDAM, B., and STARNER, T. (1994). View-based and modular eigenspaces for face recognition. *In Proceedings, IEEE Conference on Computer Vision and Pattern Recognition*.

- PERKINS, D. (1975). A definition of caricature and recognition. *Stud. Anthro. Vis. Commun.*, 2:1–24.
- PODILCHUK, C. and ZHANG, X. (1996). Face recognition using dct-based feature vectors. *ICASSP96*, pages 2144–2147.
- PRINCIPE, J. C. and XU, D. (1997). Classification with linear networks using an on-line constrained lda algorithm. *ICPR*, 2:64–67.
- RISK, M. R. M. and TAHA, A. (2002). Analysis of neural networks for face recognition systems with feature extraction to develop an eye localization based method. *IEEE*.
- ROMDHANI, S. (1996). Face recognition using principal component analysis. Master's thesis, Department of Electronics and Electrical Engineering, University of Glasgow, UK.
- RUBNER, J. and TAVAN, P. (1989). A self-organizing network for principal components analysis. *Europhysics Letters*, 10.
- RUMELHART, D. E. and ZIPSER, D. (1985). Features discovery by competitive learning. *Cognitive Science*, 9:75–112.
- SANGER, T. D. (1989a). An optimality principle for unsupervised learning. *Advances in Neural Information Processing Systems*, 1.
- SANGER, T. D. (1989b). Optimality unsupervised learning in a single-layer linear feedforward neural network. *Neural Networks*, 12.
- SANTOS, A. R. and GONZAGA, A. (2005a). Automatic clusters to face recognition. *SIBGRAPI 2005*.
- SANTOS, A. R. and GONZAGA, A. (2005b). Face recognition based on lda and som neural nets. *WVC'2005*.
- SATO, Y., YODA, I., and SAKAUE, K. (2003). Automatic face classifications by self-organization for face recognition. *Proceedings of the IEEE International Workshop on Analysis and Modeling of Faces and Gestures*, pages 165–172.
- SERGEANT, J. (1986). *Microgenesis of face perception. In aspects of face processing*. Nijhoff, Dordrecht, The Netherlands.

- SHEPHERD, J. W., DAVIES, G. M., and ELLIS, H. D. (1981). *Studies of cue saliency. In perceiving and remembering faces*. Academic Press, London, U.K.
- SIROVICH, L. and KIRBY, M. (1987). Low-dimensional procedure for the characterization of human face. *J. Opt. Soc. Am.*, 4:519–524.
- SPENCER, R. G. and SINENCIO, E. S. (1999). A fully-differential cmos implementation of oja's learning rule in dual-synapse neuron for extracting principal components for face recognition. *IEEE*.
- SPIEGEL, M. R. (1994). *Estatística*. Makron, São Paulo.
- SPIES, H. and RICKETTS, I. (2000). Face recognition in fourier space. *Vision Interface 2000: Montreal*, pages 38–44.
- STENT, G. S. (1973). A physiological mechanism for hebb's postulate of learning. *Proceedings of the National Academy of Sciences, USA*, 70.
- SULAIMAN, N., YEE, Y. M., WOEI, G. C., and SHARMINI (2000). Pattern recognition -face recognition technology. Artificial Intelligence (SAK4602) - Lecture.
- SWETS, D. L. and WENG, J. (1996). Discriminant analysis and eigenspace partition tree for face and object recognition from views. *In Proceedings, International Conference on Automatic Face and Gesture Recognition*, pages 192–197.
- SWETS, D. L. and WENG, J. (1996b). Using discriminant eigenfeatures for image retrieval. *IEEE Trans. Patt. Anal. Mach. Intell.*, 18:831–836.
- TAKAHASHI, T. and KURITA, T. (2001). A self-organization model of feature columns and face responsive neurons in the temporal cortex. *IEEE Neural Networks*, 1:82–87.
- TARR, M. J. and BULTHOFF, H. H. (1993). Is human object recognition better described by geon structural descriptions or by multiple views- comment on biederman and gerhardstein. *J. Exp. Psych.: Hum. Percep. Perf.*, 21:71–86.
- TISTARELLI, M. and GROSSO, E. (1999). Active vision-based face authentication. *Image and Vision Computing*, 18:299–314.
- TURK, M. and PENTLAND, A. (1991a). Eigenfaces for recognition. *J. Cogn. Neurosci.*, 3:72–86.

- TURK, M. and PENTLAND, A. (1991b). Face recognition using eigenfaces. *In Proceedings of IEEE Computer Vision and Pattern Recognition*, pages 586–590.
- WANG, X. and TANG, X. (2004a). Dual-space linear discriminant analysis for face recognition. *in Proceedings of CVPR*.
- WANG, X. and TANG, X. (2004b). Random sampling lda for face recognition. *IEEE Computer Society Conference on Publication - Computer Vision and Pattern Recognition (CVPR 2004)*, 2(2):259–265.
- XU, L. and YUILLE, A. L. (1995). Robust principal component analysis by self-organizing rules based on statistical physics approach. *IEEE Transactions on Neural Networks*, 6.
- YAMBOR, W. S., DRAPER, B. A., and BEVERIDGE, J. (2000). Analyzing pca-based face recognition algorithms: eigenvector selection and distance measures. Computer Science Department - Colorado State University.
- YANG, J., YU, H., and KUNZ, W. (2000). An efficient lda algorithm for face recognition. *International Conference on Automation, Robotics, and Computer Vision, ICARCV'2000*.
- ZHAO, W., CHELLAPPA, R., and KRISHNASWAMY, A. (1998a). Discriminant analysis of principal components for face recognition. *International Conference on Automatic Face and Gesture Recognition*, pages 336–341.
- ZHAO, W., CHELLAPPA, R., and KRISHNASWAMY, A. (1998b). Discriminant analysis of principal components for face recognition. *In Proceedings, International Conference on Automatic Face and Gesture Recognition*, pages 336–341.
- ZHAO, W., CHELLAPPA, R., and NANDHAKUMAR, N. (1998c). Empirical performance analysis of linear discriminant classifiers. *Proceedings of CVPR*, pages 164–169.
- ZHAO, W., CHELLAPPA, R., PHILLIPS, P. J., and ROSENFELD, A. (2003). Face recognition: a literature survey. *ACM Computing Surveys*, 35(4):399–458.
- ZHAO, W., CHELLAPPA, R., and P.J.PHILLIPS (1996). Subspace linear discriminant analysis for face recognition. *Technical Report CAR-TR-914*.

Apêndice: Algoritmos

ALGORITMO PCA ADAPTATIVO

1. Inicialize o vetor peso para frente W_j e vetor peso realimentado A_j com valores aleatórios pequenos no tempo $n=1$, onde $j=1,2,\dots,L$ e $\eta_j(n)$ e $\mu_j(n)$ devem ser um valor positivo pequeno.
2. Faça $j=1$ e para $n=1,2,3,\dots$, calcule

$$y_1(n) = W_1^T(n) \cdot X(n)$$

$$W_1(n+1) = W_1(n) + \eta(n) \cdot y_1(n) \cdot X(n) + \beta(n) \cdot \Delta W_1(n)$$
 onde $X(n)$ é o vetor de entrada. Para n grande, tem-se $W_1(n) \Rightarrow q_1$, onde q_1 é o autovetor associado com o maior autovalor λ_1 da matriz de correlação de $X(n)$.
3. Faça $j=2$ e para $n=1,2,3,\dots$, calcule

$$YR(n) = [y_1(n), y_2(n), \dots, y_{(j-1)}(n)]^T$$

$$y_j(n) = W_j^T(n) \cdot X(n) + A_j^T(n) \cdot YR(n)$$

$$W_j(n+1) = w_1(n) + \eta(n) \cdot y_j(n) \cdot X(n) + \beta(n) \cdot \Delta W_j(n)$$

$$A_j(n+1) = A_j(n) - \mu(n) \cdot y_j(n) \cdot YR(n) + \beta(n) \cdot \Delta A_j(n)$$
4. Incremente j em 1, vá para o passo 3 e continue até $j=L$, onde L é o número desejado de componentes principais. (note que $j=1$ corresponde ao autovetor as-

sociado com o maior autovalor, o qual é tratado no passo 2) Para n grande, tem-se $W^j(n) \Rightarrow q^j$, onde q^j é o autovetor associado com o maior autovalor λ_j da matriz de correlação de $X(n)$.

5. Após alcançar o valor L , vá para o passo 2 e execute o algoritmo novamente até que todos os pesos laterais sejam suficientemente pequenos (a soma de seus máximos absolutos seja menor que algum limiar ε), ou seja atingido um número máximo de iterações.

ALGORITMO DE APRENDIZAGEM SUPERVISIONADA AUTO-ORGANIZÁVEL LDA

Seja $T = \{\zeta_k^{(l)} \mid k=1, 2, \dots, nl, l=1, 2, \dots, c\}$ denota o conjunto dos padrões de entrada para o treinamento. E $T_{T_o} = \{\xi_k^{(l)}, k=1, 2, \dots, nl, l=1, 2, \dots, c\}$ denota o conjunto dos dados de treinamento normalizado obtidos pela subtração de cada padrão em T pelo vetor médio global do conjunto dos dados de treinamento T , $\xi_k^{(l)} = \zeta_k^{(l)} - m$. Forma-se também, outro conjunto de treinamento com média zero, $T_{w_o} = \{\eta_k^{(l)}, k=1, 2, \dots, nl, l=1, 2, \dots, c\}$ pela subtração de cada padrão em T pelo vetor médio correspondente a classe ($\eta_k^{(l)} = \zeta_k^{(l)} - m^{(l)}$). A informação com relação as classe somente é usada para formar o conjunto de classe para o treinamento com média zero (T_{w_o}).

Uma vez feito isto, o algoritmo de treinamento é completamente auto-organizado.

O algoritmo resume-se no seguinte:

1. Formar o conjunto de treinamento T_{w_o} e T_{T_o} a partir de T .
2. Treinar a primeira camada usando o algoritmo de aprendizado PCA tendo como entrada T_{w_o} .
3. Projetar todos os padrões de T_{w_o} usando a primeira camada de treinamento ; calcular o desvio padrão para cada neurônio de saída; escalonar todos os pesos nas conexões para cada neurônio de saída pelo desvio padrão associado a este neurônio.

$$W_{ij}^{(1)} = \frac{W_{ij}^{(1)}}{\sigma_j}, \text{ com } i, j = 1, 2, \dots, d$$
4. Formar um conjunto de dados T_H pela coleção de saídas da camada intermediária quando o conjunto de treinamento T_{T_o} é apresentado a camada de entrada.
5. Treinar a segunda camada usando o algoritmo de aprendizado PCA com os dados do conjunto T_H .

$\Phi = [W_{ij}^{(1)}]_{dxd}$ e Φ_s são as matrizes de pesos quando os passos 2 e 3 estão terminados, respectivamente. $\Psi = [W_{ij}^{(2)}]_{dxd}$ são os pesos da matriz da segunda camada depois do treinamento. O mapeamento geral que a rede executa é $o_k^{(l)} = A^T \cdot \xi_K^{(l)}$, onde $A = \Phi_s \cdot \Psi$, e $o_k^{(l)}$ é o vetor saída da rede quando um vetor entrada $\xi_K^{(l)}$ é apresentado. Concluindo, as colunas da matriz A são compostas pelos autovetores de $(\sum W)^{-1} \cdot \sum B$ correspondentes aos autovalores em ordem decrescente. Quando o processo de aprendizado é terminado, estes autovalores podem ser obtidos calculando as variâncias das unidades de saída quando o conjunto de treinamento T_{TO} é apresentado para os neurônios de entrada.

Apêndice: Testes - Atualização Constante dos Neurônios

Para avaliar a atualização dos neurônios em direção às faces de teste foram selecionadas duas pessoas - A e B com 8 faces de treino e 12 testes (figura B.1 e figura B.2) mantendo as características estruturais do algoritmo utilizado nos testes presentes no sub-item 5.6.5.



Figura B.1: Pessoa A.

Os neurônios e faces de teste de cada indivíduo são vetores unitários espalhados em uma superfície n-dimensional com raio 1 (módulo dos vetores igual a 1). A quantidade de épocas foi definida como 96 (ou seja, as 12 faces de teste se apresentam 8 vezes na



Figura B.2: Pessoa B.

entrada do modelo), as taxas de aprendizado inicial e final como sendo 0.8 e 0.09, com vizinhança gaussiana inicial igual a 5.

As distâncias entre os neurônios do indivíduo A e suas faces de teste são mostradas na figura B.3 retirada da janela de comando do Matlab, e na figura B.4 são apresentados os valores após à atualização dos neurônios em direção às amostras de entrada.

O importante nesses resultados é notar que as distâncias entre os neurônios vencedores e as faces de teste diminuem comprovando o aprendizado. E os neurônios vencedores antes e depois das atualizações são diferentes demonstrando a atualização ordenada do mapa topológico em direção aos padrões de entrada, visando minimizar as distâncias entre os neurônios e as entradas.

```
>> c_old

c_old =

Columns 1 through 7

    0.8315    0.4265    0.2971    0.3090    0.7941    0.7272    ●0.2056
    0.7692    0.3843    0.3460    0.3681    0.7093    0.6811    0.2962
    ●0.3555    0.5321    0.5888    0.6642    ●0.4547    0.3568    0.6000
    0.7541    ●0.1914    0.3596    0.4161    0.7329    0.6073    0.4068
    0.8016    0.3229    0.3490    0.3045    0.7595    0.6821    0.3693
    0.8044    0.3209    ●0.2168    ●0.2388    0.7542    0.7064    0.2158
    0.4651    0.5309    0.6199    0.7648    0.5343    ●0.3350    0.6226
    0.6791    0.6650    0.7078    0.7335    0.4912    0.7126    0.7175

Columns 8 through 12

    0.2660    0.6758    0.9059    0.6710    0.8128
    0.3410    0.5839    0.8816    0.6175    0.7493
    0.6807    0.4835    0.4860    ●0.2587    0.4291
    0.4396    0.7062    0.7728    0.6854    0.6642
    0.3495    0.6406    0.9123    0.6667    0.7493
    ●0.2484    0.6442    0.8792    0.6508    0.7841
    0.7541    0.6435    ●0.3405    0.4565    ●0.4117
    0.7534    ●0.3868    0.8682    0.5649    0.7526

● Neurônio vencedor para cada teste
```

Figura B.3: Distância entre os 8 neurônios (linha) e as 12 faces de teste (colunas) antes do treinamento.

```
>> c_new

c_new =

Columns 1 through 7

    0.8571    0.4206    0.2145    ●0.0743    0.8228    0.7766    0.2424
    0.7917    0.3398    ●0.0733    0.2408    0.7539    0.6944    ●0.0778
    0.6918    ●0.0215    0.2965    0.3920    0.6669    0.5809    0.3578
    0.4441    0.3875    0.4619    0.5495    0.3649    0.4412    0.4794
    0.3896    0.5975    0.6356    0.7078    0.2788    0.4778    0.6371
    0.2557    0.6589    0.7426    0.8092    ●0.0166    0.4572    0.7550
    ●0.0232    0.6791    0.7807    0.8485    0.2613    0.3499    0.7921
    0.3278    0.6328    0.7502    0.8473    0.4611    ●0.1443    0.7588

Columns 8 through 12

    ●0.0759    0.7241    0.9400    0.7279    0.8354
    0.2159    0.6601    0.8386    0.6350    0.7675
    0.4280    0.6313    0.7397    0.5903    0.6464
    0.5758    0.2813    0.6168    0.2501    0.5161
    0.7275    ●0.1577    0.6220    ●0.1541    0.5391
    0.8282    0.3407    0.5172    0.2683    0.4914
    0.8599    0.4971    0.3517    0.3168    0.3686
    0.8364    0.6159    ●0.2557    0.4544    ●0.1405

● Neurônio vencedor para cada teste
```

Figura B.4: Distância entre os 8 neurônios (linha) e as 12 faces de teste (colunas) após o treinamento.

As faces de teste para o indivíduo B tiveram a intensidade luminosa modificada (veja figura B.5) para demonstrar a atualização dos neurônios em direção às faces de teste. As figuras B.6 e B.7 mostram os resultados com relação as distâncias entre os vetores.



Figura B.5: Pessoa B com as imagens mais claras.

```
>> c_old
c_old =
Columns 1 through 7
    0.3156    0.3074    0.2969    0.4460    0.5699    0.2096    0.5269
    0.3324    0.3504    0.3941    0.3925    0.4884    0.3698    0.4795
    0.3827    0.1355    0.3689    0.3670    0.5619    0.2475    0.5304
    0.3350    0.2734    0.3607    0.4203    0.5455    0.3377    0.5073
    0.2864    0.2362    0.2871    0.3460    0.4701    0.2803    0.4669
    0.3141    0.2699    0.2638    0.3982    0.5250    0.2753    0.4812
    0.2455    0.3934    0.2753    0.2967    0.3659    0.3166    0.2300
    0.5028    0.4003    0.4991    0.3948    0.6515    0.3809    0.5645

Columns 8 through 12
    0.6082    0.4842    0.3808    0.3347    0.1778
    0.4951    0.4980    0.3653    0.4508    0.3844
    0.6122    0.4055    0.2591    0.3001    0.3040
    0.5618    0.5032    0.3432    0.3992    0.2943
    0.5073    0.4564    0.3049    0.3250    0.2642
    0.5620    0.4929    0.3579    0.3146    0.2439
    0.4467    0.4235    0.2879    0.3903    0.4496
    0.6816    0.2591    0.2717    0.4674    0.4985

● Neurônio vencedor para cada teste
```

Figura B.6: Distância entre os 8 neurônios (linha) e as 12 faces de teste (colunas) antes do treinamento.

A atualização dos neurônios representantes das classes de indivíduos no banco de dados pode ser útil em aplicações, onde as faces de teste divergem dos padrões definidos para as faces de treinamento com o passar do tempo. Por exemplo: os indivíduos envelhecem, a iluminação apresenta um decremento com o tempo, o indivíduo muda ligeiramente o cabelo, ou seja, o algoritmo deve estar em constante aprendizado para extrair e armazenar

as novas características apresentadas pelos indivíduos nos teste. A figura B.8 simplifica a modificação da região de identificação da classe de um indivíduo de acordo com o seu vetor (ou neurônio) representante no banco de dados.

```
>> c_new
c_new =
Columns 1 through 7
    0.3525    0.2605    0.3012    0.4084    0.5808    0.1353    0.4772
    0.4098    0.0238    0.3936    0.3742    0.5815    0.2857    0.5203
    0.3381    0.2700    0.3715    0.1905    0.4428    0.3124    0.3990
    0.3985    0.4064    0.4359    0.1295    0.4236    0.3637    0.4095
    0.1972    0.3571    0.2348    0.1964    0.3469    0.2645    0.2999
    0.0647    0.4136    0.0719    0.3791    0.3782    0.3001    0.3109
    0.2735    0.5243    0.3298    0.3394    0.2897    0.4545    0.0248
    0.3628    0.6066    0.4479    0.3491    0.0921    0.5752    0.3165

Columns 8 through 12
    0.6183    0.4426    0.3432    0.1503    0.1571
    0.6249    0.4467    0.2625    0.2443    0.3064
    0.4897    0.2555    0.0105    0.3658    0.4209
    0.4489    0.1226    0.1951    0.4337    0.4933
    0.3968    0.2809    0.1942    0.3738    0.3689
    0.4393    0.4709    0.3478    0.4154    0.3307
    0.3344    0.4799    0.3885    0.4855    0.5251
    0.0917    0.5171    0.4590    0.6207    0.6282

● Neurônio vencedor para cada teste
```

Figura B.7: Distância entre os 8 neurônios (linha) e as 12 faces de teste (colunas) após o treinamento.

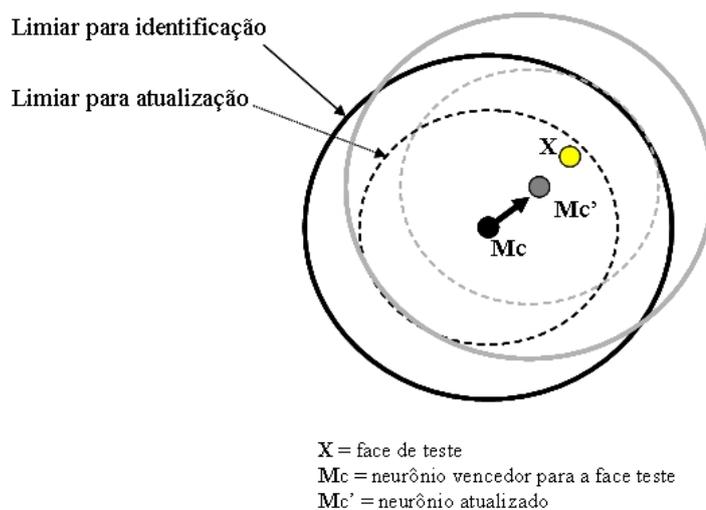


Figura B.8: Modificação da região de identificação para o neurônio vencedor com a face de teste X.

Livros Grátis

(<http://www.livrosgratis.com.br>)

Milhares de Livros para Download:

[Baixar livros de Administração](#)

[Baixar livros de Agronomia](#)

[Baixar livros de Arquitetura](#)

[Baixar livros de Artes](#)

[Baixar livros de Astronomia](#)

[Baixar livros de Biologia Geral](#)

[Baixar livros de Ciência da Computação](#)

[Baixar livros de Ciência da Informação](#)

[Baixar livros de Ciência Política](#)

[Baixar livros de Ciências da Saúde](#)

[Baixar livros de Comunicação](#)

[Baixar livros do Conselho Nacional de Educação - CNE](#)

[Baixar livros de Defesa civil](#)

[Baixar livros de Direito](#)

[Baixar livros de Direitos humanos](#)

[Baixar livros de Economia](#)

[Baixar livros de Economia Doméstica](#)

[Baixar livros de Educação](#)

[Baixar livros de Educação - Trânsito](#)

[Baixar livros de Educação Física](#)

[Baixar livros de Engenharia Aeroespacial](#)

[Baixar livros de Farmácia](#)

[Baixar livros de Filosofia](#)

[Baixar livros de Física](#)

[Baixar livros de Geociências](#)

[Baixar livros de Geografia](#)

[Baixar livros de História](#)

[Baixar livros de Línguas](#)

[Baixar livros de Literatura](#)
[Baixar livros de Literatura de Cordel](#)
[Baixar livros de Literatura Infantil](#)
[Baixar livros de Matemática](#)
[Baixar livros de Medicina](#)
[Baixar livros de Medicina Veterinária](#)
[Baixar livros de Meio Ambiente](#)
[Baixar livros de Meteorologia](#)
[Baixar Monografias e TCC](#)
[Baixar livros Multidisciplinar](#)
[Baixar livros de Música](#)
[Baixar livros de Psicologia](#)
[Baixar livros de Química](#)
[Baixar livros de Saúde Coletiva](#)
[Baixar livros de Serviço Social](#)
[Baixar livros de Sociologia](#)
[Baixar livros de Teologia](#)
[Baixar livros de Trabalho](#)
[Baixar livros de Turismo](#)