UNIVERSIDADE FEDERAL DE UBERLÂNDIA

FACULDADE DE ENGENHARIA ELÉTRICA

PÓS-GRADUAÇÃO EM ENGENHARIA ELÉTRICA

RECONHECIMENTO DE FACE UTILIZANDO TRANSFORMADA DISCRETA DO COSSENO BIDIMENSIONAL, ANÁLISE DE COMPONENTES PRINCIPAIS BIDIMENSIONAL E MAPAS AUTO-ORGANIZÁVEIS CONCORRENTES

Thayso Silva Guimarães

Maio

2010

UNIVERSIDADE FEDERAL DE UBERLÂNDIA

FACULDADE DE ENGENHARIA ELÉTRICA

PÓS-GRADUAÇÃO EM ENGENHARIA ELÉTRICA

RECONHECIMENTO DE FACE UTILIZANDO TRANSFORMADA DISCRETA DO COSSENO BIDIMENSIONAL, ANÁLISE DE COMPONENTES PRINCIPAIS BIDIMENSIONAL E MAPAS AUTO-ORGANIZÁVEIS CONCORRENTES

Thayso Silva Guimarães

Dissertação apresentada à Universidade Federal de Uberlândia para obtenção do título de Mestre em Engenharia Elétrica.

Banca examinadora:

Edna Lúcia Flôres, Dra. – Orientadora (UFU)

Cláudio Afonso Fleuri, Dr. (PUC-GO/IFG)

Antônio Claúdio Paschoarelli Veiga, Dr. (UFU)

Uberlândia, Maio de 2010.

Dados Internacionais de Catalogação na Publicação (CIP)
Sistema de Bibliotecas da UFU – MG, Brasil

G963r Guimarães, Thayso Silva, 1983-
Reconhecimento de face utilizando transformada discreta do cosseno bidimensional, análise de componentes principais bidimensional e mapas auto-organizáveis concorrentes [manuscrito] / Thayso Silva Guimarães. - 2010.
136 f. : il.

Orientadora: Edna Lúcia Flôres.

Dissertação (mestrado) – Universidade Federal de Uberlândia, Programa de Pós-Graduação em Engenharia Elétrica.
Inclui bibliografia.

1. Reconhecimento de padrões - Teses. I. Flores, Edna Lúcia, \d 1958-. II. Universidade Federal de Uberlândia. Programa de Pós-Graduação em Engenharia Elétrica. III. Título.

CDU: 681.3:007.52

RECONHECIMENTO DE FACE UTILIZANDO TRANSFORMADA DISCRETA DO COSSENO BIDIMENSIONAL, ANÁLISE DE COMPONENTES PRINCIPAIS BIDIMENSIONAL E MAPAS AUTO-ORGANIZÁVEIS CONCORRENTES

Thayso Silva Guimarães

Dissertação apresentada por Thayso Silva Guimarães à Universidade Federal de Uberlândia como parte dos requisitos à obtenção do título de Mestre em Engenharia Elétrica.

Prof. Dra. Edna Lúcia Flôres Orientadora	Prof. Dr. Alexandre Cardoso Coordenador do curso de Pós-Graduação

Dedicatória

Dedico este trabalho primeiramente a Deus, pois sem Ele, nada seria possível e não estaríamos aqui reunidos, desfrutando, juntos, destes momentos que nos são tão importantes. Aos meus pais Celso e Janete e ao meu irmão Igor pelo esforço, dedicação e compreensão, em todos os momentos desta e de outras caminhadas. Em especial, a minha esposa Cristiane pelo apoio, amor e carinho a mim dedicados. As minhas amigas de laboratório Aleandra e Eliane, pelo mútuo aprendizado de vida durante nossa convivência, no campo acadêmico e particular.

Agradecimentos

Agradeço primeiramente a Deus, pela oportunidade de fazer o mestrado e pela força e sabedoria que me auxiliou durante os meus estudos.

A Prof. Dra. Edna Lúcia Flôres, pela dedicação, paciência e ensinamentos durante todo o tempo, além disso, acreditando e confiando na minha pessoa.

A minha esposa Cristiane que sempre me incentivou e me apoiou em todos os momentos em que precisei.

A todos os colegas do Mestrado que me apoiaram e a todos aqueles que me ajudaram de alguma forma para a realização deste trabalho.

À Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), pelo apoio financeiro concedido.

À minha família que tanto amo...

Resumo

A identificação de uma pessoa pela sua face é um dos métodos não-intrusivo mais efetivo em biometria, no entanto, também é um dos maiores desafios para os pesquisadores na área; consistindo em pesquisas em psicofísica, neurociência, engenharia, reconhecimento de padrões, análise e processamento de imagens, e visão computacional aplicada no reconhecimento de faces pelos seres humanos e pelas máquinas. O algoritmo proposto nesta dissertação para reconhecimento de faces foi desenvolvido em três etapas. Na primeira etapa são obtidas as matrizes características das faces utilizando a *Two-Dimensional Discrete Cosine Transform* (2D-DCT) e a *Two-Dimensional Principal Component Analysis* (2D-PCA). O treinamento da *Concurrent Self-Organizing Map* (CSOM) é realizado na segunda etapa usando as matrizes características das faces. E finalmente, na terceira etapa obtém-se a matriz característica da imagem consulta classificando-a utilizando a rede CSOM da segunda etapa. Para verificar o desempenho do algoritmo de reconhecimento de faces proposto neste trabalho foram realizados testes utilizando três bancos de imagens bem conhecidos na área de processamento de imagens: ORL, YaleA e Face94.

Palavras-chave: Reconhecimento de Faces, 2D-DCT, 2D-PCA, SOM, CSOM..

Abstract

The identification of a person by their face is one of the most effective non-intrusive methods in biometrics, however, is also one of the greatest challenges for researchers in the area, consisting of research in psychophysics, neuroscience, engineering, pattern recognition, analysis and image processing, computer vision and applied in face recognition by humans and by machines. The algorithm proposed in this dissertation for face recognition was developed in three stages. In the first stage feature matrices are derived of faces using the Two-Dimensional Discrete Cosine Transform (2D-DCT) and Two-Dimensional Principal Component Analysis (2D-PCA). The training of the Concurrent Self-Organizing Map (Csoma) is performed in the second stage using the characteristic matrices of the faces. And finally, the third stage we obtain the feature matrix of the image consulting classifying it using the CSOM network of the second step. To check the performance of face recognition algorithm proposed in this paper were tested using three well-known image databases in the area of image processing: ORL, YaleA and Face94.

Keywords: Face recognition, 2D-DCT, 2D-PCA, SOM, CSOM..

Sumário

1. INTRODUÇÃO 15

1.1. Motivação 15

1.2. Objetivos deste Trabalho 18

1.3. Estrutura deste Trabalho 19

1.4. Considerações Finais deste Capítulo 20

2. RECONHECIMENTO DE FACES HUMANAS 21

2.1. Introdução 21

2.2. A Capacidade Humana no Reconhecimento de Faces 23

2.3. Reconhecimento de Padrões 27

2.4. Reconhecimento de Faces 28

2.5. Métodos Baseados em Características 31

2.6. Métodos Holísticos 33

2.6.1. Análise de componentes principais (PCA) 35

2.6.2. Utilização de PCA no reconhecimento de faces 37

2.6.3. Análise da componente independente (ICA) 40

2.6.4. Arquiteturas ICA utilizadas no reconhecimento de faces 42

2.7. Redes neurais aplicadas no reconhecimento de faces 46

2.8. Outros Métodos para o Reconhecimento de Faces 50

2.8.1. Transformada de Fourier 51

2.8.2. Transformada discreta do cosseno 52

2.8.3. Imagens de perfil 53

2.9. Considerações Finais deste Capítulo 53

3. MAPAS AUTO-ORGRANIZAVEIS DE KOHONEN (SOM) 55

3.1. Introdução 55

3.2. Redes Neurais Artificiais (RNAs) 56

3.3. Processos de Aprendizagem 58

3.3.1. Aprendizagem supervisionada 59

3.3.2. Aprendizagem por reforço 60

3.3.3. Aprendizagem não-supervisionada 61

3.3.4. Generalização das redes neurais 63

3.3.5. Aprendizado competitivo 65

3.3.6. Estrutura básica do SOM 67

3.4. Treinamento em redes SOM 70

3.4.1. Algoritmo convencional 70

3.4.2. Observações do processo de treinamento do SOM 72

3.5. Parâmetros de Treinamento em Redes SOM 76

3.5.1. Tamanho e dimensionalidade do mapa 76

3.5.2. Topologia da rede e conectividade entre os neurônios 77

3.5.3. Inicialização dos pesos 77

3.5.4. Função de decaimento 78

3.5.5. Taxa de aprendizagem 79

3.5.6. Número de Interações 79

3.6. Considerações Finais deste Capítulo 80

4. MÉTODO PROPOSTO 81

4.1. Introdução 81

4.2. Two-Dimensional Discrete Cosine Transform (2D-DCT) 83

4.3. Two-Dimensional Principal Component Analysis (2D-PCA) 85

4.4. Concurrent Self-Organizing Maps (CSOM) 89

4.5. Algoritmo Proposto neste Trabalho 92

4.6. Considerações Finais deste Capítulo 96

5. RESULTADOS OBTIDOS 98

5.1. Introdução 98

5.2. Banco de Imagens 99

5.2.1. ORL 99

5.2.2. YaleA 100

5.2.3. Face94 101

5.3. Testes Realizados com a Base de Dados ORL 102

5.4. Testes Realizados com a Base de Dados YaleA 110

5.5. Testes Realizados com a Base de Dados Face94 117

5.6. Conclusões 125

6. CONCLUSÕES, CONTRIBUIÇÕES DESTE TRABALHO E TRABALHOS FUTUROS 127

6.1. Introdução 127

6.2. Conclusões 127

6.3. Contribuições deste Trabalho 130

6.4. Trabalhos Futuros 131

6.5. Considerações Finais deste Capítulo 132

REFERÊNCIAS 133

Lista de Figuras

Figura 2.1: Sistema genérico de reconhecimento de padrões 27

Figura 2.2: Biometria facial 32

Figura 2.3: Exemplo de concatenação de uma imagem [19] 38

Figura 2.4: Arquitetura I [20] 43

Figura 2.5: Representação das imagens bases independentes [20] 43

Figura 2.6: Arquitetura II [20]. 45

Figura 2.7: Imagens bases desconhecidas [20] 46

Figura 2.8: Rede de auto-associação e classificação para reconhecimento de faces [21]. .. 48

Figura 2.9: Descrição do processo de amostragem local da imagem [21] 49

Figura 2.10: Diagrama geral do sistema de reconhecimento de faces. (a) Típica rede de convolução; (b) diagrama em blocos do sistema utilizado para o reconhecimento de faces. 50

Figura 3.1: (a) Modelo matemático não-linear de um neurônio; (b) neurônio do sistema nervoso central [30]. 58

Figura 3.2: Diagrama de blocos representativo da aprendizagem supervisionada [30] 60

Figura 3.3: Diagrama de blocos da aprendizagem por reforço [30] 61

Figura 3.4: Diagrama de blocos da aprendizagem não-supervisionado [30] 62

Figura 3.5: Aprendizagem: erro em relação ao conjunto de aprendizado e de teste 64

Figura 3.6: Exemplo simplificado de uma rede neural competitiva 65

Figura 3.7: Atualização de pesos em uma rede neural competitiva 67

Figura 3.8: Exemplo de SOM de dimensão retangular [7] 69

Figura 3.9: Duas configurações de *grid* e seus níveis de vizinhança [7] 69

Figura 3.10: SOM com tamanho 4 x 4 e entrada com dimensão de 3 70

Figura 3.11: Atualização do BMU e de seus vizinhos em direção ao padrão de entrada x . 74

Figura 4.1: Diagrama do método linha-coluna utilizado no cálculo da 2D-DCT..... 85

Figura 4.2: Modelo de treinamento de uma rede CSOM..... 90

Figura 4.3: Modelo de classificação de uma rede CSOM 91

Figura 4.4: 2D-DCT de uma face da base de dados da ORL e a seleção dos coeficientes mais significativos 93

Figura 4.5: Diagrama de blocos da fase de treinamento do modelo proposto neste trabalho 94

Figura 4.6: Diagrama de blocos da fase de reconhecimento/classificação do modelo proposto neste trabalho. 95

Figura 5.1: Exemplo de imagens de faces presentes no banco de faces ORL [39]. 100

Figura 5.2: Exemplo de imagens de faces presentes no banco de faces YaleA [41]..... 101

Figura 5.3: Exemplo de imagens de faces presentes no banco de faces Face94 [42]..... 102

Figura 5.4: Seleção dos coeficientes mais significativos da 2D-DCT de uma face da base de dados ORL 103

Figura 5.5: Autovalores normalizados do conjunto de treinamento da base de dados ORL. 104

Figura 5.6: Gráfico dos resultados obtidos das taxas de reconhecimento dos grupos de testes utilizados nas Tabelas 5.1 e 5.2..... 107

Figura 5.7: Gráfico dos resultados obtidos utilizando-se as redes CSOM e SOM na base de dados ORL 108

Figura 5.8: Faces identificadas incorretamente utilizando o algoritmo *Leave-one-out*. na base de dados ORL. 110

Figura 5.9: Seleção dos coeficientes mais significativos da 2D-DCT de uma face da base de dados YaleA 111

Figura 5.10: Autovalores normalizados do grupo de treinamento da base de dados YaleA]. 112

Figura 5.11: Gráfico dos resultados obtidos das taxas de reconhecimento dos grupos de testes utilizados nas Tabelas 5.4 e 5.5..... 115

Figura 5.12: Seleção dos coeficientes mais significativos da 2D-DCT de uma face da base de dados Face94.. 118

Figura 5.13: Autovalores normalizados do grupo de treinamento da base de dados Face94.. 119

Figura 5.14: Gráfico dos resultados obtidos das taxas de reconhecimento dos grupos de testes utilizados na Tabela 5.7 121

Figura 5.15: Gráfico dos resultados obtidos do algoritmo proposto neste trabalho variando o número de autovalores utilizando-se Face94. 122

Figura 5.16: Imagens das faces de um indivíduo do banco de dados Face94. 124

Lista de Tabelas

Tabela 5.1: Resultados obtidos utilizando a base de dados ORL. 105

Tabela 5.2: Resultados obtidos com uma única rede SOM utilizando a base de dados ORL 106

Tabela 5.3: Resultados obtidos nos testes variando a quantidade de imagens utilizando-se o banco de dados ORL. 109

Tabela 5.4: Resultados obtidos utilizando a base de dados YaleA. 113

Tabela 5.5: Resultados obtidos com uma única rede SOM utilizando a base de dados YaleA. 114

Tabela 5.6: Resultados obtidos variando a quantidade de imagens utilizando-se a banco de dados YaleA. 116

Tabela 5.7: Resultados obtidos utilizando a base de dados Face94. 120

Tabela 5.8: Resultados obtidos nos testes realizados variando-se a quantidade de imagens utilizando-se o banco de dados Face94. 123

Lista de Abreviaturas

2D-DCT	*Two-Dimensional Discrete Cosine Transform*
2D-PCA	*Two-Dimensional Principal Component Analysis*
CSOM	*Concurrent Self-Organizing Maps*
SOM	*Self-Organizing Maps*
ORL	*Olivetti Research Laboratory*
PCA	*Principal Component Analysis*
ICA	*Independent Component Analysis*
LDA	*Linear Discriminant Analysis*
DCT	*Discrete Cosine Transform*
CP	*Componentes Principais*
BMU	*Best Mmatch Unit*
CIS	*Computer and Information Science*
LVQ	*Learning Vector Quantisation*

Capítulo 1

INTRODUÇÃO

1.1 Motivação

Nos últimos anos foram propostas muitas linhas de pesquisa para a autenticação/reconhecimento da identidade de uma pessoa no campo da biometria, como reconhecimento de impressão digital, da retina, da voz, geometria da mão, reconhecimento de íris e também análise facial.

Todos os métodos de reconhecimento pelas características biométricas são caracterizados por limitações intrínsecas e muitos requerem uma cooperação ativa do usuário, e isto têm como conseqüência uma baixa aceitabilidade social.

O controle de acesso e identificação pelo reconhecimento de faces é uma tecnologia promissora no campo da identificação pessoal, pois traz algumas vantagens em relação a outras formas de identificação biométrica. Por exemplo:

- Em relação à impressão digital, pesquisas mostram que 5% das pessoas não possuem uma “boa” impressão digital. As razões para isto são: 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 de impressão digital é muito

menor do que a face, portanto cortes e contusões afetam mais as digitais do que as faces;

- O reconhecimento de íris possui grau de acerto praticamente 100%, pois é uma característica única de cada ser humano. Porém, o problema está na captura da íris, visto que a câmera deve estar alinhada perfeitamente em relação à íris, para isso é necessário a cooperação do indivíduo; e

- A análise de voz para reconhecimento é um meio natural e pouco inoportuno, porém o reconhecimento da face é o mais prático e cômodo. Pois permite que o usuário seja identificado simplesmente passando na frente de uma câmera.

A identificação de uma pessoa pela sua face é um dos métodos não-intrusivo mais efetivo em biometria. No entanto, também é um dos maiores desafios para os pesquisadores na área; consistindo em pesquisas em psicofísica, neurociência, engenharia, reconhecimento de padrões, análise e processamento de imagens, e visão computacional aplicada no reconhecimento de faces pelos seres humanos e pelas máquinas.

Em princípio a análise de imagens de faces parece ser o melhor modo para executar autenticação de identidade de um individuo, e também o mais aceitável pelas pessoas, porque sempre os seres humanos associam as características sociais e pessoais a uma determinada face guardada na memória desses seres, que posteriormente é lembrada por eles para o reconhecimento.

Uma das dificuldades do reconhecimento de um indivíduo por sua face é em relação à alta dimensionalidade do espaço necessário para o armazenamento e análise dos dados encontrados em imagens naturais de faces obtidas por câmeras.

Um sistema de reconhecimento de faces genérico deve considerar diversas questões técnicas relativas tanto ao tratamento das faces, quanto aos processos envolvidos na tarefa de reconhecimento em si. Existem diversas etapas associadas ao processo de reconhecimento automático de faces:

1. Detecção: considerando 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: selecionar as informações da face que serão utilizadas para representá-la; e

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 registrado primeiramente no sistema e codificado para uma futura busca. Uma face consulta é apresentada ao sistema e comparada com o banco de dados; o reconhecimento é efetuado quando a face consulta possui maior similaridade com uma face codificada da base de dados.

O desafio para os sistemas de reconhecimento de face é reconhecer a face apesar de transformações: mudança de escala, translação, rotação, iluminação, alteração de expressão, idade, óculos e disfarces (chapéu, bigode, barba, boné, peruca, etc) que são inerentes às faces.

Este capítulo apresenta os objetivos e a estrutura deste trabalho. Finalmente, são realizadas as considerações finais deste capítulo.

1.2 Objetivos deste Trabalho

O objetivo principal deste trabalho é desenvolver um algoritmo para o reconhecimento de faces em aplicações *one-to-many*. Neste algoritmo são aplicadas nas faces técnicas de redução de dimensionalidade como *Two-Dimensional Discrete Cosine Transform* (2D-DCT) e *Two-Dimensional Principal Component Analysis* (2D-PCA) e utiliza-se a rede neural *Concurrent Self-Organizing Map* (CSOM) na classificação das faces.

Nesta dissertação não são considerados os dados tridimensionais ou geometricamente detalhados da face, mas as características estatísticas da distribuição dos pixels na imagem da face em escala de cinza. Portanto, o objetivo é construir um modelo computacional para o reconhecimento de faces que seja dinâmico e eficiente em termos de taxa de reconhecimento e tempo de processamento.

Outros objetivos deste trabalho são:

- Apresentar os principais métodos de reconhecimento de faces;
- Mostrar os pontos principais da 2D-DCT, da 2D-PCA e da rede CSOM no algoritmo proposto nesta dissertação;
- Realizar testes comparando as redes SOM e CSOM; e

- Apresentar os resultados obtidos do algoritmo proposto utilizando os bancos de faces da ORL, YaleA e Face94.

1.3 Estrutura deste Trabalho

Este capítulo apresenta a motivação, os objetivos e a estrutura deste trabalho. Finalmente, são realizadas as considerações finais deste capítulo.

O capítulo 2 apresenta a capacidade humana no reconhecimento de faces, o reconhecimento de padrões, o reconhecimento de faces pelos métodos holísticos e baseados em características, redes neurais aplicadas no reconhecimento de faces e outros métodos utilizados no reconhecimento de faces. E finalmente, são realizadas considerações finais desse capítulo.

O objetivo do capítulo 3 é fornecer fundamentos sobre redes neurais artificiais enfatizando a aplicação do SOM em problemas de classificação automática de faces. É apresentada a arquitetura e o processo de treinamento do SOM, analisa-se o seu funcionamento e são discutidos os aspectos de convergência do SOM. Finalmente, são realizadas considerações finais desse capítulo.

No capítulo 4 é apresentado o algoritmo proposto nesta dissertação para o reconhecimento de faces. São apresentadas as técnicas 2D-DCT e 2D-PCA utilizadas para a obtenção das matrizes características das faces e a rede neural *Concurrent Self-Organizing Maps* (CSOM) usada na classificação das faces. Finalmente, são realizadas considerações finais desse capítulo.

No capítulo 5 são apresentados os resultados obtidos do algoritmo de reconhecimento de faces proposto neste trabalho utilizando-se três bancos de imagens bem conhecidos na área de processamento de imagens: ORL, YaleA e Face94. Finalmente, são realizadas conclusões sobre esses resultados

O capítulo 6 apresenta as conclusões e contribuições deste trabalho e os trabalhos futuros que poderão ser desenvolvidos a partir desta dissertação. Finalmente, são realizadas as considerações finais desse capítulo.

1.4 Considerações Finais deste Capítulo

Este capítulo apresentou a motivação, os objetivos e a estrutura deste trabalho.

O próximo capítulo apresenta a capacidade humana no reconhecimento de faces, o reconhecimento de padrões, o reconhecimento de faces pelos métodos holísticos e baseados em características, redes neurais aplicadas no reconhecimento de faces e outros métodos utilizados no reconhecimento de faces. E finalmente, são realizadas considerações finais desse capítulo.

Capítulo 2

RECONHECIMENTO DE FACES HUMANAS

2.1 Introdução

A identificação de uma pessoa pela sua face é um dos métodos não-intrusivo mais efetivo em biometria, no entanto, também é um dos maiores desafios para os pesquisadores na área; consistindo em pesquisas em psicofísica, neurociência, engenharia, reconhecimento de padrões, análise e processamento de imagens, e visão computacional aplicada no reconhecimento de faces pelos seres humanos e pelas máquinas.

O 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 na compreensão de imagens biométricas. Esse reconhecimento tem recebido recentemente uma atenção significativa, especialmente durante os últimos anos, e ao menos duas razões esclarecem essa tendência: a primeira é a larga escala de aplicações comerciais, militares e de segurança civil, e a segunda é a disponibilidade de tecnologias praticáveis após décadas de pesquisa.

Mesmo que os sistemas atuais de reconhecimento automático de faces 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 foi resolvido. Pode-se considerar que os sistemas atuais estão ainda distantes da potencialidade do sistema de percepção humana, 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% [1].

Como na maioria dos outros sistemas de medidas biométricas, o objetivo geral do reconhecimento de faces é conseguir um nível elevado de desempenho na identificação de uma face apresentada ao sistema em relação 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, baixo custo do equipamento, etc.

Segundo Barrett [2], existem fatores adversos como condições de luminosidade, ruído na imagem, variações na expressão facial, presença de óculos, mudanças no cabelo, translação e rotação da face em relação à postura. Estes fatores estão presentes nas aplicações estáticas relacionadas a fotografias em passaportes, cartões de crédito, carteiras de identidade, carteiras de motorista, e retrato falado, onde pode-se destacar todos esses fatores somados a complexidade de processamento de imagens para as aplicações de vigilância em tempo-real.

Este capítulo apresenta a capacidade humana no reconhecimento de faces, o reconhecimento de padrões, o reconhecimento de faces pelos métodos holísticos e baseados em características, redes neurais aplicadas no reconhecimento de faces e outros métodos utilizados no reconhecimento de faces. Finalmente, no fim do capitulo são realizadas as considerações finais.

2.2 A Capacidade Humana no Reconhecimento de Faces

Os conceitos de reconhecimento de faces considerando o sistema perceptivo e cognitivo humano é 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, onde é necessário tentar compreender a arquitetura considerando os estímulos e processos cognitivos presentes nos seres humanos para efetuar o reconhecimento de faces.

Em uma análise superficial em relação a arquitetura do sistema de reconhecimento de faces humanas pode-se observar 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 recuperar possivelmente essas informações e estabelecer um reconhecimento futuro.

Em muitas situações o conhecimento contextual com relação ao ambiente em que o individuo 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 esses estímulos, mas o cérebro humano possui uma desvantagem: sua capacidade de armazenamento para o reconhecimento preciso é pequena [3].

Existem evidências sugestivas de que a capacidade visual humana para o reconhecimento de faces é um processo dedicado consistindo de informações holísticas (características globais) e traços do rosto (características locais), não meramente uma aplicação geral de reconhecimento de objetos [4]. Isso encoraja as aplicações em reconhecimento de faces a serem também especificas, 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.

Segundo Zhao e outros [1], algumas características potencialmente relevantes para o desenvolvimento de sistemas de reconhecimento de faces são:

- Análise holística e/ou de características (traços do rosto): ambas informações são cruciais para a percepção e o reconhecimento das faces. 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 utilizadas. 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: o cabelo, o contorno da face, os olhos, e a boca (não necessariamente nesta ordem) são determinados como sendo importantes para perceber e reconhecer as faces. Diversos estudos mostraram que o nariz apresenta um papel insignificante; isto pode ser devido ao fato que quase todos esses estudos foram realizados utilizando-se imagens frontais. No reconhecimento de faces utilizando-se 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. A parte superior da face é mais útil para o reconhecimento de faces do que a parte mais inferior. Os atributos relacionados a 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 vêem em seguida, deixando as faces habituais por último, em termos de facilidade para serem reconhecidas;

- Caricaturas: uma caricatura pode ser definida formalmente 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 contém tanta informação quanto fotografias, mas podem capturar as características mais importantes de uma face;

- Distinção: 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 realizada 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 utilizados para a detecção e para a identificação realizada pelo cérebro;

- Estudo na análise de freqüência espacial: as informações contidas nas faixas de baixa freqüência espacial constituem um fator dominante no reconhecimento de faces. Dependendo da tarefa específica de reconhecimento, os componentes de baixa e alta freqüência podem estabelecer relações diferentes. A classificação de sexo, por exemplo, pode com sucesso ser realizada utilizando-se componentes de baixa freqüência, enquanto a identificação requer a utilização de componentes de alta freqüência. Os componentes de baixa freqüência contribuem para a descrição

global, e as componentes de alta freqüência contribuem para os detalhes mais finos necessários na identificação;

- O reconhecimento com relação ao ponto de vista (invariante ou não): os desenvolvimentos no reconhecimento visual de objetos foram moldados dentro de uma estrutura teórica, em que as vistas diferentes dos objetos são analisadas de maneira que permita o acesso (pela maior parte) as 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 do ponto de vista de um perfil a outro, é pobre mas a generalização de uma vista de três quartos (esquerda) a outra (direita) é muito boa;

- Expressões faciais: baseado em estudos neuropsicológicos, existem evidências de que a análise de expressões faciais são realizadas em paralelo com o reconhecimento; e

- Reconhecimento de faces em raças diferentes: os seres humanos reconhecem as pessoas de sua própria raça melhor do que as pessoas de outras raças. Isto pode ser porque esses seres codificam uma face média com atributos médios, e essas 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, como mostrado na Figura 2.1. Estes sistemas podem ser compostos pelas seguintes partes [5,6]:

- 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; e
- Classificador; analisa um padrão obtido e toma uma decisão.

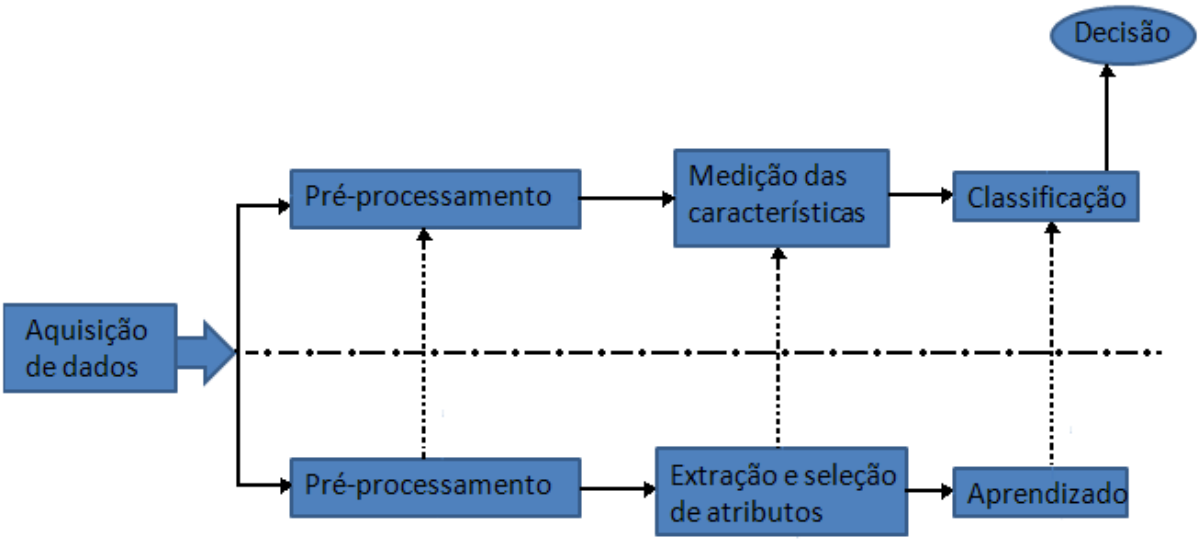


Figura 2.1 – Sistema genérico de reconhecimento de padrões.

O classificador toma decisões baseando-se no aprendizado realizado a partir de um conjunto de treinamento, que contém exemplos de padrões de todas as classes existentes no sistema. Em reconhecimento estatístico de padrões, a classificação é realizada utilizando-se estimativas de distribuições probabilísticas.

O sistema de reconhecimento de padrões é avaliado por 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: considere uma pessoa desconhecida, extrair suas características para obter seu vetor de características e determinar a classe a qual pertença essa pessoa.

2.4 Reconhecimento de Faces

As pesquisas em reconhecimento de faces utilizando imagens 2-D em escala de cinza apresentam-se geralmente em três categorias: métodos baseados em características (locais), métodos holísticos (globais) e 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 utilizada 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 que 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 [1].

Os métodos holísticos tratam os dados da imagem simultaneamente sem tentar localizar os 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. Eles 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 do reconhecimento. Entretanto, tais técnicas são sensíveis a variações em posição, escala e assim por diante.

Uma fusão dos métodos baseados em características e holísticos é chamada de método híbrido: as idéias consolidadas nesses sistemas estão embasadas no sistema de percepção humana, que utiliza características locais e a região inteira da face para efetuar o reconhecimento, sendo assim, portanto um sistema de reconhecimento artificial deveria utilizar 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 [7], 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 alinhadas, nenhum conseguiu sucesso em ambientes reais com bases de dados grandes, onde a posição e a escala da face são desconhecidas.

Sirovich e Kirby [8] 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 [9,10], demonstraram que, o erro residual presente na codificação de faces proposta por Kirby e Sirovich [11] poderia ser utilizado 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 [12].

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 em relação as bases de dados grandes (superiores a 100 imagens).

As técnicas utilizadas para o reconhecimento de faces são muitas e variadas, mas todas consistem nas mesmas três etapas básicas [13]:

1. Segmentação/ detecção da face na cena.

   Exemplos:

   - Mapeamento das bordas;

- Análise da pigmentação da pele; e
- Redes neurais.

2. Extração de características presentes na face:

   Exemplos:

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

3. Decisão:

   Exemplos:

   - Identificação;
   - Reconhecimento de faces; e
   - Categorização (raça, sexo, idade, expressão facial, etc.).

Algumas das técnicas mais importantes na área de reconhecimento de face serão discutidas brevemente nos próximos sub-itens deste capítulo.

2.5 Métodos Baseados em Características

Os sistemas que utilizam as feições, ou as características geométricas da face, usam as informações contidas nas características locais das faces para obter um vetor característico com a finalidade 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 esses componentes definem a aparência de uma face e essas variações podem também ser utilizadas para classificar e distinguir diferentes faces. Por exemplo, esses parâmetros podem ser as distâncias entre os olhos, o tamanho da boca, e a distância entre a boca e o nariz, como mostrado na Figura 2.2. Todos esses tipos de medidas são exemplos de biometria da face humana.

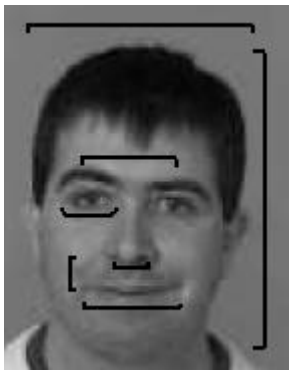


Figura 2.2 – Biometria facial

As primeiras implementações em reconhecimento de faces foram realizadas utilizando os conceitos de biometria facial. O sistema fundamenta-se na localização automática de características tais como cantos dos olhos, narinas, a boca etc., então um conjunto de dezesseis parâmetros faciais, que corresponde a essas características é utilizado para definir um vetor característico, e para o reconhecimento é realizada uma medida Euclidiana simples da distância entre o vetor de teste e os vetores do banco de dados para definir o grau de similaridade entre eles e possivelmente associar um individuo.

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 as 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 no desempenho do sistema. Desta forma, as mais infindáveis restrições impostas ao bom funcionamento do método tornam sua aplicação no mundo real pouco robusta e eficiente.

2.6 Métodos Holísticos

O reconhecimento de faces aplicando os métodos holísticos utiliza as informações globais das faces para executar o reconhecimento. Esta informação é representada por um vetor de face em um sistema de coordenadas reduzidas, e esse vetor é derivado diretamente da informação dos pixels das imagens da face. Esse vetor no sistema de coordenadas reduzidas captura distintamente a variação entre as faces individuais diferentes da média e conseqüentemente é utilizado 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 $P(m,\mathrm{n})$ pode ser representada por um vetor x e $R^{m,n}$, com os seus pixels concatenados em linhas ou colunas para formar um só vetor no espaço $R^{m,n}$.

Apesar do espaço dimensional ser alto, ele representa o mundo real e natural das imagens da face. Ele não é fácil 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 [13].

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) [9,10], Análise de Componente Independente (ICA) [14], e Análise Discriminante Linear (LDA) [15,16,17].

Os métodos PCA, ICA e LDA são baseados em análises estatísticas dos vetores de face, e em transformações lineares e não lineares dos sistemas de coordenadas onde se encontram esses vetores.

A análise de discriminantes lineares (LDA), também conhecidos como discriminantes lineares de Fisher, é uma técnica que se tornou muito comum para reconhecimento de faces, principalmente a partir de 1997, com a publicação do artigo [15]. Nesse artigo, os autores comparam o PCA com o LDA e mostram que o espaço de características criado pela transformação LDA proporcionou resultados de classificação muito melhores do que o espaço criado pela transformada PCA para o reconhecimento de pessoas em imagens com grandes variações de iluminação.

O PCA e o ICA serão brevemente explicados neste capítulo, como suas aplicações em reconhecimento de faces. 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 por 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 os vetores, como por exemplo, a distância Euclidiana, do Cosseno, da Mahalanobis, etc.

Sendo assim, definindo *X* como o conjunto de todas as faces de treinamento, ou seja, $X=(x_1, x_2, ...x_i, ...x_N)$ é uma matriz *pt.N*, onde x_i é um vetor face de dimensão *pt (concatenação da imagem m.n = pt).* Onde *pt* representa o número total de pixels na imagem da face, e N a quantidade de faces diferentes presentes 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, conforme a Equação (2.1).

$$Y=W^T.X \tag{2.1}$$

onde:

Y - matriz com a representação dos vetores originais no novo espaço vetorial (*d.N*);

X (*pt.N*) - vetores do espaço original; e

W (*pt.d*) - matriz transformação.

Pode-se verificar que *d* é a dimensão do espaço Y e *pt* é o tamanho dos vetores, lembrando que *d* sempre é menor do 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-Loêve consiste em reduzir a dimensionalidade dos dados efetuando

uma análise estatística baseada na variância, e na redundância presente nos dados. Quando essa técnica é utilizada em aplicações de faces recebe o nome de *Eigenfaces (Autofaces)*, que é derivado do prefixo alemão *eigen* (auto), que significa próprio ou individual.

Um dos problemas básicos ao analisar e visualizar um conjunto de dados multidimensional, como é o caso das faces, é a representação dos dados e a sua estruturação em um subespaço de dimensão mais baixa. A implementação geral para resolver esse problema é representar os dados de modo que as distâncias entre os pontos, que representam as amostras no subespaço de dimensão mais baixa, corresponda ao máximo possível as dissimilaridades entre os pontos no espaço dimensional original. Originando um dos mais famosos métodos nessa tarefa que é a técnica PCA

O objetivo do PCA é encontrar um subespaço 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 subespaço das principais componentes minimiza a correlação cruzada entre as amostras dos dados [18].

Teoricamente é desejável extrair as informações relevantes contidas na imagem da face, codificando essa 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 onde é realizada qualquer análise dos traços do rosto.

Estatisticamente pode-se aplicar a técnica PCA encontrando os *eigenvectors* (autovetores), que constituem as bases ortonormais do subespaço de faces e seus respectivos *eigenvalues* (autovalores) obtidos da matriz de covariância formada pela correlação entre os pixels. Em resumo, cada imagem da face no conjunto de treinamento pode ser representada

exatamente em termos de uma combinação linear dos autovetores, e os coeficientes dessa combinação são os novos representantes da face no subespaço dimensional.

2.6.2 - Utilização de PCA no reconhecimento de faces

O método proposto por Turk e Pentland [9], é uma abordagem holística, que considera o reconhecimento de face como um problema de reconhecimento 2-D. As imagens de faces são projetadas em um espaço de características ("espaço de faces") que melhor representa a variação entre as faces conhecidas. Esse espaço é definido pelas "autofaces", que são os autovetores do conjunto de faces. O reconhecimento de uma face é realizado pela sua projeção no subespaço gerado pelas autofaces e então pela comparação da posição obtida com a posição de indivíduos conhecidos.

Considerando que a Equação (2.1) é a representação de uma transformação linear e visando encontrar a matriz transformação W constituída pelos autovetores ("autofaces") tem-se: cada imagem da face definida por uma matriz de pixels com intensidades de cinza, então definindo os vetores Γ_l, Γ_2, Γ_3, ... Γ_M associados aos pixels das imagens $I\ (x,\ y)$ concatenados por linhas formando vetores $N_x.N_y$ conforme ilustrado na Figura 2.3.

$$= \begin{bmatrix} x_{11} & x_{12} & \dots & x_{1N_y} \\ x_{21} & x_{22} & \dots & x_{2N_y} \\ \vdots & \vdots & \ddots & \vdots \\ x_{N_x 1} & x_{N_x 2} & \dots & x_{N_x N_y} \end{bmatrix} \cong \begin{bmatrix} x_{11} \\ x_{12} \\ \vdots \\ x_{1N_y} \\ x_{21} \\ x_{22} \\ \vdots \\ x_{2N_y} \\ \vdots \\ x_{N_x 1} \\ x_{N_x 2} \\ \vdots \\ x_{N_x N_y} \end{bmatrix}$$

Figura 2.3 – Exemplo de concatenação de uma imagem [19].

Desta forma, seja $\Gamma = [\Gamma_1, \Gamma_2, \Gamma_3, ..., \Gamma_M]$ o conjunto de imagens de treinamento (onde estas imagens são parte de um banco de dados) e M a quantidade de imagens do conjunto de treinamento.

De acordo com Turk e Pentland [9], existe muita informação redundante na face e ela pode ser representada somente pela variância que ocorre em torno de uma média, isto é, uma imagem de face pode ser representada com poucas componentes principais devido a sua redundância. A face média é calculada pela Equação (2.2).

$$\Psi = \frac{1}{M} \sum_{n=1}^{M} \Gamma_n \tag{2.2}$$

Utilizando-se o valor definido pela media, são calculados os desvios de cada imagem em relação a média, pela Equação (2.3).

$$\phi_n = \Gamma_n - \Psi \tag{2.3}$$

O conjunto de vetores definido por ϕ está normalizado em relação a média, e pode ser utilizado no cálculo dos autovetores μ_k e autovalores λ_k da matriz de covariância obtida da Equação (2.4).

$$C = \frac{1}{M}\sum_{n=1}^{M}\phi_n\phi_n^T = A.A^T \tag{2.4}$$

onde:

A – matriz, [ϕ_1, ϕ_2, ..., ϕ_M]

Os autovetores calculados utilizando-se a matriz de covariância formam as bases ortonormais do subespaço dimensional, portanto cada face representada por um vetor no espaço original deve ser identificada no subespaço, mas primeiramente deve-se escolher de forma heurística os *M'* primeiros autovalores associados aos primeiros autovetores, que possuem valores numericamente expressivos, e consequentemente estão relacionados as maiores variâncias no conjunto de dados, formando assim um espaço vetorial comprimido com relação ao original. Assim, as faces de treinamento são projetadas pela Equação (2.5) no subespaço definido pelos autovetores selecionados.

$$\omega_n = \mu_n^T.(\Gamma - \Psi) \quad \text{para n = 1, ...M'} \tag{2.5}$$

Portanto os ω_n definem um vetor de pesos $\Omega^T(i) = [\omega_1^i, \omega_2^i, ..., \omega_{M'}^i]$, i =1, ..., M onde M é a quantidade de imagens, que expressa cada face de treinamento no subespaço definido pelos autovetores formando uma nuvem de pontos de uma determinada região do subespaço caracterizando um espaço de faces, ou seja, tem-se uma região que limita e engloba todos os pontos representados pelos vetores $\Omega^T(i)$. Cada um desses vetores descreve a contribuição de cada autovetor na representação da imagem [19].

Na identificação, o sistema encontra os pesos ω que representam a face de entrada no subespaço, isto é conseguido projetando-se o vetor Γ_{nova} no subespaço, pela Equação (2.6).

$$\omega_n = \mu_n^T.(\Gamma_{nova} - \Psi) \text{ para n = 1, 2, ..., M'} \tag{2.6}$$

Os pesos formam um vetor $\Omega^T(nova) = [\omega_1, \omega_2, ..., \omega_{M'}]$, 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 como alguma classe de treinamento (onde cada pessoa é uma classe), ou como uma face desconhecida, ou simplesmente como uma imagem da face.

2.6.3 Análise da componente independente (ICA)

O ICA assume que um conjunto de variáveis observadas pode ser expresso como combinações lineares de variáveis desconhecidas e linearmente independentes. De alguma

forma, o ICA separa os sinais gravados, em um conjunto de variáveis estatisticamente independentes.

Matematicamente, assume-se que m variáveis $x_1, x_2, ..., x_m$ são observadas, cada uma é uma combinação linear de m variáveis desconhecidas $u_1, u_2, ..., u_m$, como mostrado na Equação (2.7).

$$x_j = \sum_{i=1}^{m} a_{ji} u_i \tag{2.7}$$

A Equação (2.7) pode ser escrita na forma matricial, pela Equação (2.8).

$$x = Au \tag{2.8}$$

A Equação (2.7) descreve como o dado gravado x é gerado pelas fontes u. As fontes u_i, chamadas de componentes independentes, não podem ser observadas diretamente. A matriz dos coeficientes A, denominada de matriz de misturas, também é desconhecida.

O ICA fornece um modo de estimar tanto A quanto u. Para estimar a matriz A, o ICA possui como vantagem o fato de seus componentes serem estatisticamente independentes. A explicação de estimar A é que ele não é gaussiano e, uma vez conhecendo a matriz A e supondo que a mesma seja inversível, a sua inversa W pode ser aplicada para obter os componentes independentes: $u = Wx$.

A principal diferença entre o ICA e o PCA é que, o PCA assume que os componentes são não correlacionados e o ICA assume que os componentes são independentes. O objetivo do PCA é encontrar os componentes de maior variância dos dados, enquanto que no ICA, os

componentes independentes produzidos, fornecem um mecanismo de separação entre os dados que são assumidos serem independentes [19].

2.6.4 Arquiteturas ICA utilizadas no reconhecimento de faces

Bartleet e outros [20] apresentaram duas arquiteturas para o reconhecimento de face. Na primeira, chamada de arquitetura-I, as imagens das faces são descritas como combinação linear de um conjunto desconhecido de componentes estatisticamente independentes e os componentes obtidos nessa arquitetura podem ser vistos como um conjunto de imagens de características faciais independentes. Na arquitetura-II, as representações dos coeficientes (pesos) são estatisticamente independentes e, os componentes obtidos podem ser vistos como um conjunto de imagens com a aparência de uma face, assim como no PCA

Na arquitetura-*I*, as faces originais, *X*, são descritas como combinação linear de um conjunto desconhecido de imagens base estatisticamente independentes, *S*, com os coeficientes dados por uma matriz de mistura *A*. A matriz de pesos W_I aprendida pelo ICA é utilizada para recuperar um conjunto de imagens base (componentes) independentes nas linhas de $U = [u_1, u_2, ..., u_N]$, como mostrado na Figura 2.4.

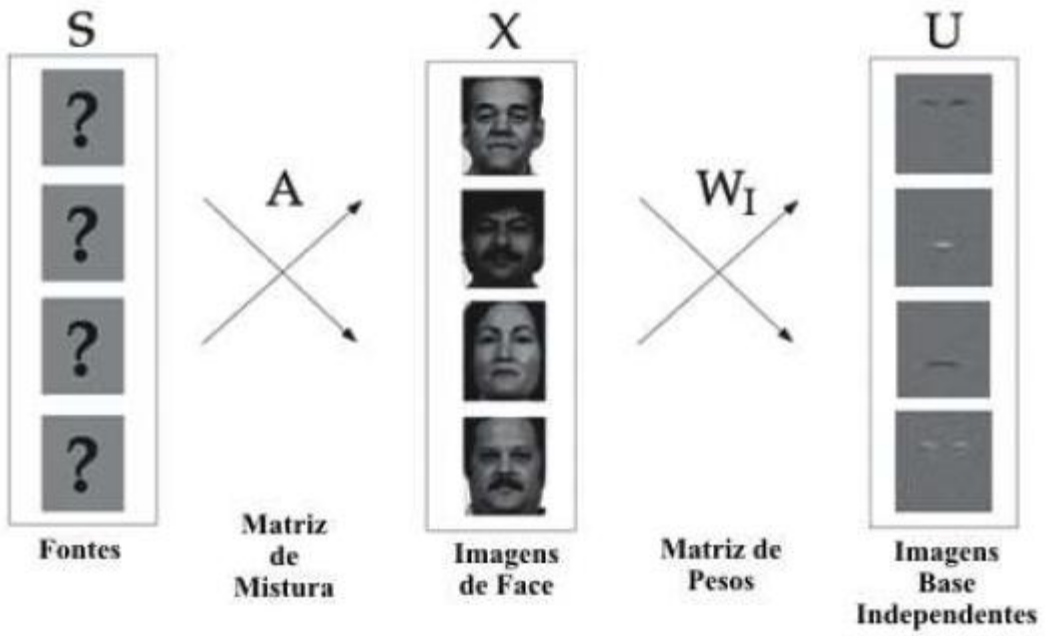


Figura 2.4 – Arquitetura I [20]

Na arquitetura-*I* a matriz de entrada, *X*, é organizada de modo que as linhas representem as imagens diferentes e as colunas representem os pixels diferentes. A projeção das imagens originais, *X*, no vetor de pesos, W_I aprendido pelo ICA, produz as imagens base independentes, *U*, tal que $U = W_I X$. A Figura 2.5 mostra as imagens base independentes obtidas nessa arquitetura.

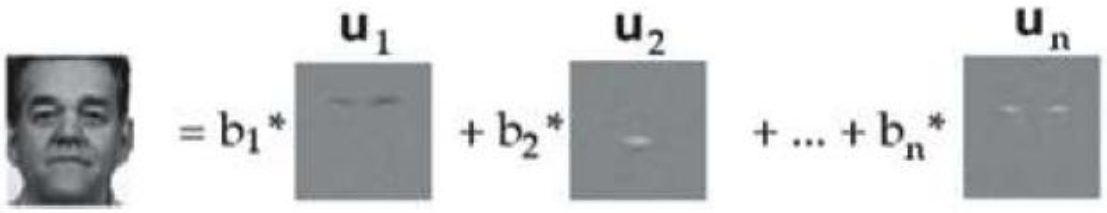


Figura 2.5 - Representação das imagens bases independentes [20]

Inicialmente, Bartlett e outros [20] aplicaram o PCA para projetar os dados em um subespaço de dimensão *m* de modo a controlar o número de componentes independentes produzidos pelo ICA, permitindo-lhe criar subespaços *m*-dimensionais.

Considere P_m uma matriz de autovetores de um conjunto de faces obtidos do PCA. Como as linhas da matriz de entrada são variáveis e as colunas são observações, o ICA é representado por $P^T{}_m$. Os m componentes independentes nas linhas de U são calculados utilizando-se a Equação (2.8).

$$U = W_I P_m^T \tag{2.8}$$

Em seguida, é calculada a matriz dos coeficientes ICA, B, como uma combinação linear das componentes independentes, U, utilizando-se a Equação (2.9).

$$B = RW_I^{-1} \tag{2.9}$$

Considerando R_m a matriz dos coeficientes PCA, tem-se a Equação (2.10).

$$R_m = XP_m \qquad (2.10$$

onde:

$$X = R_m P_m^T \tag{2.11}$$

A partir da Equação (2.8) e, supondo que W_I é inversível, obtém-se a Equação (2.12).

$$P_m^T = W_I^{-1} U \tag{2.12}$$

Cada linha de B contém os coeficientes da combinação linear das componentes estatisticamente independentes com as imagens das faces contidas nas linhas de X. Além

disso, X é a reconstrução dos dados originais em relação ao erro quadrático mínimo, como no PCA.

Quando uma nova imagem (imagem teste), X_{teste} é solicitada, ela é projetada no subespaço de face, $R_{test} = X_{test}P_m$, e é multiplicada por W_I produzindo um vetor de características ICA $B_{test} = R_{test}W_I^{-1}$. Este vetor é comparado com os vetores característicos de cada indivíduo na base de dados das faces através da distância euclidiana.

O objetivo do ICA na arquitetura-II é encontrar uma representação em que os coeficientes utilizados para representar as faces sejam estatisticamente independentes, como ilustra a Figura 2.6. Enquanto, na arquitetura-I as imagens base (componentes), u_i obtidas na arquitetura-I são estatisticamente independentes, os coeficientes, b_i que representam cada face não necessariamente são independentes.

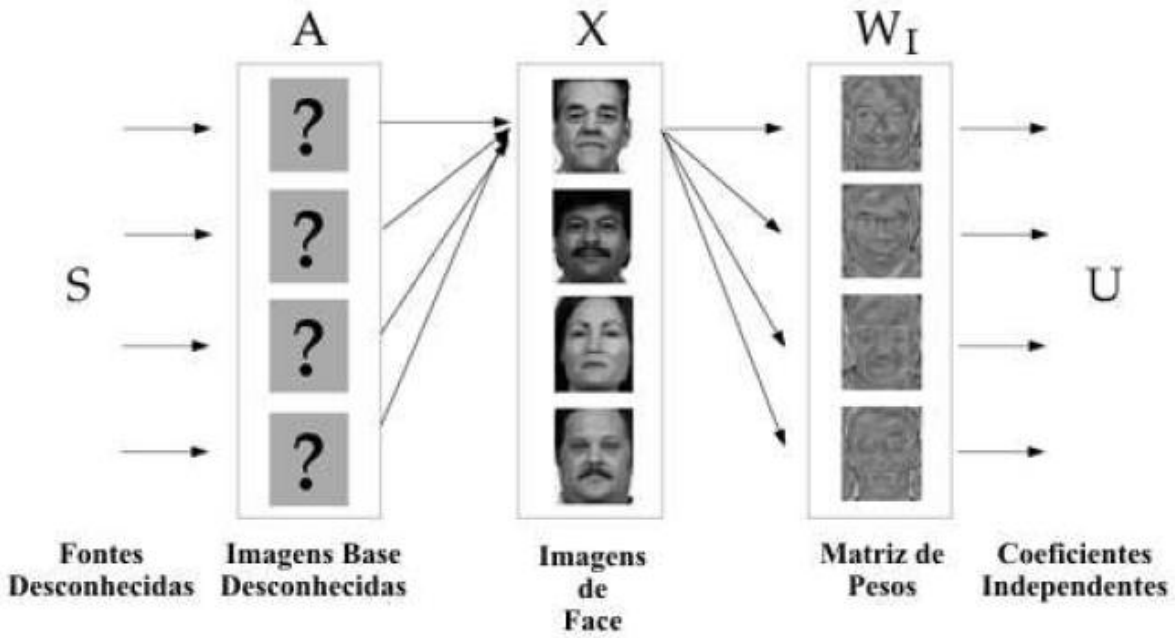


Figura 2.6 - Arquitetura II [20].

Na arquitetura-II a matriz de entrada, X, é organizada de modo que as linhas representem pixels diferentes e as colunas representem imagens diferentes. As colunas da matriz $U = W_IX$ fornecem uma representação das imagens da matriz de entrada X. Cada

coluna de *U* contém os coeficientes das imagens base em *A*, para reconstruir cada imagem de *X*, como mostrado na Figura 2.7.

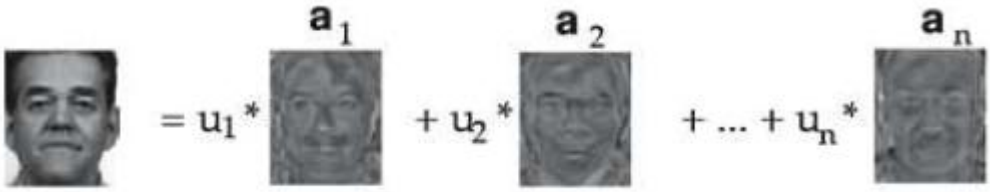


Figura 2.7 - Imagens bases desconhecidas [20]

.

2.7 Redes neurais aplicadas no reconhecimento de faces

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 que é formado por variáveis de distância entre os olhos, nariz, boca, etc. Mas existem alguns métodos onde as redes neurais são aplicadas de forma holística utilizando as características globais da face.

Uma rede neural P*erceptron 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. Por exemplo, uma imagem de baixa resolução, com 128 x 128 pixels, necessita de $N = 16385$ neurônios de entrada. A camada de saída deve ter ao menos um neurônio para cada pessoa, e a camada intermediária (*hidden-layer*) com *p* neurônios é a responsável por extrair as características dos dados de entrada. A camada intermediária deve ser escolhida de forma adequada, pois forma um vetor característico comprimido com relação ao vetor de entrada [2].

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 [21] utilizaram duas redes *Perceptron Back-Propagation*, como ilustrado na Figura 2.8.

Na Figura 2.8, a primeira rede opera no modo auto-associativo descrito por Bourlard e Kamp [22] 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 $p << n$. A rede considera o vetor de face x como entrada, e é treinada para obter uma saída y, que é a "melhor aproximação" à 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.

O sistema da Figura 2.8 continua exigindo uma alta capacidade computacional para o banco de dados, além de não apresentar uma boa generalização. E 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 das imagens de teste com o banco de dados, pois as redes são altamente dependentes dos valores dos níveis de cinza dos pixels, portanto muito sensíveis à rotação, translação e escala. Tendo em mente superar esses desafios, Lawrence e outros [23] apresentaram um sistema, que combina uma amostragem local da imagem, uma rede auto-organizável ou *Self-Organizing Map* (SOM), e uma rede neural de convolução ou *Convolution Neural Network*.

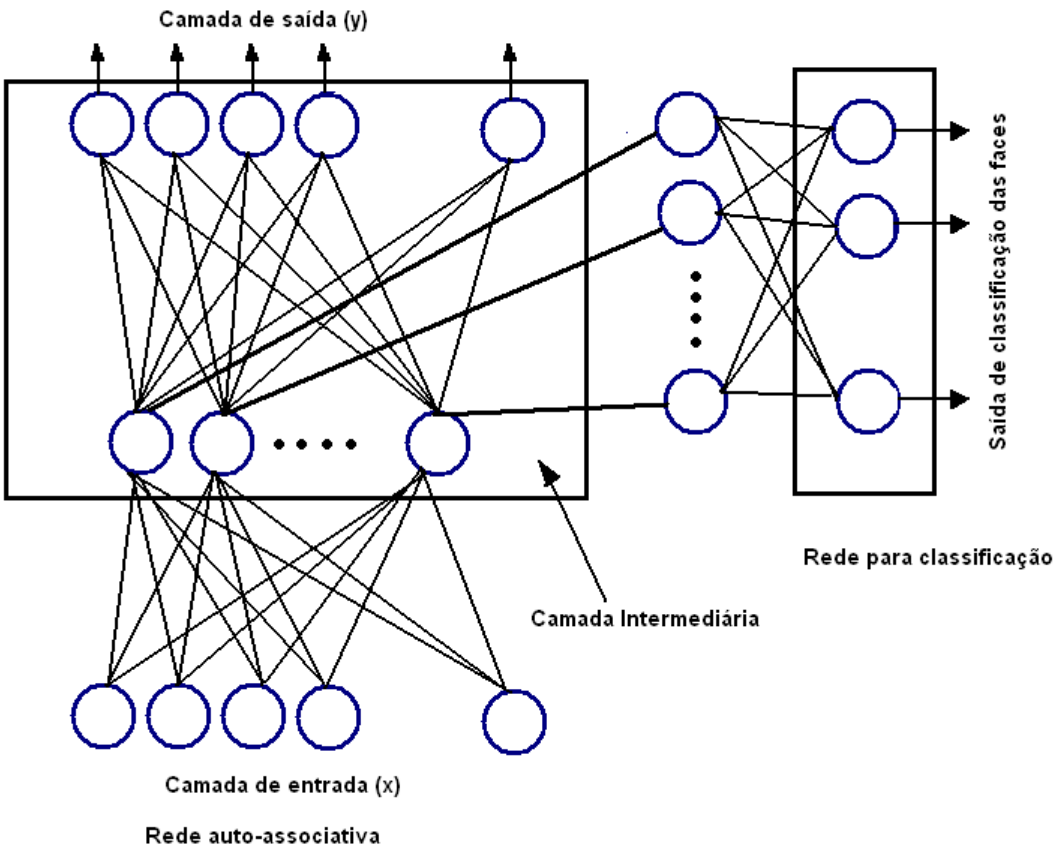


Figura 2.8 – Rede de auto-associação e classificação para reconhecimento de faces [21].

A SOM oferece uma quantização das amostras das imagens 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 mudanças nas amostras, e *a Convolution Neural Network* providencia uma invariância à translação, à rotação, à escala, e à deformação.

As imagens de treinamento (96 x 112 pixels) são amostradas com uma janela de 5 x 5 pixels deslocada sobre a imagem com passos de 4 pixels, como mostrado na Figura 2.9 [21], e as sub-amostras das imagens alimentam uma rede SOM de três dimensões com cinco nós por dimensão, perfazendo um total de 125 nós.

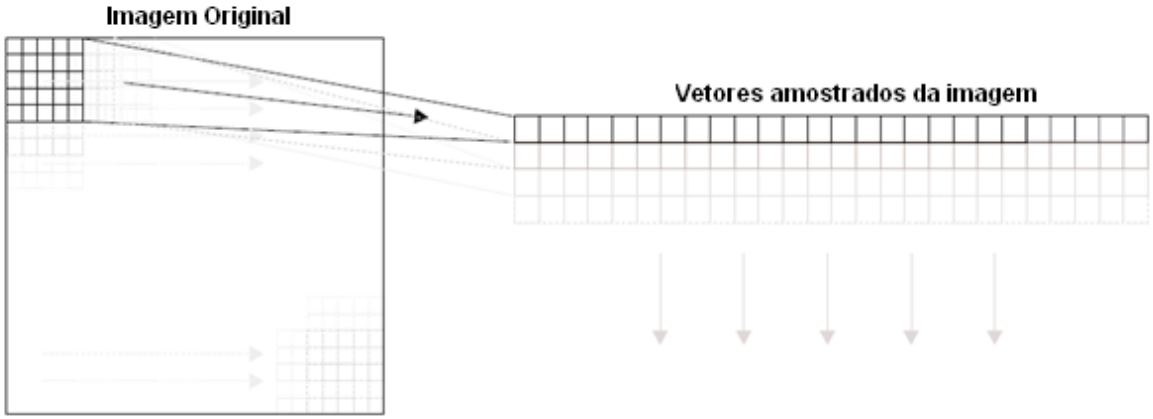


Figura 2.9 – Descrição do processo de amostragem local da imagem [21].

Agora as imagens de treinamento são novamente sub-amostradas e apresentadas à rede SOM para identificar o neurônio vencedor em cada local da 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 23 x 28 pixels (quantidade de locais sub-amostrados na imagem devido ao janelamento 5 x 5 pixels com passos de 4 pixels). Esses 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.

Contrell e Fleming [21] mostraram que a rede apresentou erro de 3,83% para o banco de dados da *Olivetti Research Laboratory* (ORL)[1], com cinco faces de treinamento e cinco faces de teste para cada pessoa. A Figura 2.10 mostra a estrutura da rede e do sistema utilizado por Contrell e Fleming [21].

Este banco de dados possui 10 imagens diferentes de 40 pessoas.[1]

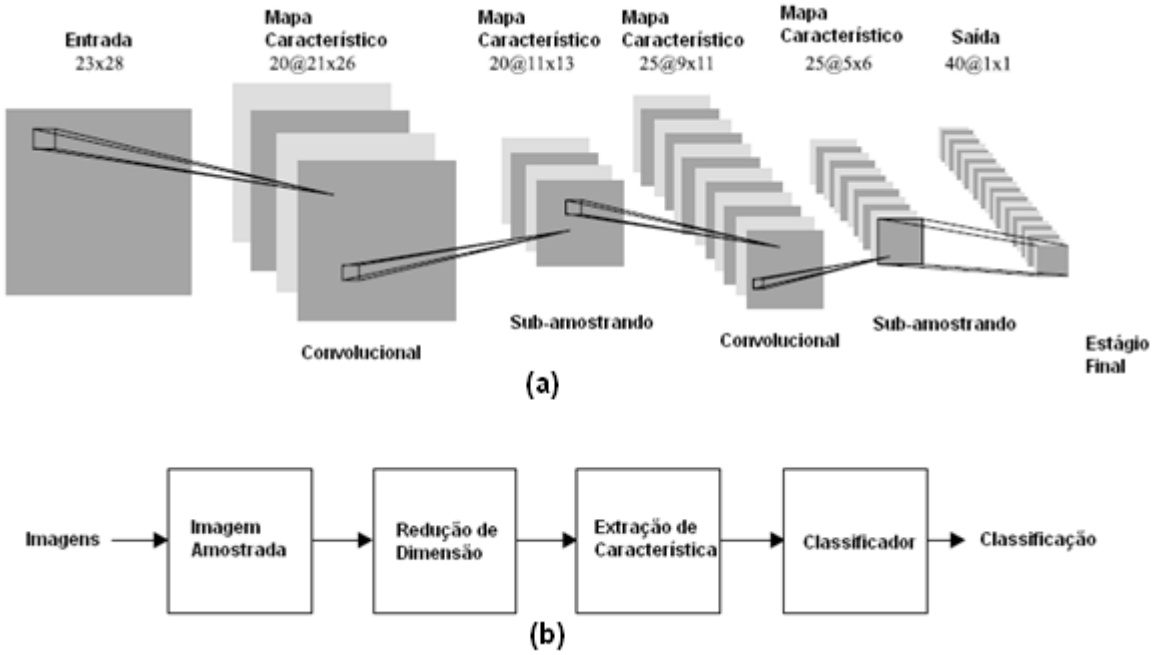


Figura 2.10 – Diagrama geral do sistema de reconhecimento de faces. (a) Típica rede de convolução; (b) diagrama em blocos do sistema utilizado para o reconhecimento de faces.

No trabalho de Sato e outros [24], 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 próxima a vista frontal.

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 a descrição de vários métodos são na maioria das vezes uma tarefa muito complexa, pois os resultados são obtidos utilizando-se diferentes bancos de dados.

2.8.1 Transformada de Fourier

A transformada de Fourier é um método muito conhecido nas aplicações de engenharia. Ela consiste em analisar o espectro de freqüência de um sinal. Com o objetivo de analisar os espectros de freqüências, que constituem uma imagem de face, Spies e Ricketts [25] desenvolveram um sistema de reconhecimento de faces baseado nessa transformada para comprimir e identificar uma face em um banco de dados visando minimizar os efeitos da translação e da rotação presentes no reconhecimento.

O reconhecimento de face é obtido encontrando no banco de dados o vetor característico, que contém os coeficientes de Fourier nas freqüências 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 freqüência, pois a transformada de Fourier plota as energias das freqüências presentes nas imagens, e essas freqüências estão relacionadas com a distribuição espacial dos níveis de cinza na imagem. No reconhecimento de face isso é realizado pelas distâncias entre as áreas de brilho particular da imagem, como o tamanho global da cabeça, ou a distância dos olhos e a variação dessas intensidades na escala de cinza.

Freqüências mais altas descrevem os detalhes finos da imagem, e os pesquisadores relatam que esses são menos úteis para a identificação de uma pessoa. Eles também sugerem que, os seres humanos podem reconhecer uma face com um breve olhar sem focalizar em detalhes pequenos.

No experimento prático Spies e Ricketts [25] utilizaram o banco de dados da ORL com cinco faces de treinamento e cinco de teste para cada pessoa, e foram escolhidas 27 freqüências do espectro de Fourier para formar o vetor característico e eles obtiveram um resultado de 95% a 98% no reconhecimento de face. Além disso, esse pequeno vetor de característica combinado com a Transformada Rápida de Fourier (FFT) possibilita esse sistema uma melhora em termos de tempo de processamento.

2.8.2 Transformada discreta do cosseno

Podilchuk e Zang [26] propuseram um método, que define um vetor característico para as imagens da face utilizando a transformada discreta de cosseno (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 especificas. Um vetor característico é montado para cada pessoa do grupo de treinamento.

Podilchuk e Zang [26] relataram que os coeficientes da DCT reduziram 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 as imagens digitais) comprimido baseado na DCT sem ter que decodificar a imagem. Esse método tem grande aplicabilidade, 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 polícia.

Liposcak e Loncaric [27] trabalharam com imagens de perfil em vez de imagens frontais. O método é 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 utilizando-se as operações de dilatação e erosão.

Liposcak e Loncaric [27] simularam o perfil de cada individuo com cabelo comprido e curto produzindo duas novas imagens para cada um. Portanto, essas imagens de perfil de cada pessoa formam um vetor característico de treinamento, que é utilizado para efetuar o reconhecimento de uma imagem de teste pela medida de distância Euclidiana [27].

2.9 Considerações Finais deste Capítulo

Neste 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 na arquitetura dos algoritmos.

A literatura cientifica apresenta 3 abordagens principais para o reconhecimento de faces: métodos holísticos, característicos e híbridos. Nesse capitulo foram discutidos superficialmente os métodos característicos e híbridos, mas foi amplamente discutido o método holístico enfatizando a técnica PCA que é a base de parte deste trabalho. Neste capítulo também foi apresentado um histórico dos principais trabalhos realizados nos últimos anos aplicando as técnicas holísticas, e alguns trabalhos significantes utilizando redes neurais.

O próximo capitulo apresenta com detalhes os fundamentos sobre as redes neurais enfatizando principalmente o *Self-Organizing Map* (SOM), que é o tipo de rede utilizada neste trabalho.

Capítulo 3

MAPAS AUTO-ORGRANIZAVEIS DE KOHONEN (SOM)

3.1 Introdução

Este capítulo apresenta os mapas auto-organizáveis (*Self-Organizing Maps – SOM*) que é o tipo básico de rede neural artificial utilizada neste trabalho.

O SOM foi desenvolvidos pelo Prof. Teuvo Kohonen [7,28] e faz parte de um dos principais paradigmas em redes neurais artificiais. O SOM foi motivado pelo modo em que as informações sensoriais são mapeadas no córtex cerebral.

O SOM é um algoritmo não supervisionado que aproxima a densidade de probabilidade dos estímulos de entrada ao mesmo tempo em que reduz a dimensionalidade, tentando manter ao máximo as relações topológicas entre os dados.

O objetivo deste capítulo é fornecer fundamentos sobre redes neurais artificiais enfatizando a aplicação do SOM em problemas de classificação automática de faces. O SOM é uma ferramenta básica em que os algoritmos desenvolvidos nesta dissertação são aplicados. É apresentada a arquitetura e o processo de treinamento do SOM, analisa-se o funcionamento e são discutidos os aspectos de convergência do SOM. Finalmente, são realizadas considerações finais deste capítulo.

3.2 Redes Neurais Artificiais (RNAs)

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 cerebral no reconhecimento e no 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, tem respostas lentas, e muitas vezes específicas na aplicação, portanto não superam as qualidades cerebrais no processamento genérico dos estímulos.

A visão humana 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 em volta do ser humano e, mais importante que isso, fornecer a informação que é necessária para interagir com ele.

O cérebro realiza rotineiramente tarefas de reconhecimento perceptivo em aproximadamente 200 *ms*, no entanto esta tarefa é complexa para ser executada em um computador convencional, mas o computador ganha na quantidade de informação armazenada [29].

Segundo Haykin [30], uma rede neural é um processador distribuído paralelamente constituído de unidades de processamento simples, que tem 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 por um processo de aprendizagem; e
- Forças de conexão entre os neurônios, conhecidas como pesos sinápticos, são utilizadas para armazenar o conhecimento adquirido.

Assim, uma rede neural baseada nos sistemas neurais biológicos pode ser definida como a generalização 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 ligações transmitem as informações (sinais) entre os neurônios;
- Cada conexão possui um peso a ela associado; e
- 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.

Na Figura 3.1 é possível visualizar a semelhança entre um neurônio celular e um neurônio de uma rede neural existente em muitas arquiteturas de redes neurais. Uma rede neural é caracterizada por seu padrão de conexão entre os neurônios (arquitetura ou topologia

de rede), seu método para determinar os pesos das conexões (algoritmo de treinamento ou aprendizagem) e pela função de ativação de seus neurônios.

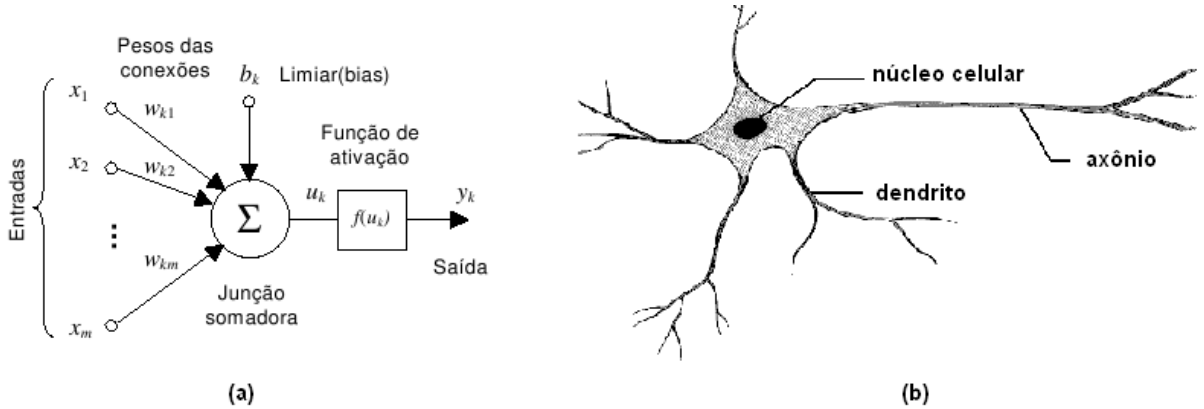


Figura 3.1 – (a) Modelo matemático não-linear de um neurônio; (b) neurônio do sistema nervoso central [30].

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. As redes neurais e suas características únicas são um ótimo auxílio para a implementação de sistemas de reconhecimento de faces e visão computacional [30].

3.3 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 pela aprendizagem. Portanto, o problema do aprendizado em redes neurais é simplesmente o problema de encontrar por um processo interativo (relativo a uma aplicação na qual cada entrada provoca uma resposta), e

conectivo (reiterado) um conjunto de parâmetros livres que possibilite a rede o desempenho desejado.

Segundo Haykin [30] a aprendizagem é um processo pelo qual os parâmetros livres de uma rede neural são adaptados por um processo de estimulação pelo ambiente no qual a rede está inserida. O tipo de aprendizagem é determinado pela maneira pela qual a modificação dos parâmetros ocorre. Esta definição implica nos seguintes eventos:

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

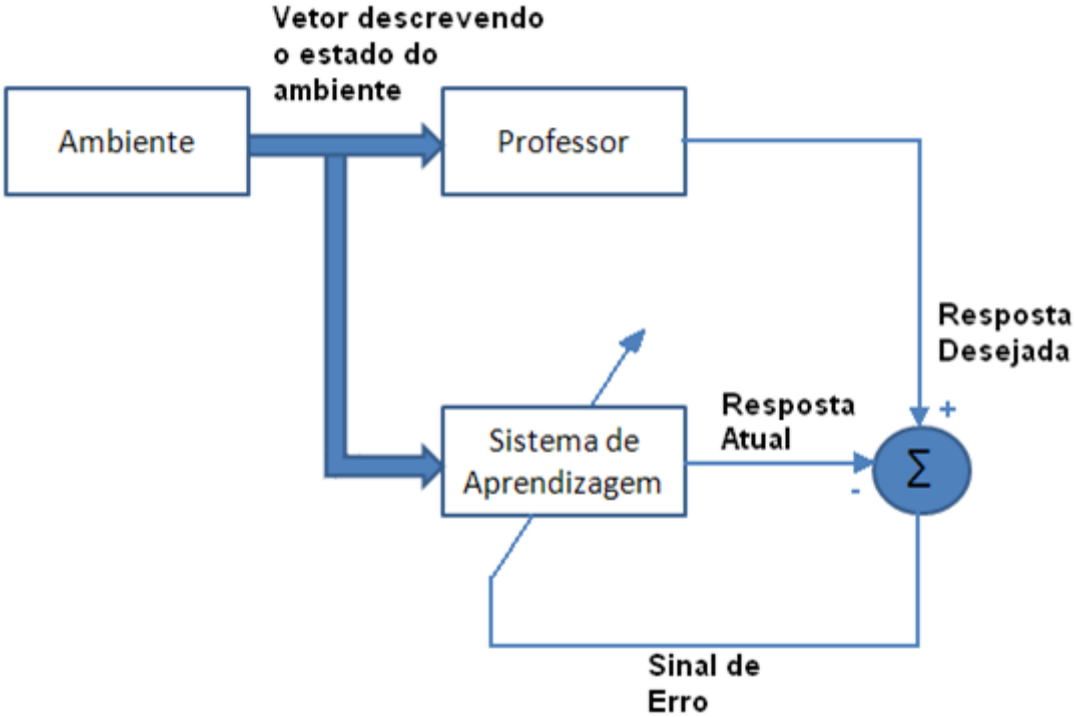
3.3.1 Aprendizagem supervisionada

Em uma aprendizagem supervisionada, a rede neural é treinada com o auxilio 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 padrão de entrada.

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.

A 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 3.2 apresenta o diagrama de blocos representativo desse sistema.



Fgiura 3.2 – Diagrama de blocos representativo da aprendizagem supervisionada [30].

3.3.2 Aprendizagem por reforço

Na aprendizagem por reforço não é utilizado um supervisor, mas sim um crítico (ou juiz). Sendo assim, o aprendizado de um mapeamento entrada/saída é desempenhado pela interação contínua com o ambiente buscando minimizar um índice escalar de desempenho.

O crítico (ou juiz) possui apenas indicações imprecisas (por exemplo: sucesso/insucesso da rede) sobre o comportamento final desejado. Para ser mais exato, nesse 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).

A Figura 3.3 mostra o diagrama de 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.

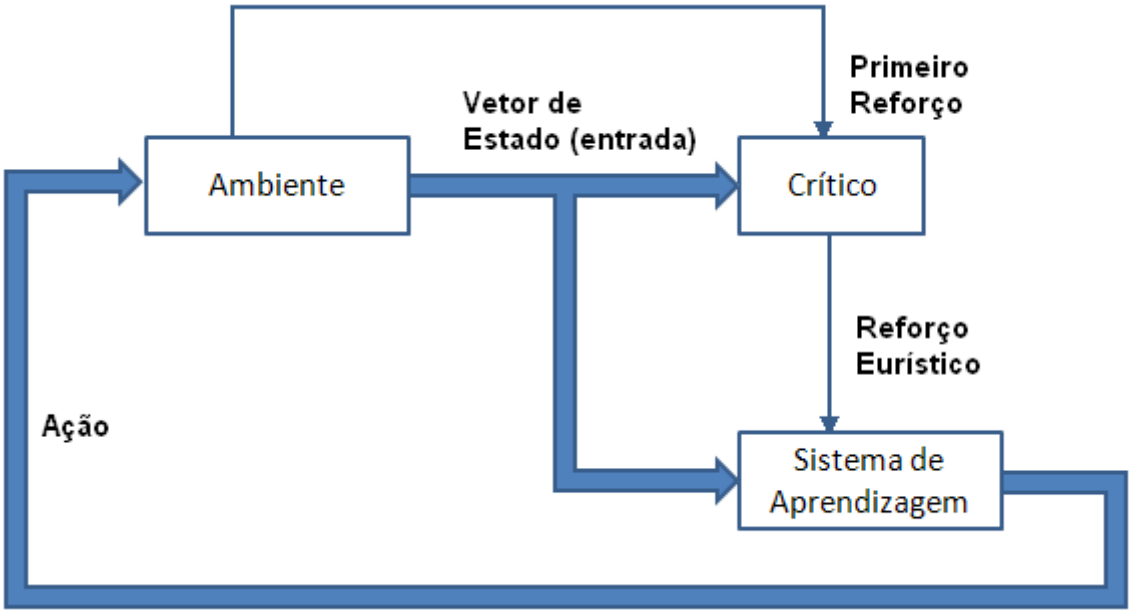


Figura 3.3 – Diagrama de blocos da aprendizagem por reforço [30].

3.3.3 Aprendizagem não-supervisionada

Em algoritmos não-supervisionados ou auto-organizados não existe 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 utilizar supervisores para o seu treinamento.

A rede trabalha os valores de entrada e se organiza de modo que acabe classificando-os, utilizando para isso os seus próprios critérios. Esse tipo de rede utiliza os neurônios como classificadores, e os dados de entrada como os elementos para a 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 esse tipo de rede como detector de características, devido a 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 inicia quando um exemplo é mostrado à rede e termina quando todas as conexões estão devidamente 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 de processamento será utilizado para ajustar as conexões e melhor será o seu treinamento. A Figura 3.4 apresenta um diagrama de blocos desse sistema.

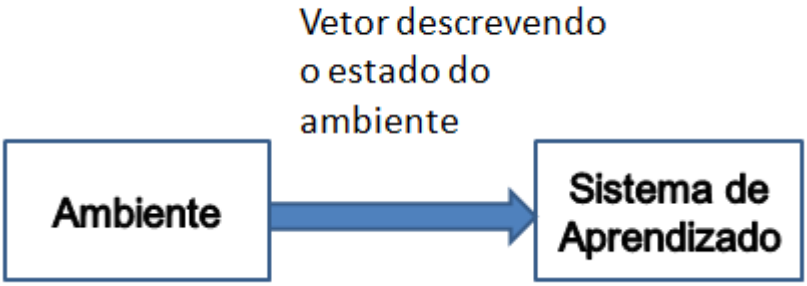


Figura 3.4 – Diagrama de blocos da aprendizagem não-supervisionada [30].

3.3.4 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 por ela no teste. Generalização é a capacidade de um modelo de aprendizagem responder corretamente aos exemplos que lhe são apresentados, onde esses 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 [30].

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, a translação, a escala, a iluminação e as mudanças na integridade física do rosto (envelhecimento, mudança no corte de cabelo, uso de óculos, etc).

A Figura 3.5 apresenta um gráfico típico da evolução do erro durante a aprendizagem de uma rede neural, comparando a curva do erro referente a base de aprendizado (amostras de treinamento) com a curva do erro da base de teste de generalização.

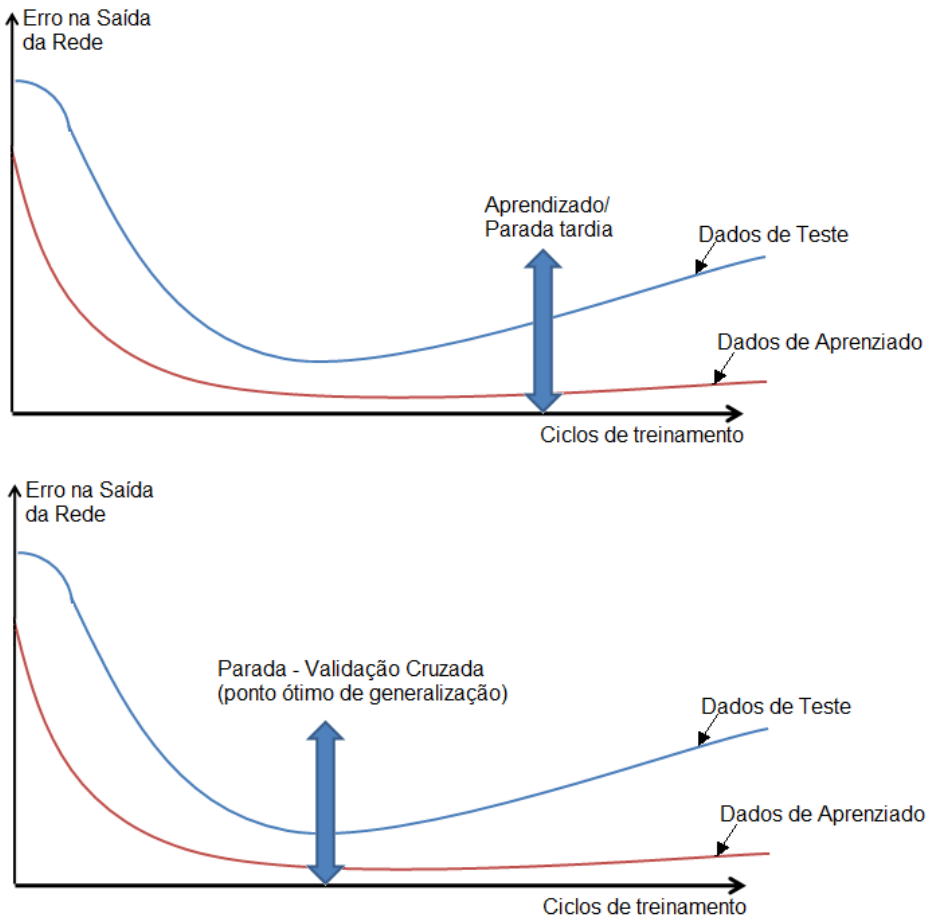


Figura 3.5 – Aprendizagem: erro em relação ao conjunto de aprendizado e de teste.

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-training* pode ser evitado utilizando-se um teste de generalização por validação cruzada (*cross-validation*).

3.3.5 Aprendizado competitivo

O objetivo do aprendizado competitivo é fazer com que os neurônios se especializem nos estímulos apresentados de forma não supervisionada. Isto é, nenhuma informação sobre a classe do estímulo apresentado é utilizada no processo de ajuste dos pesos sinápticos.

Todos os neurônios recebem o mesmo conjunto de entradas e competem, por uma dinâmica que usa conexões laterais, com todos os outros neurônios, conforme pode ser visto na Figura 3.6. Essas conexões laterais podem ser positivas (no caso da auto-realimentação) ou negativas (inibitórias).

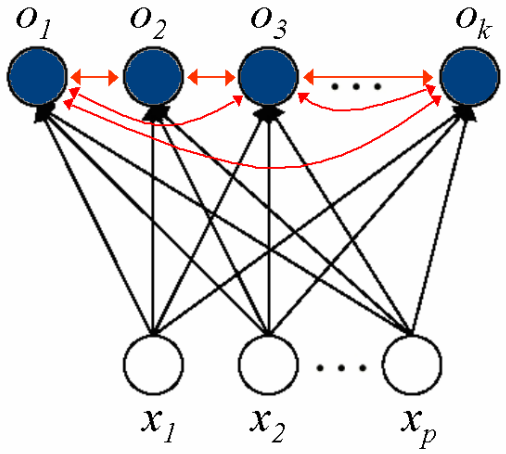


Figura 3.6 – Exemplo simplificado de uma rede neural competitiva.

Existem três elementos básicos em regras de aprendizado competitivo [30]:

1. O conjunto de neurônios com a mesma função de ativação, apenas diferindo inicialmente pela aleatoriedade da distribuição dos pesos sinápticos, o que os farão responder diferentemente para um determinado conjunto de padrões;

2. Os vetores de pesos que conectam neurônios entre as camadas de entrada e de saída são limitados, por exemplo, || m || = 1; e

3. Um mecanismo que permita aos neurônios competir pelo direito de responder a um determinado subconjunto de entradas. No final do processo de competição, apenas um neurônio estará ativo. Esta é a regra *winner-takes-all*.

Tais condições citadas podem em muitos casos serem estendidas ou mesmo desconsideradas. Existem regras em que se pode considerar mais de um neurônio vencedor para um determinado padrão, ou mesmo não necessitar normalizar os pesos como descrito pela condição 2.

No aprendizado competitivo, o ajuste é restrito ao neurônio vencedor a cada apresentação de um padrão, onde esse neurônio é o neurônio que possui um conjunto de pesos mais semelhantes ao padrão apresentado. O neurônio vencedor é identificado utilizando-se um critério de distância ou dissimilaridade, entre o padrão $\boldsymbol{x}$ e o neurônio $\boldsymbol{m_i}$, como mostrado na Equação (3.1).

$$d(x, m_c) = \min_i \{d(x, m_i)\} \tag{3.1}$$

A adaptação ocorre apenas aos pesos sinápticos do neurônio vencedor *c*, como apresentado na Equação (3.2).

$$\Delta m_c(t+1) = \alpha(t).[x(t) - m_c(t)] \tag{3.2}$$

Para os demais neurônios, $i = 1, \ldots, k, i \neq c$, $\Delta m_i(t+1)=0$. A taxa de aprendizado, $\alpha(t)$ pode ser uma função (ou seqüência de valores) decrescente, monotônica, limitada em $0 < \alpha(t) < 1$. O efeito desta regra é o deslocamento dos pesos do neurônio vencedor na direção do padrão *x* como mostrado na Figura 3.7.

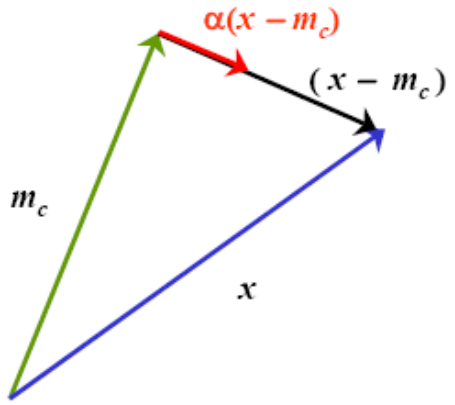


Figura 3.7 – Atualização de pesos em uma rede neural competitiva.

3.3.6 Estrutura básica do SOM

O SOM é visto atualmente como uma ferramenta de software para a visualização de dados. Geralmente os dados de aplicações reais são multi-dimensionais o que torna difícil a visualização das relações entre eles.

O SOM implementa uma projeção não-linear de um espaço de dados sensoriais ou de atributos $\Re^p$, geralmente de dimensionalidade elevada (p >> 2), em um conjunto discreto de neurônios, geralmente dispostos na configuração de um vetor ou uma matriz. Relações estatísticas complexas e não-lineares entre os dados são convertidas em relações geométricas simples sobre um "*display*" de menor dimensionalidade.

Apesar de ter sido relacionado a componentes principais (CP), existem diferenças fundamentais entre esses dois métodos, como por exemplo o espaço discreto de saída do SOM, enquanto que em CP o espaço é contínuo. Além disso, as componentes principais são obtidas por combinações lineares das variáveis. Enquanto que o SOM é obtido matematicamente por regressão recursiva e não paramétrica [31], e ao levar em conta a topologia da rede resulta em uma projeção não linear da densidade de probabilidade p(x) dos sinais de entrada sobre a grade de neurônios.

Os dois objetivos principais do SOM são reduzir a dimensionalidade dos dados ao mesmo tempo em que se tenta preservar as relações métricas e topológicas do espaço de entrada. A dimensionalidade do mapa é obtida pelas relações de vizinhança entre os neurônios.

Um mapa auto-organizável consiste de duas camadas de neurônios: a camada de entrada, *I*, e a camada de saída (ou de Kohonen), *U*. As entradas da rede são vetores no espaço *p*-dimensional, geralmente no espaço $\Re^p$. Cada neurônio *i* da camada de Kohonen possui um vetor também no espaço $\Re^p$ associado, $m_i = [m_{i1}, m_{i2}, ..., m_{ip}]^T$.

Os neurônios na camada de Kohonen são conectados aos neurônios adjacentes por uma relação de vizinhança que descreve a estrutura do mapa. No caso bidimensional pode-se ter vizinhança tipo 4-conectados (mapas retangulares) ou 6-conectados (mapas hexagonais). A escolha da topologia, assim como o tamanho do mapa, *N x M*, dependem da aplicação. Geralmente são utilizados SOMs retangulares por razões de simplicidade. A Figura 3.8 ilustra um SOM para um vetor de entrada de dimensão 3, e a camada *U* possui tamanho 7 *x* 10.

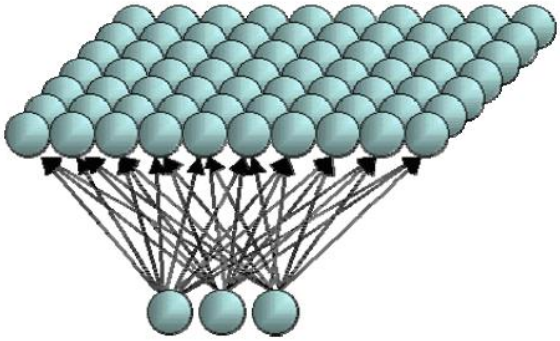


Figura 3.8 – Exemplo de SOM de dimensão retangular [7].

Durante a fase de treinamento no SOM os neurônios que terão os pesos sinápticos atualizados, serão aqueles que tiverem dentro de um raio (decrescente com o tempo) do neurônio vencedor, como mostra a Figura 3.9.

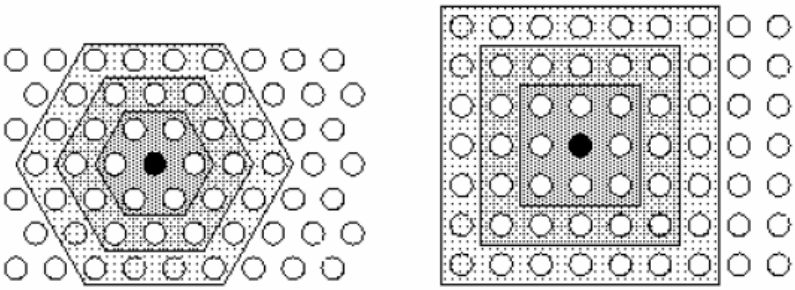


Figura 3.9 – Duas configurações de *grid* e seus níveis de vizinhança [7].

Outra distância importante refere-se à distância no espaço de pesos, ou seja, a distância entre um padrão x e um neurônio m_i. Este tipo de distância depende da métrica a ser utilizada no problema. Em geral utiliza-se a distância Euclidiana ou a de Hamming.

Na Figura 3.10 é possível visualizar uma rede SOM, onde é mostrado que cada neurônio possui um "endereço" no *grid* e todos recebem o mesmo sinal vindo da entrada. Pode-se também observar que cada neurônio possui um vetor com a mesma dimensão da entrada.

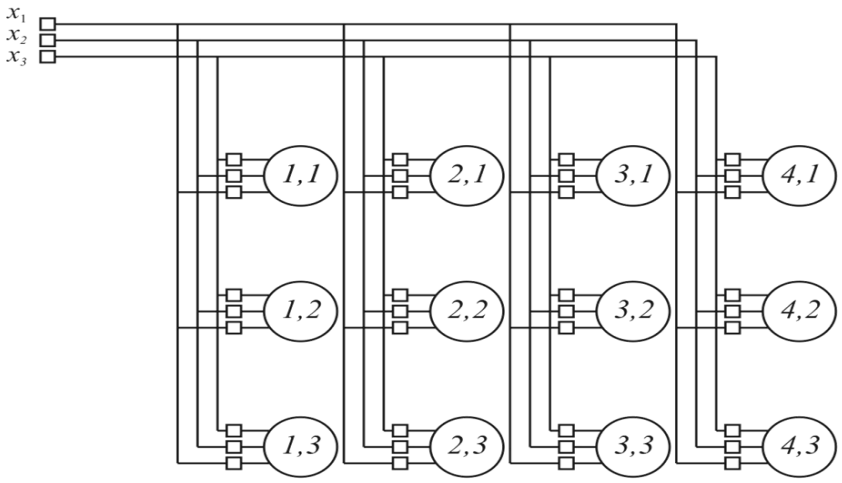


Figura 3.10 – SOM com tamanho 4 x 4 e entrada com dimensão de 3.

3.4 Treinamento em redes SOM

3.4.1 Algoritmo convencional

Uma das razões do sucesso do SOM é a simplicidade do algoritmo de treinamento. Geralmente, o processo de treinamento do SOM consiste em apenas quatro etapas, além da inicialização:

1. Inicialização da rede. Pode-se utilizar as estratégias descritas na sub-seção 3.5.3 deste capítulo para gerar os vetores de pesos iniciais m_i, $i = 1, \ldots, n$, onde n é o número de neurônios.

2. Um padrão de entrada $x_k = (\xi_1, \xi_2, \ldots, \xi_p)$, $x_k \in \Re^p$, é selecionado aleatoriamente de todo conjunto de padrões.

3. Uma função de ativação é utilizada para calcular o estado de cada neurônio *i* em relação ao padrão x_k. No caso da distância Euclidiana, utiliza-se a Equação (3.3).

$$d(m_i, x_k) = \sqrt{\sum_{j=1}^{p} \left[x_{kj}(t) - m_{ij}(t) \right]^2} \tag{3.3}$$

4. O neurônio vencedor, *c*, é escolhido de acordo com a Equação (3.4).

$$\| x - m_c \| = \min_i \{ \| x - m_i \| \} \tag{3.4}$$

onde ||.|| significa a distância utilizada, neste trabalho utiliza-se a distância Euclidiana.

5. Os pesos sinápticos do neurônio vencedor, *c* , como também os pesos dos neurônios que estão dentro da vizinhança de *c*, são atualizados utilizando-se as Equações (3.5) e (3.6).

$$m_i(t+1) = m_i(t) + h_{ci}(t) \cdot [x(t) - m_i(t)] \tag{3.5}$$

Onde $h_{ci}(t)$ é o núcleo de vizinhança ao redor do neurônio vencedor *c* no passo *t*. Este termo é uma função decrescente com o tempo e com a distância do neurônio *i* ao neurônio vencedor *c*, e geralmente é formado por dois componentes: a taxa de aprendizado α(t) e a função de vizinhança *h*(*d*,*t*).

$$h_{ci}(t) = \alpha(t) \cdot h(\| r_c - r_i \|, t) \tag{3.6}$$

Onde r_i é a posição do neurônio *i* na camada de Kohonen.

O processo definido pelos passos 2 a 5 é repetido iterativamente, o que leva a um mapeamento gradual de preservação da topologia dos sinais de entrada à medida que o algoritmo de treinamento converge.

No final do treinamento é esperado que o mapa esteja topologicamente ordenado. Basicamente isto significa que n_i padrões que estejam próximos no espaço p-dimensional de atributos devem ser mapeados em neurônios que estejam próximos no espaço do *grid*, geralmente no mesmo neurônio ou em neurônios vizinhos.

3.4.2 Observações do processo de treinamento do SOM

Um dos principais problemas no treinamento do SOM é a ausência de um funcional ou de um critério de otimização. A falta das informações referentes à saída desejada para cada padrão de entrada no treinamento faz com que a alternativa de adaptação dos pesos seja realizada de forma não supervisionada, ou seja, por um processo competitivo.

O objetivo é "sintonizar" os neurônios às entradas fazendo com que um neurônio vencedor (BMU[2]) de um determinado padrão de entrada responda mais fortemente àquele padrão no próximo ciclo. O grau de adaptação, em uma iteração t, depende da taxa de aprendizado $\alpha(t)$, da função de vizinhança $h(d, t)$ e da distância entre o neurônio i, representado pelo seu vetor sináptico m_i, e o vetor x(t).

A etapa que mais consome tempo no treinamento é a busca pelo neurônio vencedor. Geralmente utiliza-se a busca seqüencial, o que torna o processo extremamente árduo quando o número de neurônios é elevado. Variações do SOM têm sido propostas como a

[2] Do inglês *best match unit.*

implementação de uma estrutura em árvore para tentar diminuir esse tempo [32]. Uma alternativa é a paralelização da busca, utilizando-se processadores especiais e/ou paralelos.

No decorrer do treinamento os padrões são apresentados à rede, a taxa de aprendizado $\alpha(t)$ decresce gradualmente a um valor pré-especificado, geralmente próximo a zero, de acordo com uma determinada função de decaimento. Isto garante o término do processo de aprendizado em um tempo finito. Um exemplo para $\alpha(t)$ pode ser mostrado na Equação (3.7).

$$\alpha(t) = \alpha(0) \cdot e^{-\left(\frac{t}{\lambda}\right)} \tag{3.7}$$

onde:

$\alpha(0) < 1$, é a taxa de aprendizado inicial;

λ - parâmetro responsável pela taxa de redução desejada.

A vizinhança ao redor do neurônio vencedor, $h_{ci}(t)$, Equação (3.6), decresce com o tempo a uma taxa previamente especificada similar a taxa de aprendizagem. Essa vizinhança é considerada o principal agente da auto-organização no SOM [33,34] responsável pela preservação das relações topológicas.

Pode-se definir uma região de ativação circular, ou retangular, como mostrado na Figura 3.9, como um conjunto de neurônios $N_c(t)$ em uma região do mapa centrado no vencedor, c. No início do treinamento, $N_c(t)$ compreende grande parte do SOM e à medida que o aprendizado prossegue $N_c(t)$ encolhe até que apenas o neurônio vencedor seja atualizado. Geralmente atualizam-se apenas os pesos dos neurônios dentro da área excitatória, ou seja, $h_{ci}(t) = 0$ para os neurônios $m_i \notin N_c(t)$ (ver as Equações (3.5) e (3.6)).

Uma vez encontrado o BMU, os pesos sinápticos do BMU e dos neurônios pertencentes à vizinhança são deslocados na direção do padrão x de acordo com a Equação (3.5), conforme pode ser visto na Figura 3.11. A magnitude do deslocamento é influenciada pela taxa de aprendizado $\alpha(t)$ e pela função de vizinhança $h(d, t)$.

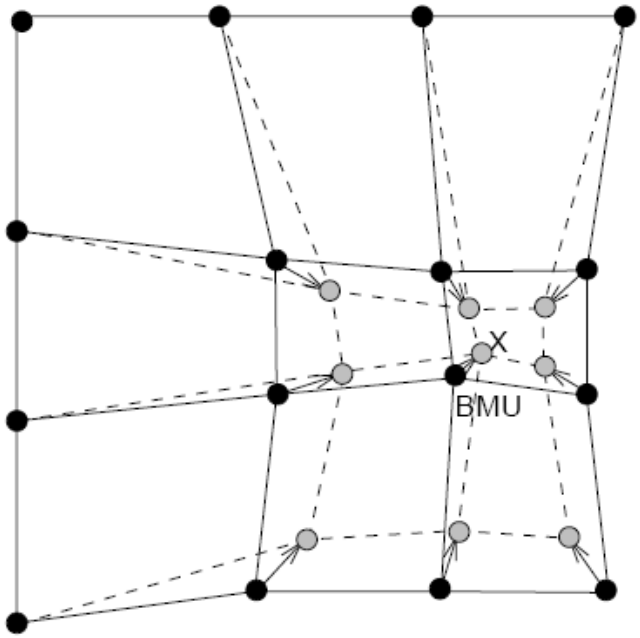


Figura 3.11 – Atualização do BMU e de seus vizinhos em direção ao padrão de entrada x.

O BMU possui o vetor de pesos sinápticos com maior similaridade ao estímulo apresentado, para verificar tal similaridade geralmente utiliza-se a distância Euclidiana. Apesar de não ser obrigatório, grande parte das aplicações na literatura normalizam os dados antes de apresentarem ao SOM, limitando o espaço de entrada a uma hiperesfera p-dimensional de raio 1. A norma Euclidiana de um vetor $x = (x_1, x_2, \ldots, x_p)^T$, é obtida pela Equação (3.8).

$$norma = \sqrt{\sum_{k=1}^{p} x_k^2} \tag{3.8}$$

e os padrões de entrada podem ser normalizados utilizando-se a Equação (3.9).

$$x_k^{'} = x_k \,/\, norma \tag{3.9}$$

A função de ativação baseada no produto interno é mostrada na Equação (3.10). Os padrões de entrada podem ser normalizados antes da apresentação ao SOM.

$$f_i(t) = x^T(t) \cdot m_i(t) = \sum_{k=1}^{p} x_k(t) \cdot m_{ik}(t) \tag{3.10}$$

Na Equação (3.4) o BMU foi descrito como o neurônio, *c*, com menor distância. Como a Equação (3.10) define uma função de similaridade entre dois vetores, o BMU será definido como o neurônio mais similar ao estímulo *x*(*t*), o que levaria a Equação (3.4) a ser descrita pela Equação (3.11).

$$\| x^T \cdot m_c \| = \max_i \left\{ \| x^T \cdot m_i \| \right\} \tag{3.11}$$

A Equação (3.4) é rescrita na forma da Equação (3.11) sempre que os pesos sinápticos se mantiverem normais. Isso pode ser realizado re-normalizando os pesos de todos os neurônios que devem ser atualizados para um determinado instante de tempo *t*.

3.5 Parâmetros de Treinamento em Redes SOM

A escolha certa dos parâmetros de treinamento é tarefa difícil e de muita importância na etapa de classificação. O resultado final do treinamento geralmente pode ser diferente, mesmo quando o treinamento é realizado com os mesmos parâmetros. Isso ocorre devido basicamente a dois fatores aleatórios, como a inicialização dos pesos e a seqüência de apresentação dos padrões de treinamento.

3.5.1 Tamanho e dimensionalidade do mapa

Geralmente utiliza-se mapas de dimensão 1 ou 2, este último é mais freqüente devido à capacidade de visualização do mapeamento dos dados na forma de um display. Porém, na maioria dos problemas de agrupamentos, por exemplo, no trabalho de Bezdek e outros [35] foi utilizado um *grid* unidimensional, no qual o número de neurônios é igual ao número esperado de agrupamentos.

Em muitos casos práticos, pode ser útil dimensionar o mapa supondo que alguns neurônios serão inativos, isto é, não serão BMUs para nenhum padrão do conjunto de treinamento. O SOM funciona como uma grade elástica tentando concentrar mais neurônios em regiões mais densas do espaço de atributos.

Mapas de tamanhos grandes tornam o aprendizado muito lento, enquanto que mapas de tamanhos muito pequenos, geram *U-matrizes*[3] também pequenas, o que irá dificultar bastante a utilização de métodos de processamento e segmentação de imagens.

---

[3] Matrizes onde os elementos são distâncias entre os neurônios, calculadas no espaço de pesos.

3.5.2 Topologia da rede e conectividade entre os neurônios

Como descrito na sub-seção 3.3.6 deste capítulo, no caso bidimensional, geralmente utiliza-se um arranjo retangular no qual os elementos são conectados aos vizinhos nos padrões de conectividade 4 ou 6, implicando na topologia retangular ou hexagonal. Os elementos que situam-se nas extremidades do mapa podem, ainda, serem conectados ou não. Neste trabalho, assume-se que em todos os exemplos foram usados mapas com topologias e conectividades retangulares, a menos que se mencione o contrário.

3.5.3 Inicialização dos pesos

Apesar do SOM ter uma inicialização relativamente robusta, a escolha do método de inicialização dos pesos de forma adequada pode facilitar a convergência do algoritmo. A forma mais comum de inicialização é a randômica, onde atribuí-se valores pequenos aleatórios aos pesos. Pode-se inicializar vetores considerando-se aleatoriamente padrões no conjunto de dados, e copiando-se esses padrões aos vetores de pesos. Outra maneira seria considerar valores máximos e mínimos em cada dimensão e inicializar, de forma ordenada, a grade de neurônios, interpolando o espaço de acordo com o número de neurônios.

A inicialização linear é preferível em relação às descritas anteriormente neste capítulo porque ela possibilita ir diretamente para a fase de convergência do algoritmo de treinamento conforme descrito por Kohonen [7].

Considerando-se o caso de um SOM bidimensional (k = 2), a inicialização linear é realizada utilizando-se os k componentes principais da matriz de autocorrelação do conjunto de dados X. Os pesos, $m_i(0)$ são inicializados de forma ordenada na direção do sub-espaço linear obtido pelos autovetores, ortogonais, correspondentes aos dois maiores autovalores, sendo seu centróide coincidente com a média do conjunto de dados X. Seja y_1 e y_2 os autovetores escolhidos. A coordenada do neurônio (i, j) pode ser expressa pela seguinte expressão.

$$s \cdot \left[\left(i - \frac{\max_i}{2} \right) \cdot y_1 + \left(j - \frac{\max_j}{2} \right) \cdot y_2 \right]$$

Os valores de *max_i* e *max_j* representam o tamanho do mapa bidimensional, escolhido pelo usuário, e s é uma constante selecionada de forma adequada.

3.5.4 Função de decaimento

O algoritmo original do SOM [30] define uma vizinhança do tipo bolha ou circular, onde todos os neurônios dentro de um determinado raio possuem h igual a 1. Seja N_c(t) o conjunto de neurônios satisfazendo essa condição. Desta forma, h_{ci}(t) = α(t) para todos os neurônios $i \in N_c$(t) e h_{ci}(t) = 0, caso contrário, $i \notin$ Nc(t). A Figura 3.9 ilustra o efeito no tempo de N_c(t).

Outra forma bastante utilizada é a mostrada na Equação (3.12), onde utiliza-se a vizinhança do tipo Gaussiana,

$$h_{ci}(t) = \alpha(t) \cdot \exp\left(\frac{\| r_c - r_i \|^2}{\sigma^2(t)} \right) \tag{3.12}$$

onde:

r_c e r_i – posições dos neurônios *c* e *i* no *grid* de neurônios, onde r_c e $r_i \in \Re^2$; e

σ – raio de $N_c(t)$.

O raio inicial, N*c*(0), deve ser relativamente grande, geralmente, a metade do diâmetro da rede. O raio inicial pequeno faz com que o algoritmo não gere uma ordenação adequada [30].

3.5.5 Taxa de aprendizagem

Inicialmente deve-se escolher α(t) próximo a 1, possibilitando seu decaimento com o tempo, que pode ser realizado, por exemplo, de forma exponencial, linear, ou inversamente proporcional a *t*. Para convergência do algoritmo, é necessário que $h_{ci}(t) \rightarrow 0$ quando $t \rightarrow \infty$. Um exemplo para $\alpha(t)$ pode ser visto na Equação (3.7).

3.5.6 Número de interações

O algoritmo original proposto por Kohonen estabelece duas fases, uma de ordenação inicial, e outra de convergência. A primeira dura, por exemplo, cerca de 1000 interações, enquanto a convergência deve ser mais demorada, por exemplo, 100.000 interações. O

número de interações depende da aplicação e da quantidade de padrões a serem apresentados à rede na fase de treinamento.

3.6 Considerações Finais deste Capítulo

Este capítulo descreveu os principais conceitos de redes neurais artificiais, enfatizando o SOM.

Sobre os SOMs, foram tratados os aspectos do treinamento e seus parâmetros, foram destacados que a melhor alternativa atual no treinamento é a utilização da inicialização linear, que implica em começar o treinamento em um estágio relativamente avançado de ordenação em relação à inicialização convencional, que é a aleatória.

Em condições gerais para obter bons mapas de Kohonen é necessário treinar a rede com número de interações relativamente elevado, e na fase inicial manter tanto o raio de vizinhança do neurônio vencedor como a taxa de aprendizado relativamente elevada, permitindo decaimento suave desses fatores com o tempo.

O próximo capitulo apresenta o método proposto para o reconhecimento de faces explicando as técnicas utilizadas para a extração de características/informação das faces bem como a rede *Concurrent Self-Organizing Maps* (CSOM) utilizada para o reconhecimento/classificação das faces.

Capítulo 4

MÉTODO PROPOSTO

4.1 Introdução

Os softwares de reconhecimento de faces humanas são desenvolvidos com base em algoritmos matemáticos com o objetivo de autenticar ou identificar indivíduos, utilizando-se técnicas de extração de características na base de dados de treinamento para o reconhecimento/classificação de faces apresentadas ao sistema.

Na autenticação é necessário que o usuário digite algum código ou passe, por exemplo, um crachá ou cartão para que o sistema capture a face e compare com a face indexada pelo código no banco de faces. Esse método é conhecido como verificação um para um (*one-to-one*), pois o sistema apenas verifica se a pessoa é ela mesma. Na identificação é preciso um sistema mais sofisticado, pois é necessário fazer a comparação de 1 para "M" (*one-to-many*) indivíduos dentro do banco de faces.

Cada algoritmo biométrico possui vantagens e desvantagens, e são utilizados em aplicações específicas. Em locais de alta densidade de indivíduos, como aeroportos e shoppings, o reconhecimento por meio da face é uma ótima tecnologia para separar uma pessoa da multidão para posterior averiguação. Mas dificilmente ele será usado para autorizar saque em bancos, pois o reconhecimento por meio de face apresenta deficiência em relação às

taxas de falsa aceitação (permite uma pessoa não autorizada) e falsa rejeição (não permite alguém que esteja cadastrado).

A proposta deste trabalho está baseada no reconhecimento de faces utilizando as características holísticas da imagem, pois elas são eficazes quando aplicadas em banco de dados com grandes variações na iluminação, rotação e escala de faces na cena, considerando que esses fatores estão presentes nas aplicações do mundo real.

Os métodos holísticos em reconhecimento de faces consideram como vetor característico os pixels relacionados à face na imagem. Para evitar o problema da alta dimensionalidade, diversos algoritmos estatísticos de redução de dimensionalidade podem ser aplicados, como por exemplo: *Principal Component Analysis* (PCA), *Two-Dimensional Principal Component Analysis* (2D-PCA) e Redes Neurais. O PCA é o mais popular, por ser de fácil implementação, freqüentemente ele é utilizado na etapa de pré-processamento de imagens para melhorar o desempenho. E a classificação é realizada de várias maneiras, geralmente utilizando-se métodos estáticos ou redes neurais

Nesta dissertação não será realizado o estágio de detecção automática de faces presentes em sistemas de reconhecimento de faces automático. Toda a pesquisa será voltada para a fase de classificação/identificação.

O algoritmo proposto nesta dissertação para a classificação de faces foi desenvolvido em três etapas. Na primeira etapa são obtidas as matrizes características das faces utilizando a *Two-Dimensional Discrete Cosine Transform* (2D-DCT) e a *Two-Dimensional Principal Component Analysis* (2D-PCA). O treinamento da *Concurrent Self-Organizing Map* (CSOM) é realizado na segunda etapa usando as matrizes características das faces. E finalmente, na terceira etapa obtém-se a matriz característica da imagem consulta, classificando-a utilizando a rede CSOM da segunda etapa.

Este capítulo apresenta as técnicas 2D-DCT e 2D-PCA usadas neste trabalho para a obtenção das matrizes características e a rede neural CSOM utilizada na classificação das faces. Em seguida é mostrado como são usadas a 2D-DCT, 2D-PCA e a rede CSOM no algoritmo proposto nesta dissertação para o reconhecimento de faces. Finalmente, são realizadas as considerações finais deste capítulo.

4.2 *Two-Dimensional Discrete Cosine Transform* (2D-DCT)

A transformada discreta de cosseno (DCT, do inglês, *Discrete Cosine Transform*) vem sendo utilizada pela comunidade de processamento de imagens, principalmente nas áreas de compressão, filtragem e extração de características [36].

A Equação (4.1) mostra a DCT aplicada em uma seqüência $f(x)$ unidimensional de tamanho N.

$$C(u) = \alpha(u)\sum_{x=0}^{N-1} f(x)\cos\left[\frac{\pi(2x+1)u}{2N}\right] \qquad u = 0, 1, 2, 3, ..., N\text{-}1 \qquad (4.1)$$

De forma análoga, a Equação (4.2) mostra a transformada inversa da DCT de uma seqüência unidimensional.

$$f(x) = \sum_{u=0}^{N-1} \alpha(u)C(u)\cos\left[\frac{\pi(2x+1)u}{2N}\right] \qquad x = 0, 1, 2, 3, ..., N\text{-}1 \qquad (4.2)$$

A Equação (4.3) mostra o valor de *α(u)* utilizado nas Equações (4.1) e (4.2).

$$\alpha(u) = \begin{cases} \sqrt{\dfrac{1}{N}} & para \quad u = 0 \\ \sqrt{\dfrac{2}{N}} & para \quad u \neq 0 \end{cases} \tag{4.3}$$

O primeiro coeficiente da DCT, isto é *C(0)*, é conhecido na literatura como coeficiente DC por ser o valor médio da seqüência de amostras, os demais coeficientes são denominados de coeficientes AC [36].

A transformada discreta de cosseno bidimensional (2D-DCT, do inglês, *Two-Dimensional Discrete Cosine Transform*) é utilizada para transformar uma representação da informação no domínio do espaço em uma representação no domínio da freqüência. A Equação (4.4) mostra o cálculo da 2D-DCT.

$$C(u,v) = \alpha(u)\alpha(v)\sum_{x=0}^{N-1}\sum_{y=0}^{N-1} f(x,y)\cos\left[\frac{\pi(2x+1)u}{2N}\right]\cos\left[\frac{\pi(2y+1)v}{2N}\right] \tag{4.4}$$

para *u, v = 0, 1, 2, ..., N-1* e os valores de *α(u)* e *α(v)* utilizados são mostrados na Equação (4.3).

A Equação (4.5) ilustra o cálculo da inversa da transformada discreta de cosseno bidimensional.

$$f(x,y) = \sum_{u=0}^{N-1}\sum_{v=0}^{N-1} \alpha(u)\alpha(v)C(u,v)\cos\left[\frac{\pi(2x+1)u}{2N}\right]\cos\left[\frac{\pi(2y+1)v}{2N}\right] \tag{4.5}$$

para x, *y = 0, 1, 2, ..., N-1*.

Como a DCT é uma transformada separável, a 2D-DCT pode ser encontrada calculando-se a DCT unidimensional das linhas, e o resultado desse cálculo é utilizado para obter a DCT das colunas. O resultado desse último cálculo são os coeficientes da 2D-DCT de uma imagem qualquer. A Figura 4.1 mostra o diagrama do método linha–coluna utilizado no cálculo da 2D-DCT [36].

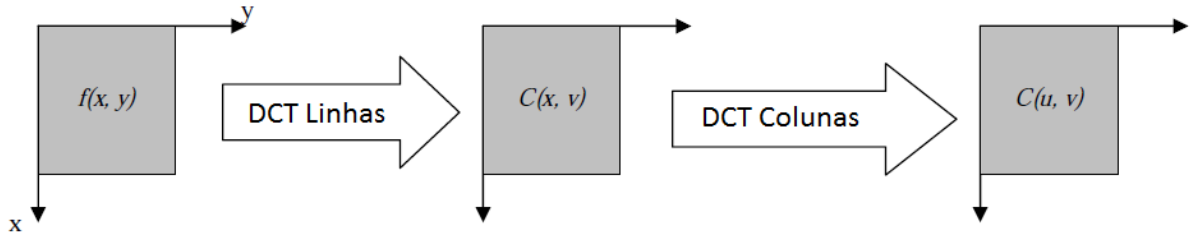


Figura 4.1 – Diagrama do método linha-coluna utilizado no cálculo da 2D-DCT.

O algoritmo proposto neste trabalho para o reconhecimento de faces utiliza a 2D-DCT porque ela tem a propriedade de compactação de energia em torno da origem. Com essa compactação, após a execução da 2D-DCT, muitos coeficientes distantes da origem apresentam valores próximos de zero e, portanto, podem ser desconsiderados sem que ocorra perda significativa de informações contidas na imagem.

4.3 *Two-Dimensional Principal Component Analysis* (2D-PCA)

No método PCA, descrito no Capitulo 2 deste trabalho, as matrizes de cada imagens precisam ser transformadas em vetores unidimensionais. Como resultado, os vetores-imagem resultantes da transformação trabalham em espaços de dimensões muito altas, tornando difícil analisar a matriz de covariância precisamente, em função do seu tamanho.

Em 2004, Yang e outros [37] desenvolveram um método de projeção direto conhecido como 2D-PCA (do inglês, *two-dimensional principal component analysis*), com o objetivo de extrair características de imagens. Nesse método, diferente da representação utilizada na PCA, as imagens são representadas por matrizes bidimensionais. Desta forma, a etapa de transformação da matriz em um vetor não é utilizada. Além disso, diferente da matriz de covariância obtida pela técnica PCA a matriz gerada pela 2D-PCA é bem menor. Como resultado, torna-se mais fácil avaliar a matriz de covariância precisamente e conseqüentemente é necessário menos tempo de processamento para a determinação dos autovetores.

Considere A uma imagem de uma face de tamanho $m \times n$ pixels representada por uma matriz qualquer e um vetor X n-dimensional sobre o qual a imagem é projetada utilizando-se a transformação linear da Equação (4.6).

$$\boldsymbol{Y} = \boldsymbol{AX} \tag{4.6}$$

Com a transformação linear da imagem mostrada na Equação (4.6) obtém-se o vetor Y m-dimensional que é denominado vetor característico da imagem A. A avaliação da qualidade da projeção realizada pelo vetor pode ser realizada utilizando-se a medida de dispersão total das imagens projetadas. A dispersão total dos vetores produzidos na projeção das imagens pode ser medida pelo traço da matriz de covariância gerada a partir desses vetores. Desta forma, pode-se adotar o critério da Equação (4.7).

$$J(X) = tr(S_x) \tag{4.7}$$

Na Equação (4.7) S_x corresponde à matriz de covariância dos vetores de características obtidos das imagens da base de treinamento e $tr(S_x)$ corresponde ao traço da matriz de covariância. A maximização do critério definido na Equação (4.7) resulta em encontrar um vetor

X, sobre o qual as imagens serão projetadas, e que maximize a dispersão total dos vetores de características obtidos. A matriz de covariância S_x pode ser obtida utilizando-se a Equação (4.8).

$$\begin{aligned} S_x &= E(Y-EY)(Y-EY)^T = E[AX-E(AX)][AX-E(A)]^T \\ &= E[(A-EA)X][(A-EA)X]^T \end{aligned} \tag{4.8}$$

Portanto,

$$\begin{aligned} tr(S_x) &= X^T[E(A-EA)^T(A-EA)]X \\ G_T &= E(A-EA)^T(A-EA) \\ tr(S_x) &= X^T G_T X \end{aligned} \tag{4.9}$$

A matriz G_t definida na Equação (4.9) representa a matriz de covariância das imagens do grupo de treinamento. Pode-se verificar nessa equação que G_t é uma matriz quadrada de tamanho $n \times n$. A matriz G_t pode ser obtida diretamente das imagens de treinamento. Supondo que tenha M amostras de imagens no grupo de treinamento, a j-ésima imagem de treinamento é representada por uma matriz de tamanho m x n denominada A_j ($j = 1, 2, ..., M$) e a média de todas as imagens de treinamento é representada por $\bar{A}$. Com isso, pode-se definir G_t pela Equação (4.10).

$$G_t = \frac{1}{M}\sum_{j=1}^{M}(A_j-\bar{A})^T(A_j-\bar{A}) \tag{4.10}$$

Desse modo o critério definido na Equação (4.7) pode ser reescrito obtendo-se a Equação (4.11).

$$J(X) = X^T G_T X \tag{4.11}$$

O critério mostrado na Equação (4.11) é conhecido como critério generalizado de dispersão total (do inglês, *generalized total scatter criterion*). O vetor coluna X que maximiza essa equação é o vetor que maximiza a dispersão total das imagens projetadas sobre ele.

O vetor de projeção ótima X_{opt} é o vetor que maximiza $J(X)$, isto é, ele é o autovetor correspondente aos maiores autovalores. Assim, como utilizado no método PCA, em geral escolhe-se um conjunto de vetores ótimos para a extração dos pesos. Estes vetores devem ser ortonormais e maximizar o critério originalmente definido na Equação (4.7). Desta forma, os vetores de projeção ótima, X_1, X_2, X_3, ..., X_d, são os autovetores de G_T correspondentes aos d maiores autovalores.

Os vetores de projeção da técnica 2D-PCA, X_1, X_2, X_3, ..., X_d, são utilizados para a extração de características das imagens das faces. Para uma imagem A pode-se extrair os vetores característicos utilizando-se a Equação (4.12).

$$Y_k = AX_k, \qquad k = 1, 2, ..., d \tag{4.12}$$

Utilizando-se a Equação (4.12), obtém-se um conjunto de vetores característicos Y_1, Y_2, ..., Y_d, chamados de vetores de componentes principais da imagem A. É importante notar que cada componente principal extraída com essa técnica é um vetor, enquanto que na técnica PCA cada componente é um valor escalar (coeficiente).

Os vetores de componentes principais extraídos usando-se a Equação (4.12) são utilizados para construir a matriz $B = [Y_1, Y_2, ..., Y_d]$ de tamanho $m \times d$ conhecida como matriz característica (*feature matrix*) da imagem.

Na classificação proposta por Yang e outros [37] a imagem a ser testada, I_t, é projetada utilizando-se os vetores de projeção, X_1, X_2, X_3, ..., X_d, obtendo-se assim a matriz característica B_t. Usando-se um classificador baseado na distância Euclidiana, a distância calculada entre as

matrizes $B_t = [Y_1^{(t)}, Y_2^{(t)}, ..., Y_d^{(t)}]$ e $B_x = [Y_1^{(x)}, Y_2^{(x)}, ..., Y_d^{(x)}]$ é obtida utilizando-se a Equação (4.13), onde B_x representa uma matriz característica de uma imagem pertencente ao grupo de treinamento.

$$d(B_t, B_x) = \sum_{k=1}^{d} \| Y_k^{(t)} - Y_k^{(x)} \| \tag{4.13}$$

Onde $\| Y_k^{(t)} - Y_k^{(x)} \|$ representa a distância Euclidiana entre os dois vetores característicos $Y_k^{(t)}$ e $Y_k^{(x)}$.

O motivo de utilizar a 2D-PCA ao invés da PCA para obtenção das matrizes características no algoritmo proposto neste trabalho, antes do treinamento e da classificação da rede CSOM, é porque a 2D-PCA é simples em extração de característica/informações de imagens, melhor em taxa de reconhecimento e mais eficiente computacionalmente.

4.4 *Concurrent Self-Organizing Maps (CSOM)*

O *Sel-Organizing Map* (SOM) (também conhecido como rede de Kohonen) é caracterizada pelo fato dos neurônios vizinhos pertencentes à mesma rede se desenvolverem adaptativamente em um detector especifico de diferentes padrões de vetores. Os neurônios tornam-se especificamente ajustados para várias classes de padrões por um processo de aprendizado competitivo e não supervisionado (*self-organizing learning*). Somente um neurônio ou grupo de neurônios ao mesmo tempo obtém resposta ativa para o padrão de entrada apresentado à rede [7].

A localização espacial do neurônio pertencente à rede (especificado pela sua coordenada) corresponde a um padrão de vetor de entrada particular pertencente a uma determinada classe. Uma característica importante do SOM é que ela pode simultaneamente extrair as estatísticas do vetor de entrada e realizar a classificação.

Os mapas auto-organizáveis (SOM) foram utilizados com sucesso em vários algoritmos relacionados a análise de *cluster*, reconhecimento de fala, classificação de faces, quantização de vetores, etc. A principal finalidade da rede SOM é memorizar/transformar padrões de dimensões arbitrárias em respostas de neurônios arranjados em 1-D ou 2-D, preservando o mapeamento em relação à vizinhança dos padrões de um espaço para outro.

Neagoe e Repot [38] desenvolveram uma rede neural conhecida como *Concurrent Self-Organizing Maps* (CSOM). A CSOM é uma coleção de pequenos SOMs, que utiliza a estratégia global *winner-takes-all*. Cada rede SOM é utilizada para classificar corretamente somente os padrões de uma determinada classe, onde a quantidade de redes SOMs é equivalente ao número de classes.

O treinamento de uma rede CSOM é supervisionado, onde cada SOM utiliza um algoritmo de treinamento individual usando somente os padrões de determinada classe. A Figura 4.2 mostra o modelo de treinamento de uma CSOM com “n”classes de padrões.

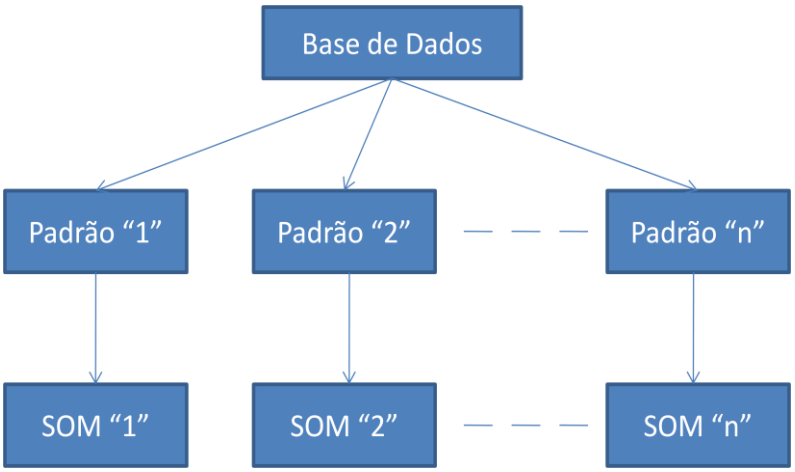


Figura 4.2 – Modelo de treinamento de uma rede CSOM.

Para o reconhecimento utilizando uma rede CSOM o padrão usado para testes é aplicado em paralelo em cada rede SOM já treinada. O padrão de entrada de teste é classificado como pertencente à classe da rede SOM cujo neurônio vencedor é pertencente. A Figura 4.3 mostra o modelo de classificação de padrões em uma rede CSOM com “n”classes.

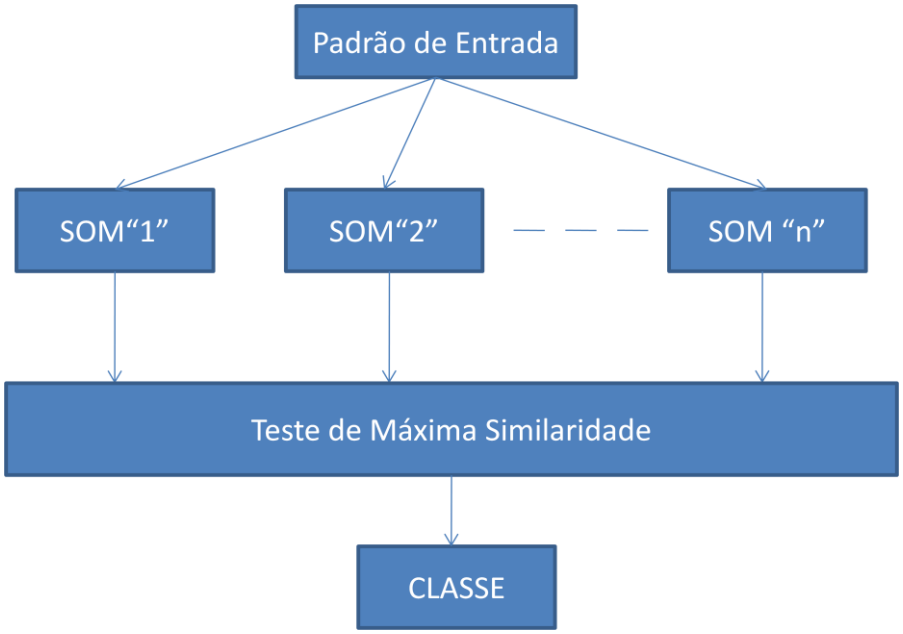


Figura 4.3 – Modelo de classificação de uma rede CSOM.

No trabalho de Neagoe e Repot [38] foi mostrado que ao utilizar uma rede CSOM consegue-se uma boa taxa de reconhecimento devido ao treinamento separado de cada SOM Assim, é necessário um tempo de treinamento menor porque cada vetor característico de entrada é aplicado somente a uma pequena SOM especifica de sua classe. Outra característica da rede CSOM é que o raio da vizinhança dela é menor, se comparado com uma única grande rede SOM. Baseado nesses conceitos a rede CSOM foi escolhida neste trabalho para ser utilizada na etapa de classificação ao invés de uma única rede SOM.

4.5 Algoritmo Proposto neste Trabalho

No algoritmo proposto neste trabalho [39] é realizada uma etapa de pré-processamento utilizando a 2D-DCT e a 2D-PCA com a finalidade de obter uma melhor extração de características (informações) das imagens e aumentar com isso a eficiência computacional na etapa de classificação.

A arquitetura utilizada na classificação foi o modelo CSOM implementado por Neagoe e Ropot [38]. Neste modelo cada rede SOM é treinada individualmente, obtendo dessa maneira resultados melhores do que se for treinada uma única rede SOM para todas as classes. O número de redes SOM utilizadas no treinamento é igual ao número de indivíduos.

Na fase de treinamento aplica-se a 2D-DCT em cada imagem da base de treinamento. Com o objetivo de diminuir a quantidade de coeficientes para representar a matriz característica de cada face, armazena-se em uma matriz quadrada apenas os coeficientes mais significativos (em termos energéticos) que estão localizados em torno da origem, como mostrado na Figura 4.4. Esta figura mostra os coeficientes da 2D-DCT que foi aplicada em uma face obtida da base de dados da "*The ORL Database of Faces*" disponibilizada pelo Laboratório AT&T da Universidade de Cambridge [40].

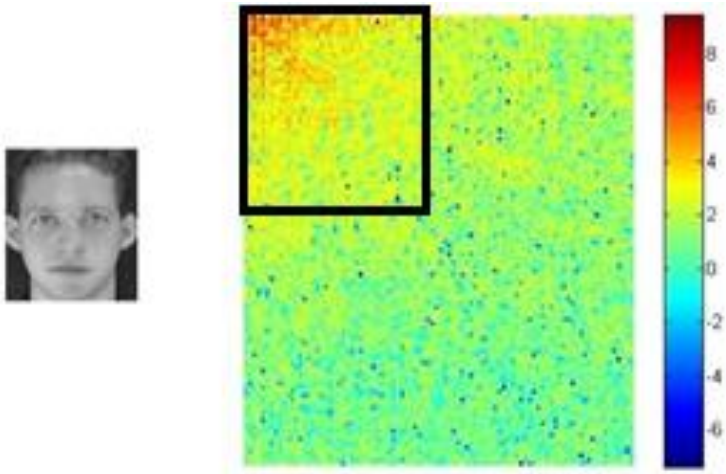


Figura 4.4 – 2D-DCT de uma face da base de dados da ORL e a seleção dos coeficientes mais significativos.

A rede CSOM é formada por *n* SOMs, onde *n* é o número de indivíduos/classe que o algoritmo é treinado para a classificação. Cada rede SOM é treinada com as matrizes características de cada indivíduo. Essas matrizes são encontradas aplicando-se a 2D-PCA nas matrizes quadradas dos coeficientes mais significativos da 2D-DCT de cada face do grupo de treinamento.

O diagrama de blocos do modelo proposto neste trabalho para o treinamento é mostrado na Figura 4.5 [39]. Este diagrama mostra as etapas relacionadas à obtenção da matriz característica de cada face utilizando-se a 2D-DCT e a 2D-PCA, e o estágio de treinamento individual das redes neurais SOM.

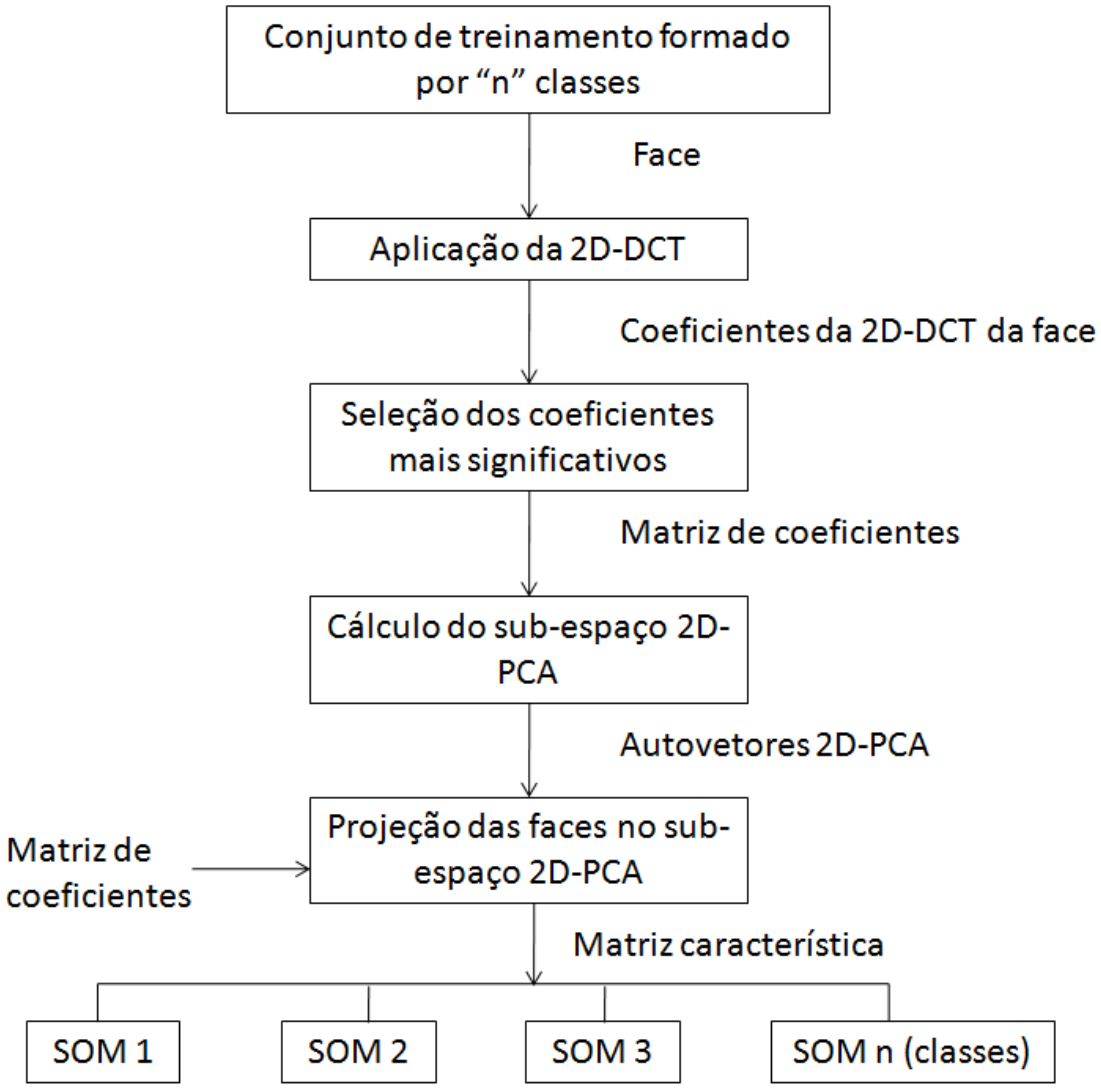


Figura 4.5 – Diagrama de blocos da fase de treinamento do modelo proposto neste trabalho.

O diagrama de blocos da fase de reconhecimento/classificação do modelo proposto nesta dissertação é mostrado na Figura 4.6 [39]. Na classificação, a matriz característica da imagem consulta é obtida da mesma maneira como foi encontrada as matrizes características das imagens das faces na fase treinamento.

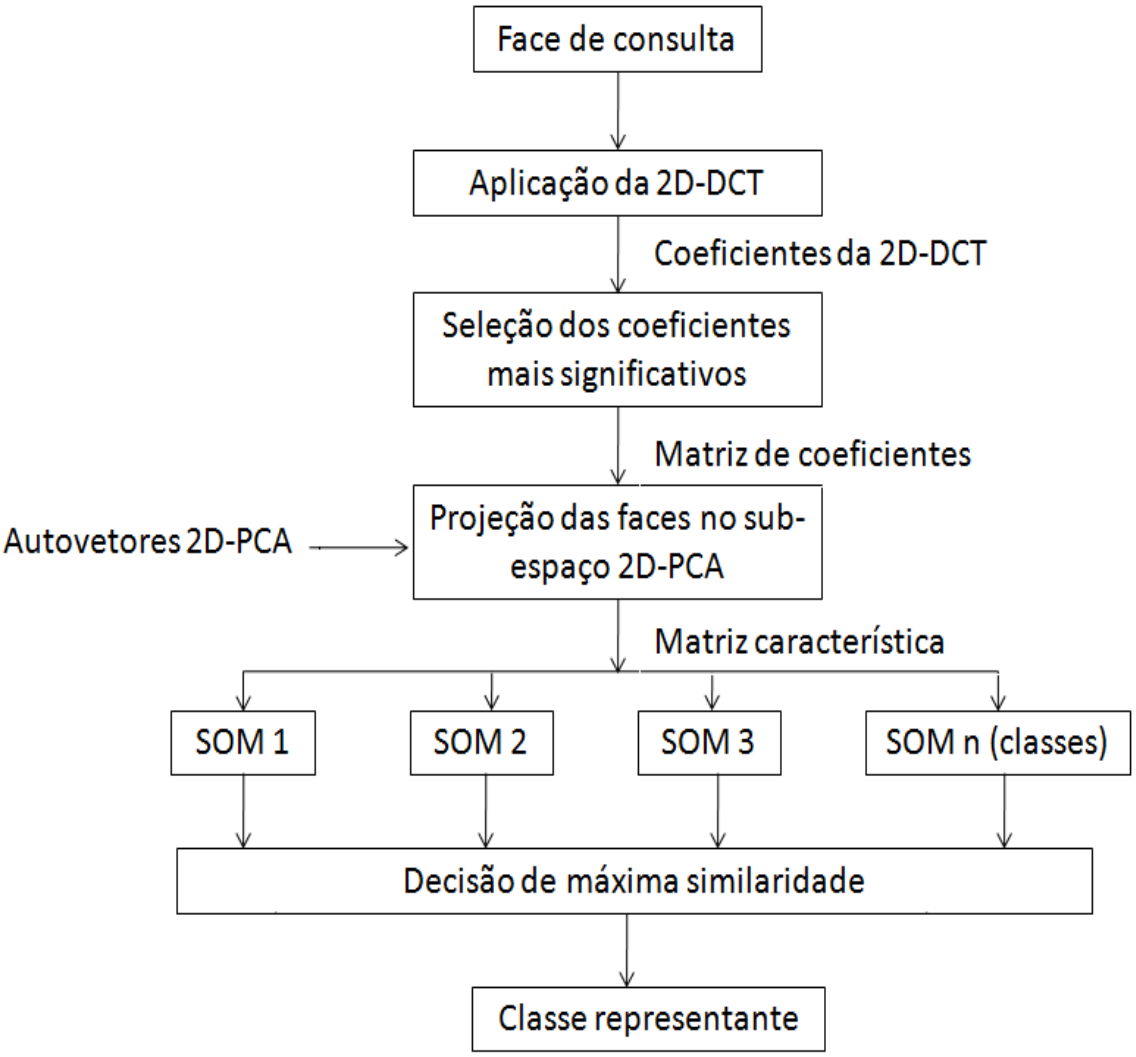


Figura 4.6 – Diagrama de blocos da fase de reconhecimento/classificação do modelo proposto neste trabalho.

Após o cálculo da similaridade, que é realizado utilizando a distancia Euclidiana, a face apresentada será classificada como pertencente a classe da rede SOM do neurônio mais próximo da matriz característica.

4.6 Considerações Finais Deste Capítulo

Este capítulo apresentou as técnicas 2D-DCT e 2D-PCA utilizadas neste trabalho para a obtenção das matrizes características e a rede neural CSOM usada na classificação das faces. Em seguida foi mostrado o algoritmo proposto nesta dissertação para o reconhecimento de faces.

O motivo de utilizar a 2D-PCA no algoritmo ao invés do PCA antes do treinamento e da classificação da rede CSOM, foi porque a 2D-PCA é uma técnica simples em extração de característica/informações de imagens, apresenta melhor taxa de reconhecimento e é mais eficiente computacionalmente. Entretanto, a 2D-PCA não é tão eficiente em termos de armazenamento, pois são necessários mais coeficientes para a representação da imagem.

Devido a deficiência apresentada da 2D-PCA em termos de armazenamento, a matriz característica no algoritmo proposto é obtida aplicando-se a 2D-DCT antes da 2D-PCA, reduzindo desse modo o espaço de armazenamento e os coeficientes de representação das imagens sem diminuir a taxa de reconhecimento.

Analisando o trabalho de Neagoe e Repot [38], pode-se observar que a rede CSOM obtém uma boa taxa de reconhecimento por treinar separadamente cada SOM, com um tempo de treinamento menor porque cada vetor característico de entrada é aplicado somente a um pequeno SOM específico de sua classe. Outra característica é que o raio da vizinhança é menor no CSOM, se comparado com um único grande SOM. Baseado nesses conceitos a rede CSOM foi escolhida nesta dissertação para ser utilizada na etapa de classificação ao invés de uma única rede SOM.

O próximo capitulo apresenta as bases de dados utilizadas nos testes com o algoritmo proposto nessa dissertação para o reconhecimento/classificação de face e os resultados obtidos nesses testes.

Capítulo 5

RESULTADOS OBTIDOS

5.1 Introdução

Neste capítulo são apresentados os resultados obtidos do algoritmo de reconhecimento de faces proposto neste trabalho utilizando-se três bancos de imagens bem conhecidos na área de processamento de imagens: ORL, YaleA e Face94. Finalmente, são realizadas conclusões sobre esses resultados.

Com o objetivo de mostrar a eficiência da rede neural CSOM sobre a rede SOM, também são realizados alguns testes utilizando a mesma técnica de obtenção da matriz característica do algoritmo proposto nesta dissertação, porém é usada na etapa de treinamento e classificação somente uma rede SOM.

Para os vários testes realizados neste trabalho, o conjunto de imagens das bases de dados foi dividido em um conjunto de treinamento e em um conjunto de testes, variando a quantidade de imagens utilizadas em cada conjunto. A finalidade desses testes foi verificar a resposta do algoritmo em relação a quantidade de imagens utilizadas no grupo de treinamento e no grupo de teste.

Também é realizado um processo de validação cruzada, que é menos tendencioso na avaliação de erro na classificação, mas também é o que mais exige recursos computacionais. Esse processo é conhecido como *Leave-one-out* [30].

O algoritmo proposto neste trabalho e todos os testes foram realizados utilizando o software Matlab 2009b com o SOM Toolbox desenvolvido pelo L*aboratory of Computer and Information Science* (CIS) da *Helsinki University of Technology* [41]. Eles foram realizados em um computador pessoal Intel Core 2 Duo de 2.33 GHz com 2 GB de RAM com sistema operacional Microsoft Windows Vista.

5.2 Banco de Imagens

Para avaliar o desempenho do algoritmo proposto nesta dissertação foram utilizados os bancos de faces da ORL, Face94 e YaleA. Estes bancos são utilizados comumente em testes de algoritmos de reconhecimento facial.

5.2.1 ORL

O banco de faces da ORL foi produzido pelo *Olivetti Research Laboratory* em *Cambrige,*UK. Este banco é formado por 10 imagens diferentes de 40 indivíduos distintos, com um total de 400 imagens. Existem variações na expressão facial e nos detalhes faciais do mesmo indivíduo. As imagens de cada indivíduo foram capturadas em horas diferentes, com condições de iluminação não padronizada e em alguns indivíduos foram adquiridas com ou sem óculos [40].

Todas as imagens foram adquiridas sob uma cena de fundo homogênea e escura com os indivíduos em posição frontal e ereta apresentando uma tolerância de mais ou menos 20

graus com uma resolução de 92 x 112 pixels. A Figura 5.1 mostra imagens de faces de alguns indivíduos presentes no banco de faces ORL.

Figura 5.1 – Exemplo de imagens de faces presentes no banco de faces ORL [40].

5.2.2 YaleA

A base de dados YaleA [42] é composta por 165 imagens em escala de cinza de 15 indivíduos, com onze imagens diferentes de cada indivíduo. Cada face é caracterizada por diferentes expressões faciais ou aspectos: feliz, normal, triste, sonolento, surpreso e piscando os olhos; com e sem óculos; e, com três direções de iluminação: luz central, luz lateral esquerda e luz lateral direita.

Um ponto negativo da base de imagens YaleA é que as faces não estão bem centralizadas, com as faces localizadas mais a direita ou a esquerda. As imagens são disponibilizadas com a resolução de 320 x 243 pixels. A Figura 5.2 mostra imagens de faces de alguns indivíduos presentes nesse banco de faces.

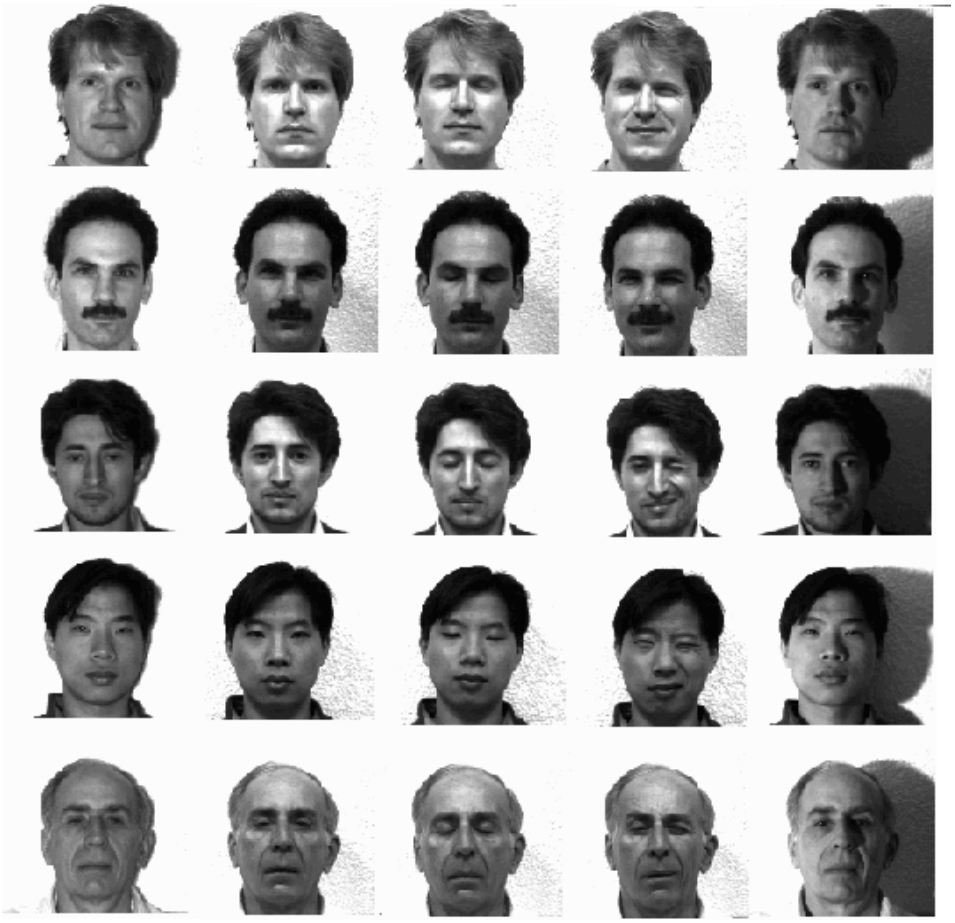


Figura 5.2 – Exemplo de imagens de faces presentes no banco de faces YaleA [42].

5.2.3 Face94

A base de dados Face 94 foi produzida pelo Dr Libor Spaceké da Universidade de Essex do Reino Unido. Ela é composta por 153 indivíduos, com 20 imagens colorida de cada indíviduo. Cada face é caracterizada por diferentes variações nas expressões faciais, com pouca variação na intensidade luminosa e com alguns indivíduos com ou sem óculos. As imagens foram obtidas

com o fundo no tom de verde com resolução de 180 x 200 pixels. A Figura 5.3 mostra imagens de faces de alguns indivíduos presentes nesse banco de faces [43].

Figura 5.3 – Exemplo de imagens de faces presentes no banco de faces Face94 [43].

5.3 Testes Realizados com a Base de Dados ORL

No primeiro teste realizado com a base de dados ORL foram escolhidas aleatoriamente 5 imagens de treinamento e 5 para testes de cada indivíduo, com um total 200 imagens para treinamento e 200 imagens para teste.

Conforme o diagrama de blocos da fase de treinamento do modelo proposto mostrado na Figura 4.5 do Capítulo 4 deste trabalho, foi aplicada inicialmente a 2D-DCT em cada imagem do grupo de treinamento, armazenando seus coeficientes em uma matriz quadrada. Pode-se observar na Figura 5.4 que as componentes mais significativas agrupam-se próximo a origem. Devido a isso, pode-se utilizar apenas os 20 x 20 coeficientes da 2D-DCT localizados próximos à origem para representar cada face. A utilização dessa quantidade de componentes reduz o tempo de processamento do algoritmo proposto nesta dissertação nas próximas etapas.

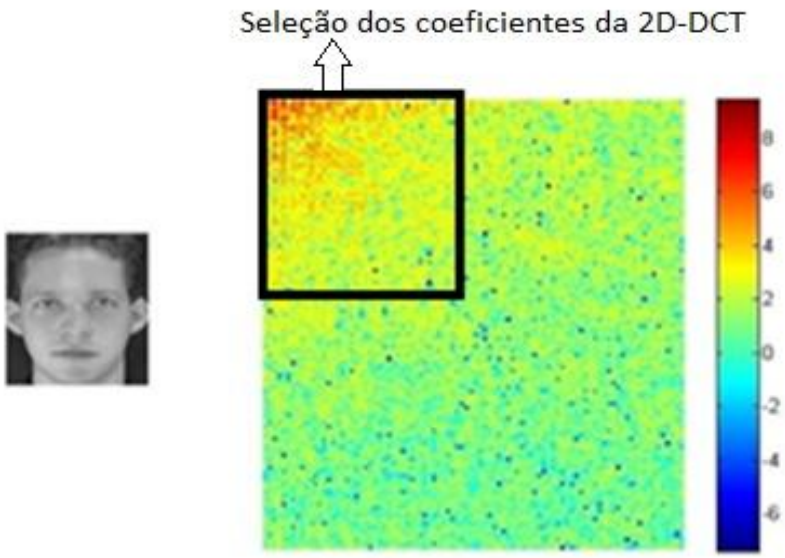


Figura 5.4 – Seleção dos coeficientes mais significativos da 2D-DCT de uma face da base de dados ORL.

Com os coeficientes selecionados da 2D-DCT das faces do grupo de treinamento, calcula-se a matriz característica de cada um delas utilizando-se os maiores autovalores da matriz de covariância da 2D-PCA. Como são usados 20 x 20 dos coeficientes da 2D-DCT de cada imagem do grupo de treinamento para criar a matriz de covariância da 2D-PCA, têm-se um total de 20 autovalores. A Figura 5.5 mostra esses autovalores normalizados em ordem decrescente.

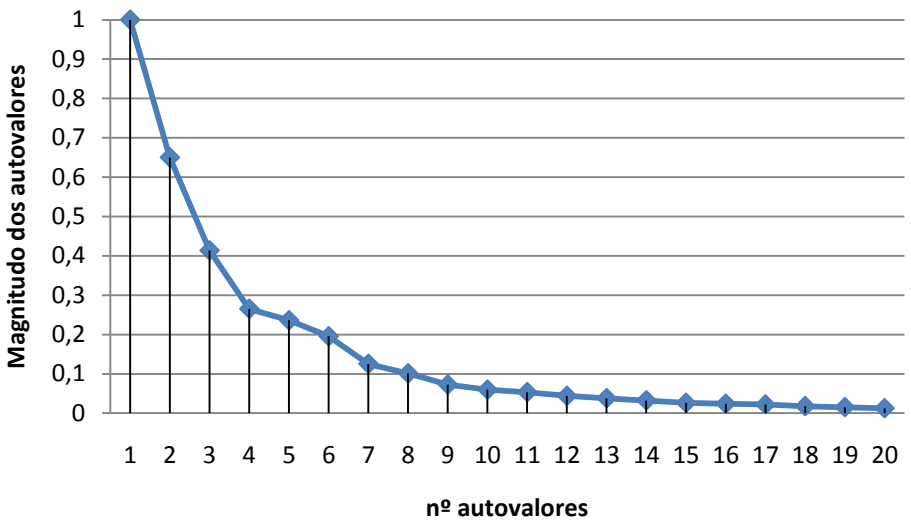


Figura 5.5 – Autovalores normalizados do conjunto de treinamento da base de dados ORL.

Pode-se observar na Figura 5.5 que as componentes mais significativas estão localizadas nos 10 primeiros autovalores da matriz de covariância. Assim, são utilizados esses 10 autovalores dessa matriz, resultando em uma matriz característica 20 x 10, ou seja, um vetor característico de 200 componentes para representar cada imagem na fase de treinamento e de classificação do algoritmo proposto neste trabalho.

O classificador utilizado nesta dissertação (última etapa desse algoritmo) foi a rede CSOM treinada com as matrizes características obtidas nas etapas anteriores. Cada SOM foi treinada separadamente para representar um indivíduo específico do banco de dados da ORL. A rede CSOM é formada por 40 SOMs, pois essa base de dados contém 40 indivíduos.

Para os testes mostrados na Tabela 5.1 foram escolhidas aleatoriamente 5 imagens de treinamento e 5 para testes de cada indivíduo, com um total de 200 imagens no grupo de treinamento e a mesma quantidade no grupo de teste. Os testes foram realizados variando a quantidade de neurônios em cada SOM, com o objetivo de verificar o desempenho do algoritmo proposto neste trabalho em relação a quantidade de neurônios em cada SOM.

Cada rede SOM foi treinada individualmente utilizando a estrutura topológica linear de Kohonem, com o método de atualização gaussiano mostrado na Equação (3.13) do Capítulo 3 desta dissertação.

Na Tabela 5.1 são mostradas as taxas de reconhecimento dos grupos de treinamento e de teste, o tempo de treinamento e o número de épocas por SOM, em relação a quantidade de neurônios de cada SOM da rede CSOM utilizada nos testes.

Tabela 5.1 – Resultados obtidos utilizando a base de dados ORL.

Nº de Neurônios/SOM	Taxa de reconhecimento do Grupo de Treinamento (%)	Taxa de reconhecimento do Grupo de Teste %	Tempo de treinamento (s)	Nº de épocas/SOM
2	99	87,5	5,68404	300
4	100	91	6,901368	400
6	100	91	7,909604	500
8	100	91	10,190012	600
10	100	93,5	12,147408	700
12	100	93	14,225628	800
14	100	93	16,96998	900
16	100	93	20,392592	1000
18	100	93,5	23,861656	1100
20	100	93,5	29,041812	1200
22	100	93,5	32,926228	1300
24	100	92,5	37,516844	1400
26	100	94	41,5398	1500
28	100	94	50,18284	1600
30	100	93,5	55,53924	1700
32	100	94	61,88136	1800
34	100	94	77,03928	1900
36	100	93,5	76,57204	2000
38	100	95	87,18408	2100
40	100	93,5	93,774	2200

Pode-se observar na Tabela 5.1 que a maior taxa de reconhecimento no grupo de teste foi 95% para 38 neurônios em cada SOM, com um tempo de treinamento de 87,18 segundos.

Outro fator que é possível verificar na Tabela 5.1 é que a taxa de reconhecimento do grupo de teste oscila entre 93% a 94% quando é utilizado acima de 10 neurônios em cada rede SOM.

Com o objetivo de mostrar a eficiência da rede neural CSOM sobre a rede SOM, foram realizados testes utilizando a mesma técnica de obtenção da matriz característica do algoritmo proposto neste trabalho, porém na etapa de treinamento e classificação foram usados somente uma rede SOM. Nestes testes a estrutura topológica e o método de atualização dos neurônios foram os mesmos utilizados nos testes mostrados na Tabela 5.1. A Tabela 5.2 ilustra os resultados obtidos com uma única rede SOM usando a base de dados ORL.

Tabela 5.2 – Resultados obtidos com uma única rede SOM utilizando a base de dados ORL.

Nº de Neurônios da SOM	Taxa de Reconhecimento do Grupo de Treinamento (%)	Taxa de Reconhecimento do Grupo de Teste (%)	Tempo de treinamento (s)	Nº de épocas da SOM
80	85	69,5	2,265796	300
160	96,5	81,5	9,069004	400
240	100	89	20,64048	500
320	99,5	88,5	41,34748	600
400	100	89	81,78972	700
480	100	91,5	131,19548	800
560	100	91,5	190,27132	900
640	100	91,5	270,80756	1000
720	100	91,5	368,03988	1100
800	100	91,5	514,3412	1200
880	100	91,5	651,006	1300
960	100	91,5	840,2388	1400
1040	100	91,5	1080,0196	1500
1120	100	91,5	1211,8012	1600
1200	100	91,5	1443,8464	1700
1280	100	91,5	1727,3732	1800
1360	100	91,5	2041,2492	1900
1440	100	91,5	2457,392	2000
1520	100	91,5	2886,0288	2100
1600	100	91,5	3250,8784	2200

Pode-ser verificar na Tabela 5.2 que a maior taxa de reconhecimento no grupo de teste foi 91,5% utilizando 480 neurônios na rede SOM, com um tempo de treinamento de 131,2 segundos. E que a taxa de reconhecimento aumentou a medida que aumentou-se a quantidade de neurônios até 480 neurônios da rede SOM, depois dessa quantidade a taxa tornou-se constante mesmo aumentando a quantidade de neurônios.

A Figura 5.6 mostra os resultados das taxas de reconhecimento dos grupos de testes utilizados das Tabelas 5.1 e 5.2.

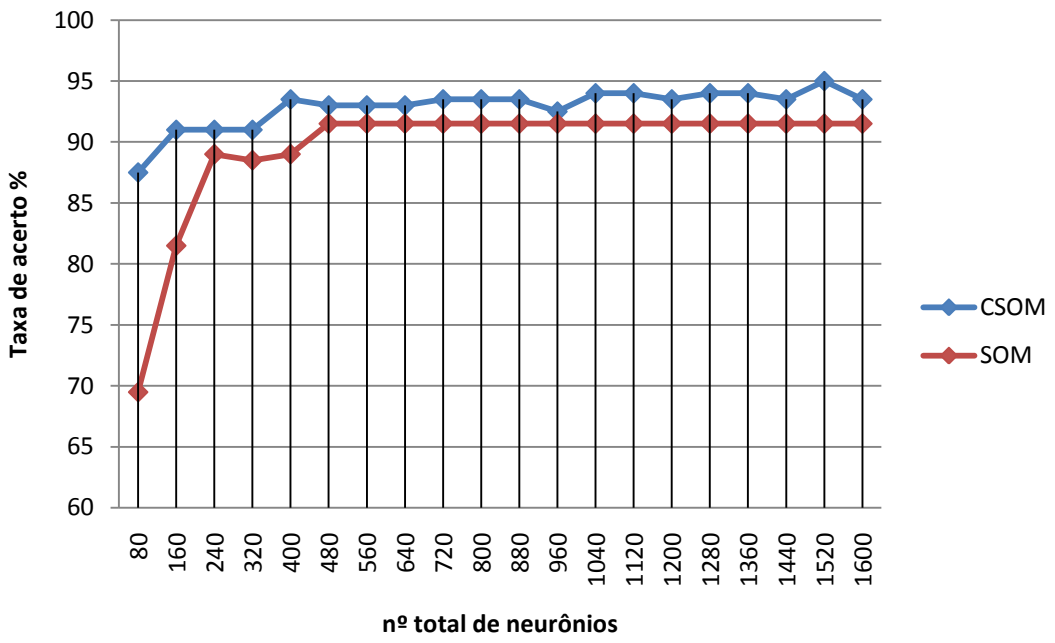


Figura 5.6 – Gráfico dos resultados obtidos das taxas de reconhecimento dos grupos de testes utilizados nas Tabelas 5.1 e 5.2.

Comparando os resultados obtidos nas Tabelas 5.1 e 5.2 pode-se observar que ao utilizar a rede CSOM na classificação obtém-se uma taxa de reconhecimento maior e um tempo de processamento menor do que quando utiliza-se uma única rede SOM na classificação para a mesma quantidade de neurônios.

A Figura 5.7 mostra os resultados obtidos nos testes variando a quantidade de autovetores na obtenção da matriz característica. Nesses testes foram escolhidas

aleatoriamente 5 imagens de treinamento e 5 para testes de cada indivíduo, formando dessa maneira os grupos de treinamento e de teste compostos por 200 imagens cada (5 x 40). Nesses testes na etapa de classificação as redes CSOM e SOM possuem a mesma quantidade de neurônios.

Na Figura 5.7 cada rede SOM do algoritmo proposto neste trabalho foi treinada individualmente utilizando a estrutura topológica linear de Kohonem com 10 neurônios cada, utilizando o método de atualização gaussiano mostrado na Equação (3.13) do Capitulo 3 desta dissertação.

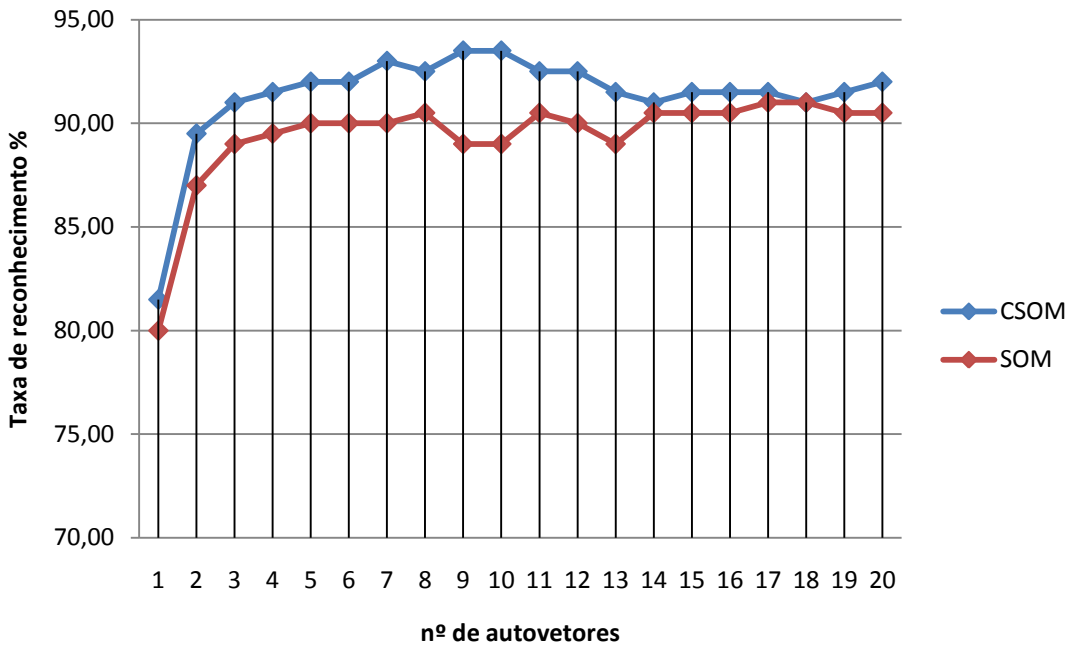


Figura 5.7 – Gráfico dos resultados obtidos utilizando-se as redes CSOM e SOM na base de dados ORL.

Pode-se observar na Figura 5.7, que em todos os testes a rede CSOM teve uma taxa de reconhecimento do grupo de teste melhor do que a rede SOM. A maior taxa de reconhecimento do grupo de teste utilizando a rede CSOM foi de 93,5 % para 9 e 10 neurônios.

A Tabela 5.3 mostra os resultados obtidos nos testes utilizando-se o algoritmo proposto neste trabalho, variando-se a quantidade de imagens usadas nos grupos de treinamento e de teste. A dimensão da matriz característica e a quantidade de neurônios em todos os testes foram mantidas iguais. Foi usado em cada rede SOM a estrutura topológica linear de Kohonem com 10 neurônios cada.

Tabela 5.3 – Resultados obtidos nos testes variando a quantidade de imagens utilizando-se o banco de dados ORL.

nº de faces de treinamento /indivíduo	Total de faces do grupo de treinamento	Total de faces do grupo de teste	Taxa de reconhecimento %
2	80	320	81,88
3	120	280	85,00
4	160	240	88,75
5	200	200	93,50
6	240	160	97,50
7	280	120	96,67
8	320	80	93,75
9	360	40	95,00

Pode-se observar na Tabela 5.3 que ao utilizar 6 faces por indivíduo no grupo de treinamento e 4 faces por indivíduo no grupo de testes, obteve-se uma taxa de reconhecimento 97,50%, ou seja, das 160 imagens utilizadas para testes apenas 4 foram identificadas incorretamente.

Foi realizado um teste com o algoritmo proposto neste trabalho utilizando o algoritmo *Leave-one-out*. Este algoritmo é bastante dispendioso computacionalmente e é usado quando é pequena a quantidade de dados para treinamento.

O treinamento do algoritmo *Leave-one-out* é repetido 400 vezes (quantidade de faces do banco de dados ORL). Em cada treinamento são utilizadas 399 faces e a única que não foi usada no treinamento é utilizada na classificação. No final é calculada a taxa média de reconhecimento dos 400 testes.

A taxa de reconhecimento ao utilizar o algoritmo *Leave-one-out* foi de 98,25 %, ou seja, das 400 faces apenas 7 faces foram identificadas incorretamente. A Figura 5.8 mostra as faces que foram identificadas incorretamente utilizando o *Leave-One-Out* no algoritmo proposto neste trabalho.

Figura 5.8 – Faces identificadas incorretamente utilizando o algoritmo *Leave-one-out*. na base de dados ORL.

5.4 Testes Realizados com a Base de Dados YaleA

No primeiro teste utilizando a base de dados YaleA foram escolhidas aleatoriamente 5 imagens de treinamento e 6 para testes de cada indivíduo, com um total de 75 e 90 imagens para treinamento e classificação, respectivamente.

Conforme o diagrama de blocos da fase de treinamento do modelo proposto neste trabalho mostrado na Figura 4.5 do Capítulo 4 deste trabalho, foi aplicada inicialmente a 2D-DCT em cada imagem do grupo de treinamento, armazenando seus coeficientes em uma matriz quadrada. Pode-se observar na Figura 5.9 que as componentes mais significativas agrupam-se próximo a origem. Devido a isso, pode-se usar apenas as 50 x 50 componentes da 2D-DCT localizadas próximas a origem para representar cada face. A utilização dessa quantidade de componentes reduz o tempo de processamento do algoritmo proposto nesta dissertação nas próximas etapas.

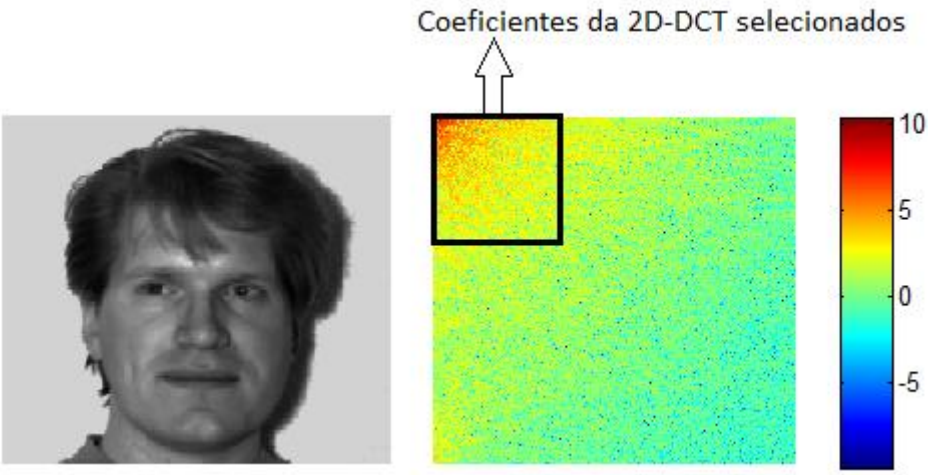


Figura 5.9 – Seleção dos coeficientes mais significativos da 2D-DCT de uma face da base de dados YaleA.

Com os coeficientes selecionados da 2D-DCT de cada individuo do grupo de treinamento, calculou-se a matriz característica de cada um deles utilizando-se os maiores autovalores da matriz de covariância da 2D-PCA. Como são usados 50 x 50 dos coeficientes da 2D-DCT de cada imagem do grupo de treinamento para criar a matriz de covariância da 2D-PCA, têm-se um total de 50 autovalores. A Figura 5.10 mostra esses autovalores normalizados em ordem decrescente.

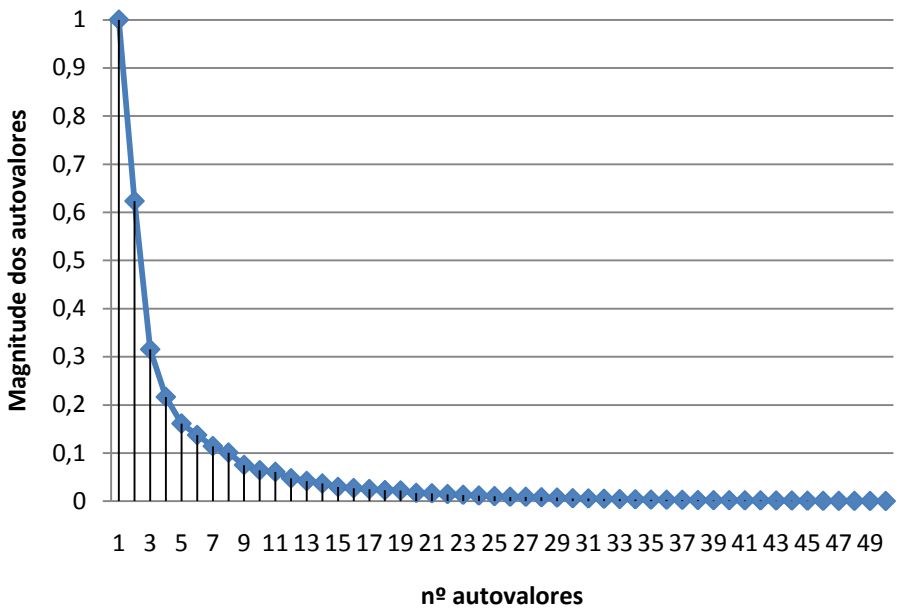


Figura 5.10 – Autovalores normalizados do grupo de treinamento da base de dados YaleA.

Pode-se observar na Figura 5.10 que as componentes mais significativas estão localizadas nos 15 primeiros autovalores da matriz de covariância. Assim, são usados esses 15 autovalores dessa matriz, resultando em uma matriz característica 50 x 15, ou seja, um vetor característico de 750 componentes para representar cada imagem na fase de treinamento e de classificação do algoritmo proposto neste trabalho.

O classificador utilizado nesta dissertação (última etapa desse algoritmo) foi a rede CSOM treinada com as matrizes características obtidas nas etapas anteriores. Cada SOM foi treinada separadamente para representar um indivíduo especifico do banco de dados da YaleA. A rede CSOM é formada por 15 SOMs, pois essa base de dados contém 15 indivíduos.

Para os testes mostrados na Tabela 5.4 foram escolhidas aleatoriamente 5 imagens de treinamento e 6 para teste de cada indivíduo, com um total de 75 imagens no grupo de treinamento e 90 imagens no grupo de teste. Os teses foram realizados variando a quantidade

de neurônios em cada SOM, com o objetivo de verificar o desempenho do algoritmo proposto neste trabalho em relação a quantidade de neurônios em cada SOM.

Cada rede SOM foi treinada individualmente utilizando a estrutura topológica linear de Kohonem, com o método de atualização gaussiano como mostrado na Equação (3.13) do Capítulo 3 desta dissertação.

Na Tabela 5.4 são mostradas as taxas de reconhecimento dos grupos de treinamento e de teste, o tempo de treinamento e o número de épocas por SOM, em relação a quantidade de neurônios de cada SOM da rede CSOM utilizada nos testes.

Tabela 5.4 – Resultados obtidos utilizando a base de dados YaleA.

Nº de Neurônios/SOM	Taxa de Reconhecimento do Grupo de Treinamento (%)	Taxa de Reconhecimento do Grupo de Teste (%)	Tempo de treinamento (s)	Nº de épocas/SOM
2	100	78,8889	8,0390	300
4	100	81,1111	9,0651	400
6	100	78,8889	10,8421	500
8	100	82,2222	13,2727	600
10	100	80,0000	15,0743	700
12	100	81,1111	18,0913	800
14	100	81,1111	22,2479	900
16	100	83,3333	25,5974	1000
18	100	81,1111	30,5291	1100
20	100	80,0000	38,2961	1200
22	100	81,1111	47,3695	1300
24	100	81,1111	49,4806	1400
26	100	80,0000	55,6330	1500
28	100	81,1111	58,9514	1600
30	100	81,1111	74,7814	1700

Pode-se observar na Tabela 5.4 que a maior taxa de reconhecimento no grupo de teste foi 83,33% para 16 neurônios em cada SOM, com um tempo de treinamento de 25,60

segundos. E que a taxa de reconhecimento hora aumentou hora diminuiu com o aumento do número de neurônios em cada rede SOM.

Com o objetivo de mostrar a eficiência da rede neural CSOM sobre a rede SOM, foram realizados testes utilizando a mesma técnica de obtenção da matriz característica do algoritmo proposto neste trabalho, porém na etapa de treinamento e classificação foram uados somente uma rede SOM. Nestes testes a estrutura topológica e o método de atualização dos neurônios foram os mesmos utilizados nos testes mostrados na Tabela 5.4. A Tabela 5.5 ilustra os resultados obtidos com uma única rede SOM usando a base de dados YaleA.

Tabela 5.5 – Resultados obtidos com uma única rede SOM utilizando a base de dados YaleA.

Nº de Neurônios/SOM	Taxa de Reconhecimento do Grupo de Treinamento (%)	Taxa de Reconhecimento do Grupo de Teste (%)	Tempo de treinamento (s)	Nº de épocas/SOM
30	97,33	76,67	4,35	300,00
60	100,00	77,78	16,79	400,00
90	100,00	78,89	50,31	500,00
120	100,00	77,78	77,77	600,00
150	100,00	77,78	135,55	700,00
180	100,00	78,89	213,86	800,00
210	100,00	77,78	346,10	900,00
240	100,00	78,89	461,31	1000,00
270	100,00	77,78	612,59	1100,00
300	100,00	78,89	848,80	1200,00
330	100,00	77,78	1072,73	1300,00
360	100,00	77,78	1274,21	1400,00
390	100,00	77,78	1574,75	1500,00
420	100,00	77,78	1903,45	1600,00
450	100,00	77,78	2208,45	1700,00

Pode-ser verificar na Tabela 5.5 que a maior taxa de reconhecimento no grupo de teste foi 78,89% para 90, 180 e 300 neurônios na rede SOM, com os tempos de treinamento de

50,31, 213,86 e 848,80 segundos, respectivamente. E que a taxa de reconhecimento hora aumentou hora diminuiu com o aumento do número de neurônios na rede SOM.

A Figura 5.11 mostra os resultados obtidos das taxas de reconhecimento dos grupos de utilizados nas Tabelas 5.4 e 5.5.

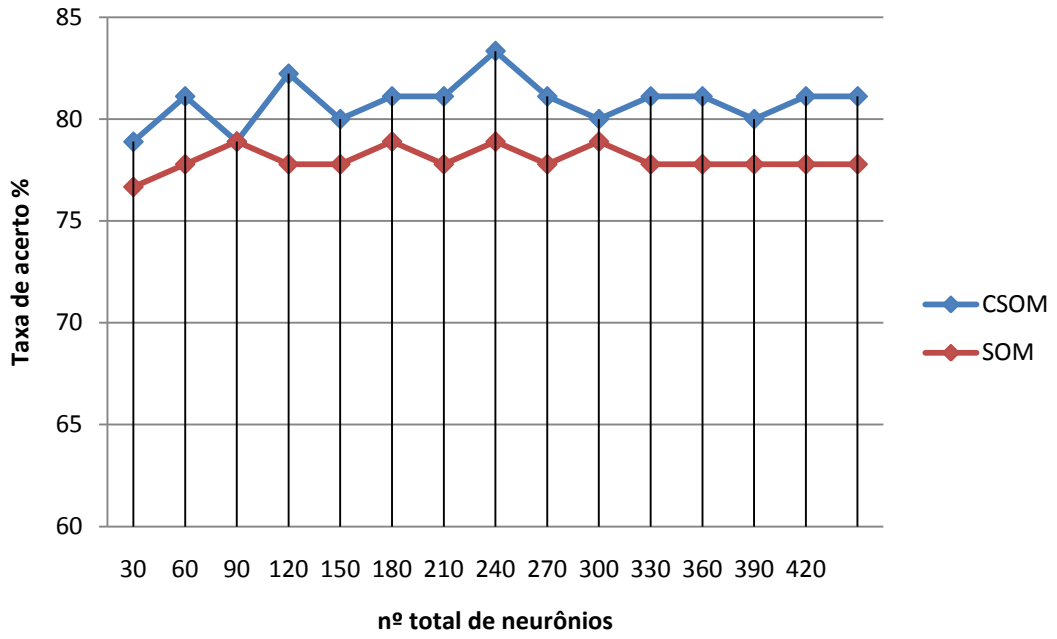


Figura 5.11 – Gráfico dos resultados obtidos das taxas de reconhecimento dos grupos de testes utilizados nas Tabelas 5.4 e 5.5.

Comparando os resultados obtidos nas Tabelas 5.4 e 5.5 pode-se observar que ao utilizar a rede CSOM na classificação obtém-se uma taxa de reconhecimento maior e um tempo de processamento menor do que quando utiliza-se uma única rede SOM na classificação para a mesma quantidade de neurônios.

A Tabela 5.6 mostra os resultados obtidos nos testes utilizando-se o algoritmo proposto neste trabalho, variando-se a quantidade de imagens usadas nos grupos de treinamento e de teste. A dimensão da matriz característica e a quantidade de neurônios em todos os testes foram mantidas iguais. Foi usada em cada rede SOM a estrutura topológica linear de Kohonem com 10 neurônios cada.

Tabela 5.6 – Resultados obtidos variando a quantidade de imagens utilizando-se a banco de dados YaleA.

Nº de faces de treinamento /indivíduo	Total de faces do grupo de treinamento	Total de faces do grupo de teste	Taxa de acerto %
2	30	135	85,9259
3	45	120	85,0000
4	60	105	82,8571
5	75	90	77,7778
6	90	75	77,3333
7	105	60	70,0000
8	120	45	75,5556
9	135	30	66,6667
10	150	15	100,0000

Pode-se observar na Tabela 5.6 que ao utilizar 2 faces por indivíduo no grupo de treinamento e 9 faces por indivíduo no grupo de teste, obteve-se uma taxa de reconhecimento 85,93%. Aumentando-se a quantidade de imagens no grupo treinamento a taxa de reconhecimento do grupo de teste diminui. Utilizando-se 10 imagens no grupo de treinamento para cada individuo, ou seja, um total de 150 faces no grupo de treinamento e 15 faces no grupo de teste, obteve-se uma taxa de 100% no reconhecimento das faces do grupo de teste.

Foi realizado um teste com o algoritmo proposto neste trabalho utilizando o algoritmo *Leave-one-out*.

O treinamento do algoritmo *Leave-one-out* é repetido 165 vezes (quantidade de faces do banco de dados YaleA). Em cada treinamento são usadas 164 faces e a única que não foi utilizada no treinamento é usada para a classificação. No final é calculada a taxa de reconhecimento dos 165 testes.

A taxa de reconhecimento ao utilizar o algoritmo *Leave-one-out* foi de 81,21%, ou seja, das 165 faces apenas 31 faces foram identificadas incorretamente.

A base de dados YaleA não apresentou bons resultados nos testes realizados com o algoritmo proposto, porque as faces desse banco não estão centralizadas no rosto dos indivíduos, e ocorre sombra no fundo de algumas imagens.

Para melhorar o desempenho do algoritmo proposto neste trabalho utilizando a base de dados YaleA é necessário realizar um pré-processamento nas imagens localizando as faces. Como o objetivo desta dissertação é a classificação/identificação de um sistema de reconhecimento de faces, não foi realizado o estágio de detecção automática da face.

Com os testes realizados utilizando-se a base de dados YaleA foi possível verificar a importância da fase de detecção de face em um sistema de reconhecimento de face. Diferente das faces desse banco, as faces da base de dados ORL e Face94 estão bem centralizadas. Assim, para os bancos de dados ORL e Face94 não é necessário a etapa de detecção automática de face.

5.5 Testes Realizados com a Base de Dados Face94

No primeiro teste realizado com a base de dados Face94 foram escolhidas aleatoriamente 10 faces de treinamento e 10 para testes de cada indivíduo, com um total de 1500 imagens para treinamento e 1500 imagens para teste. Antes de utilizar essa base de dados no algoritmo proposto deste trabalho, suas imagens coloridas foram convertidas para tons de cinza para diminuir a quantidade de informação em seu armazenamento.

Conforme o diagrama de blocos da fase de treinamento do modelo proposto neste trabalho mostrado na Figura 4.6 do Capítulo 4 deste trabalho, foi aplicada inicialmente a 2D-DCT em cada imagem do grupo de treinamento, armazenando seus coeficientes em uma

matriz quadrada. Pode-se observar na Figura 5.12 que as componentes mais significativas agrupam-se próximo a origem. Devido a isso, pode-se usar apenas as 20 x 20 componentes da 2D-DCT localizadas próximas a origem para representar cada face. A utilização dessa quantidade de componentes reduz o tempo de processamento do algoritmo proposto nesta dissertação nas próximas etapas.

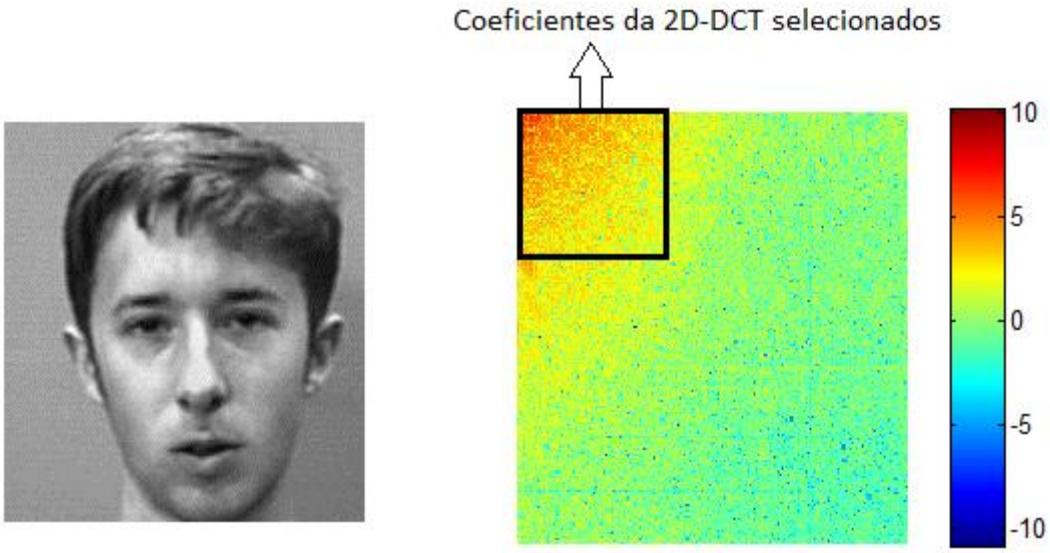


Figura 5.12 – Seleção dos coeficientes mais significativos da 2D-DCT de uma face da base de dados Face94.

Com os coeficientes selecionados da 2D-DCT de cada individuo do grupo de treinamento, calculou-se a matriz característica de cada um deles utilizando-se os maiores autovalores da matriz de covariância da 2D-PCA. Como são usados 20 x 20 dos coeficientes da 2D-DCT de cada imagem do grupo de treinamento para criar a matriz de covariância da 2D-PCA, têm-se um total de 20 autovalores. A Figura 5.13 mostra esses autovalores normalizados em ordem decrescente.

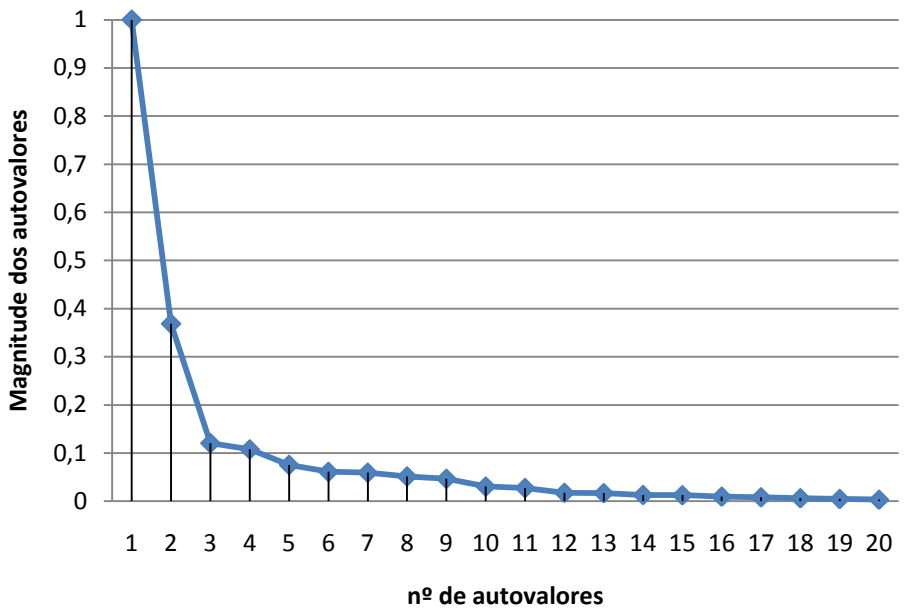


Figura 5.13 – Autovalores normalizados do grupo de treinamento da base de dados Face94.

Pode-se observar na Figura 5.13 que as componentes mais significativas estão localizadas nos 10 primeiros autovalores da matriz de covariância. Assim, são usados esses 10 autovalores dessa matriz, resultando em uma matriz característica 20 x 10, ou seja, um vetor característico de 200 componentes para representar cada imagem na fase de treinamento e de classificação do algoritmo proposto neste trabalho.

O classificador utilizado nesta dissertação (última etapa desse algoritmo) foi a rede CSOM treinada com as matrizes características obtidas nas etapas anteriores. Cada SOM foi treinada separadamente para representar um indivíduo especifico do banco de dados da Face94. A rede CSOM é formada por 150 SOMs, pois foram usados 150 indivíduos dessa base de dados.

Para os testes mostrados na Tabela 5.7 foram escolhidas aleatoriamente 10 imagens de treinamento e 10 para teste de cada indivíduo, com um total de 1500 imagens no grupo de treinamento e a mesma quantidade no grupo de teste. Os teses foram realizados variando a

quantidade de neurônios em cada SOM, com o objetivo de verificar o desempenho do algoritmo proposto neste trabalho em relação a quantidade de neurônios em cada SOM.

Cada rede SOM foi treinada individualmente utilizando a estrutura topológica linear de Kohonem, com o método de atualização gaussiano mostrado na Equação (3.13) do Capítulo 3 desta dissertação.

Tabela 5.7 –Resultados obtidos utilizando a base de dados Face94.

Nº de Neurônios/SOM	Taxa de Reconhecimento do Grupo de Treinamento (%)	Taxa de Reconhecimento do Grupo de Teste (%)	Tempo de treinamento (s)	Nº de épocas/SOM
2	99,8667	97,0000	19,1116	300
4	100	98,1333	20,8819	400
6	100	98,2000	25,6923	500
8	100	98,3333	25,3236	600
10	100	98,9333	29,9714	700
12	100	98,5333	32,1341	800
14	100	98,5333	38,5324	900
16	100	98,6000	39,6983	1000
18	100	98,6000	48,2737	1100
20	100	98,6000	52,6330	1200
22	100	98,6000	57,8013	1300
24	100	98,9333	65,9309	1400
26	100	98,6000	74,1595	1500
28	100	98,6000	83,1386	1600
30	100	98,6000	96,3029	1700
32	100	98,9333	101,9635	1800
34	100	98,6000	117,7375	1900
36	100	98,6000	130,1156	2000
38	100	98,6000	133,0583	2100
40	100	98,6000	159,6298	2200

Pode-se verificar na Tabela 5.7 que a maior taxa de reconhecimento no grupo de teste foi 98,93 % para 10 e 32 neurônios em cada SOM com um tempo de treinamento de 29,97 e

101,96 segundos, respectivamente. E que a taxa de reconhecimento foi praticamente a mesma independente da quantidade de neurônios em cada rede SOM.

A Figura 5.14 mostra os resultados obtidos das taxas de reconhecimento dos grupos de testes utilizados na Tabela 5.7.

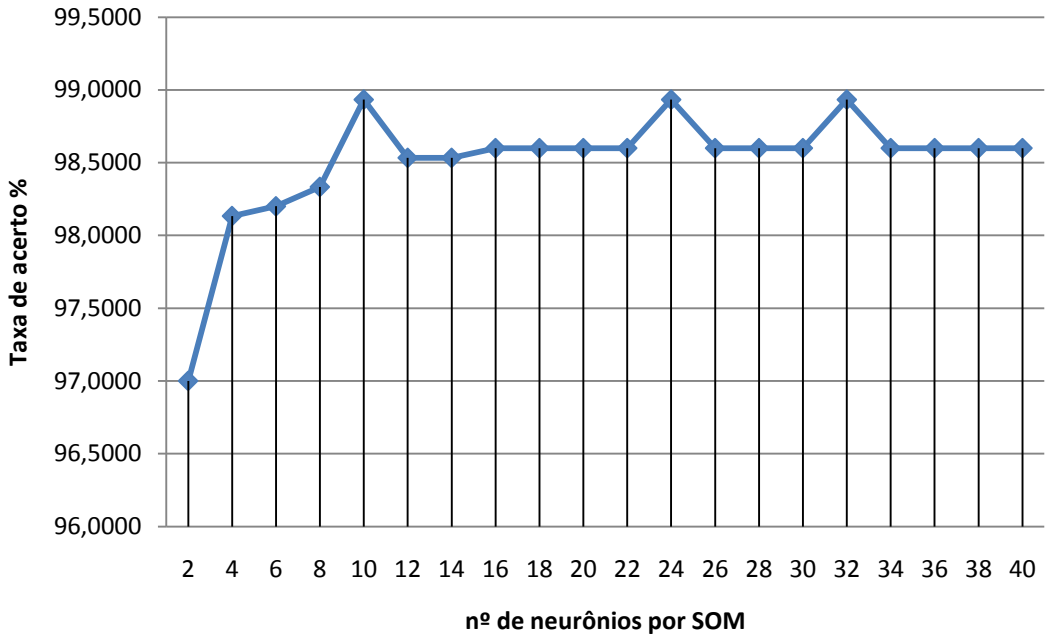


Figura 5.14 – Gráfico dos resultados obtidos das taxas de reconhecimento dos grupos de testes utilizados na Tabela 5.7.

Pode-se observar no gráfico da Figura 5.14 que a rede CSOM com 2 neurônios em cada SOM obteve uma taxa de reconhecimento de 97%. Mesmo aumentando a quantidade de neurônios, o algoritmo proposto neste trabalho não obteve um aumento nessa taxa de reconhecimento do grupo de teste. Acima de 10 neurônios por cada rede SOM a taxa de reconhecimento do grupo de teste oscila entre 98,93% e 98,6%.

A Figura 5.15 mostra os resultados obtidos nos testes variando a quantidade de autovetores na obtenção da matriz característica. Nesses testes foram escolhidas aleatoriamente 10 imagens de treinamento e 10 imagens para teste de cada indivíduo, formando dessa maneira os grupos de treinamento e de teste compostos por 15000 imagens

cada (10 x 150). Nesses testes na etapa de classificação as redes CSOM e SOM possuem a mesma quantidade de neurônios.

Na Figura 5.15 cada rede SOM do algoritmo proposto neste trabalho foi treinada individualmente utilizando a estrutura topológica linear de Kohonem com 10 neurônios cada, com o método de atualização gaussiano mostrado na Equação (3.13) do Capítulo 3 desta dissertação.

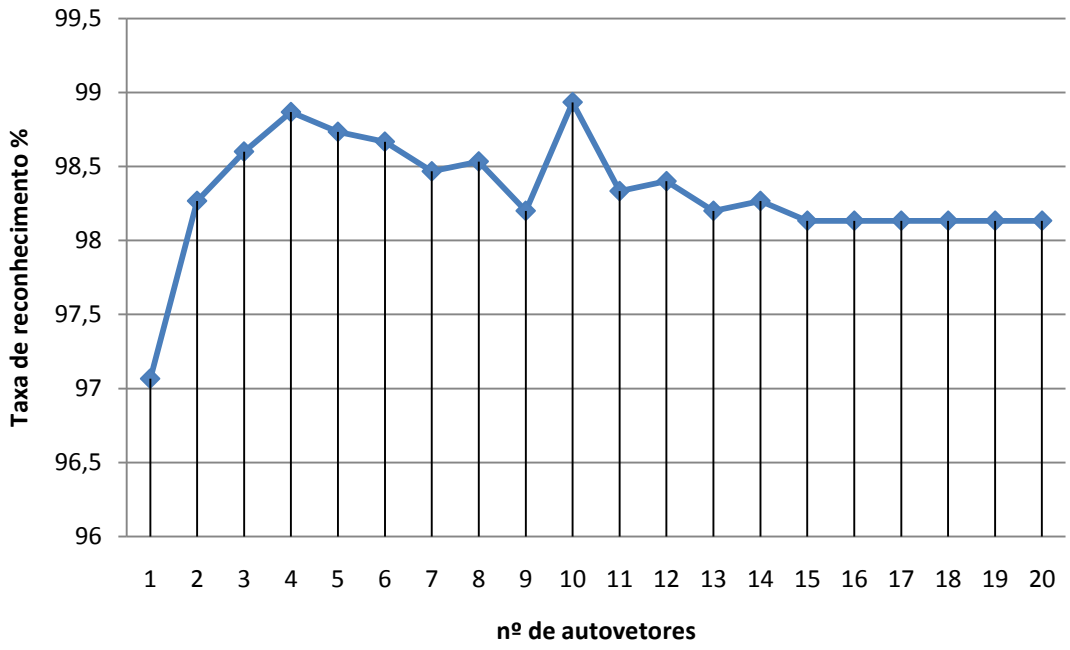


Figura 5.15 – Gráfico dos resultados obtidos do algoritmo proposto neste trabalho variando o número de autovalores utilizando-se Face94.

Pode-se verificar na Figura 5.15, que o algoritmo proposto neste trabalho teve uma boa taxa de reconhecimento no grupo de teste e a maior taxa foi de 98,93 % utilizando-se 10 neurônios em cada rede SOM.

A Tabela 5.8 mostra os resultados obtidos nos testes realizados utilizando-se o algoritmo proposto neste trabalho, variando-se a quantidade de imagens usadas nos grupos de treinamento e de teste. A dimensão da matriz característica e a quantidade de neurônios em

todos os testes foram mantidas iguais. Foi utilizada em cada rede SOM a estrutura topológica linear de Kohonem com 10 neurônios cada.

Tabela 5.8 – Resultados obtidos nos testes realizados variando-se a quantidade de imagens utilizando-se o banco de dados Face94.

Nº de faces de Treinamento /indivíduo	Total de Faces para Treinamento	Total de Faces para Teste	Taxa de acerto (%)
2	300	2700	95,5926
3	450	2550	95,7255
4	600	2400	95,8750
5	750	2250	96,2222
6	900	2100	96,7619
7	1050	1950	96,8205
8	1200	1800	97,3333
9	1350	1650	98,0606
10	1500	1500	98,9300
11	1650	1350	98,8889
12	1800	1200	99,4167
13	1950	1050	99,5238
14	2100	900	99,4444
15	2250	750	99,4667
16	2400	600	100
17	2550	450	100
18	2700	300	100
19	2850	150	100

Pode-se observar na Tabela 5.8, que acima de 16 faces por indivíduo no treinamento, a taxa de reconhecimento do grupo de teste foi de 100%.

Como a base de dados da Face94 possui 3000 imagens não foi realizado o teste com o algoritmo *Leave-one-out*, pois teriam que ser realizado 3000 testes.

A base de dados Face94 apresentou ótimos resultados nos testes realizados, pois ela possui grande quantidade de imagens por individuo. Outros fatores que ajudaram a obter um

bom desempenho foi as imagens estarem bem focalizadas na face, e não existir muita variação em suas expressões em um mesmo indivíduo, como mostrado na Figura 5.16.

Figura 5.16 – Imagens das faces de um indivíduo do banco de dados Face94.

5.6 Conclusões

Neste capítulo foram apresentados os resultados obtidos do algoritmo de reconhecimento de faces proposto neste trabalho utilizando-se três bancos de imagens bem conhecidos na área de processamento de imagens: ORL, YaleA e Face94.

Nos testes realizados utilizando-se a base de dados ORL o algoritmo proposto nesta dissertação teve um bom desempenho. No primeiro teste onde foram escolhidas aleatoriamente 5 imagens de treinamento e 5 para testes de cada indivíduo, com um total 200

imagens para treinamento e 200 imagens para teste, a maior taxa de reconhecimento no grupo de teste foi 95% com um tempo de treinamento de 87,18 segundo.

Nos testes variando a quantidade de autovetores para a obtenção da matriz característica utilizando a base de dados ORL, a maior taxa de reconhecimento do grupo de teste utilizando o algoritmo proposto nesta dissertação foi de 93,5 % utilizando-se 9 e 10 neurônios.

A taxa de reconhecimento do algoritmo *Leave-one-out* usando a base de dados ORL, de 98,25 %, ou seja, das 400 faces apenas 7 faces foram identificadas incorretamente.

A base de dados YaleA não apresentou bons resultados nos testes realizados com o algoritmo proposto neste trabalho porque as faces desse banco não estão centralizadas no rosto dos indivíduos, e ocorre sombra no fundo de algumas imagens.

Para melhorar o desempenho do algoritmo proposto neste trabalho utilizando a base de dados YaleA é necessário realizar um pré-processamento nas imagens localizando as faces. Como o objetivo desta dissertação é a classificação/identificação de um sistema de reconhecimento de faces, não foi realizado o estágio de detecção automática da face.

Com os testes realizados utilizando-se a base de dados YaleA foi possível verificar a importância da fase de detecção de face em um sistema de reconhecimento de face. Diferente das faces desse banco, as faces das bases de dados ORL e Face94 estão bem centralizadas. Assim, para os bancos de dados ORL e Face94 não é necessário a etapa de detecção automática de face.

A base de dados Face94 apresentou ótimos resultados nos testes realizados, pois ela possui grande quantidade de imagens por individuo. Outros fatores que ajudaram a obter um bom desempenho foram as imagens estarem bem focalizadas na face, e não existir muita variação em suas expressões em um mesmo indivíduo. No primeiro teste onde foram

escolhidas aleatoriamente 10 imagens de treinamento e 10 para teste de cada indivíduo, a maior taxa de reconhecimento no grupo de teste foi 98,83 % para 10 e 32 neurônios em cada rede SOM com um tempo de treinamento de 29,97 e 101,96 segundos, respectivamente.

Nos testes realizados variando-se a quantidade de autovetores para obtenção da matriz característica utilizando-se a base de dados Face94, a maior taxa de reconhecimento do grupo de teste usando-se o algoritmo proposto neste trabalho foi de 98,93 % utilizando-se 10 neurônios em cada SOM.

Nos testes realizados neste capítulo, a taxa de reconhecimento da rede CSOM em relação a rede SOM foi melhor em todas as bases de dados utilizadas.

No próximo capítulo são apresentadas as conclusões e considerações finais desta dissertação e sugestões para trabalhos futuros.

Capítulo 6

CONCLUSÕES, CONTRIBUIÇÕES DESTE TRABALHO E TRABALHOS FUTUROS

6.1 Introdução

O objetivo principal deste trabalho foi desenvolver um algoritmo para o reconhecimento de faces em aplicações *one-to-many*. Neste algoritmo foram aplicadas nas faces técnicas de redução de dimensionalidade como *Two-Dimensional Discrete Cosine Transform* (2D-DCT) e *Two-Dimensional Principal Component Analysis* (2D-PCA) e utiliza-se a rede neural *Concurrent Self-Organizing Map* (CSOM) na classificação das faces.

Neste capítulo são apresentadas as conclusões e contribuições deste trabalho e os trabalhos futuros que poderão ser desenvolvidos a partir desta dissertação. Finalmente, são realizadas as considerações finais deste capítulo.

6.2 Conclusões

No algoritmo proposto neste trabalho foi realizada uma etapa de pré-processamento utilizando a 2D-DCT e a 2D-PCA com a finalidade de obter uma melhor extração de

características (informações) das imagens e aumentar com isso a eficiência computacional na etapa de classificação. A arquitetura utilizada na classificação foi o modelo CSOM implementado por Neagoe e Ropot [38].

Na fase de treinamento aplicou-se a 2D-DCT em cada imagem da base de treinamento. Com o objetivo de diminuir a quantidade de coeficientes para representar a matriz característica de cada face, armazenou-se em uma matriz quadrada apenas os coeficientes mais significativos que estavam localizados em torno da origem.

O motivo de utilizar a 2D-PCA no algoritmo ao invés do PCA antes do treinamento e da classificação da rede CSOM, foi porque a 2D-PCA é uma técnica simples em extração de característica/informações de imagens, apresenta melhor taxa de reconhecimento e é mais eficiente computacionalmente. Entretanto, a 2D-PCA não é tão eficiente em termos de armazenamento, pois são necessários mais coeficientes para a representação da imagem.

Devido a deficiência apresentada da 2D-PCA em termos de armazenamento, a matriz característica no algoritmo proposto neste trabalho foi obtida aplicando-se a 2D-DCT antes da 2D-PCA, reduzindo desse modo o espaço de armazenamento e os coeficientes de representação das imagens sem diminuir a taxa de reconhecimento.

A rede CSOM forneceu uma boa taxa de reconhecimento por treinar separadamente cada SOM, com um tempo de treinamento menor porque cada vetor característico de entrada é aplicado somente a um pequeno SOM especifico de sua classe. Outra característica é que o raio da vizinhança é menor no CSOM, se comparado com um único grande SOM. Baseado nesses conceitos a rede CSOM foi escolhida nesta dissertação para ser utilizada na etapa de classificação ao invés de uma única rede SOM.

Os testes com o algoritmo proposto neste trabalho foram realizados utilizando-se três bancos de imagens bem conhecidos na área de processamento de imagens: ORL, YaleA e Face94.

Nos testes utilizando a base de dados ORL e Face94 obteve-se um bom desempenho. A base de dados YaleA não apresentou bons resultados nos testes realizados com o algoritmo proposto neste trabalho porque as faces desse banco de dados não estão centralizadas no rosto dos indivíduos, e ocorre sombra no fundo de algumas imagens.

Para melhorar o desempenho do algoritmo proposto nesta dissertação utilizando-se a base de dados YaleA é necessário realizar um pré-processamento nas imagens localizando as faces. Como o objetivo deste trabalho é a classificação/identificação de um sistema de reconhecimento de faces, não foi realizado o estágio de detecção automática da face.

Com os testes realizados utilizando-se a base de dados YaleA foi possível verificar a importância da fase de detecção de face em um sistema de reconhecimento de face. Diferente das faces desse banco, as faces das bases de dados ORL e Face94 estão bem centralizadas. Assim, para esses bancos não é necessário a etapa de detecção automática de face.

Nos testes realizados variando-se a quantidade de neurônios pôde-se perceber que quantidade não é qualidade, pois acima de uma determinada quantidade a taxa de reconhecimento não aumentou, o mesmo ocorreu com relação a quantidade de autovalores utilizados para a obtenção da matriz característica de cada face.

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, principalmente na base de dados YaleA. Para evitar esses erros, é essencial que as etapas de

detecção e pré-processamento sejam eficientes, pois estas etapas influênciam diretamente na taxa de reconhecimento.

Assim, 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. Portanto, um estudo criterioso do espalhamento dos vetores de faces no espaço altamente dimensional pode ajudar a elaborar algoritmos mais eficientes.

6.3 Contribuições deste Trabalho

Nesta dissertação foi proposto um algoritmo de reconhecimento de faces, onde as principais contribuições foram:

- A proposta de um modelo para o reconhecimento de faces rápido e eficiente em termos de taxa de reconhecimento;

- Avaliação do desempenho do algoritmo em relação aos parâmetros da 2D-PCA e da rede CSOM;

- Estudo comparativo entre as redes SOM e CSOM; e.

- Pesquisa e avaliação de base de dados de faces utilizadas em trabalhos de reconhecimento automático de faces.

6.4 Trabalhos Futuros

Algumas sugestões de trabalhos que poderão ser desenvolvidos a partir deste trabalho são:

- Adicionar ao algoritmo proposto nesta dissertação algum método de detecção de faces e utilizá-los em um sistema de reconhecimento para controle de acesso de pessoas;

- Realizar testes com o algoritmo proposto neste trabalho utilizando-se outros bancos de faces;

- Utilizar o algoritmo proposto neste trabalho com outros métodos biométricos para implementar um sistema de autenticação e reconhecimento mais confiável;

- Realizar um modelamento matemático para definir os parâmetros ótimos de interação entre a 2D-DCT e 2D-PCA;

- Utilizar outros métodos de classificação como: *Learning Vector Quantisation* (LVQ) e lógica Fuzzy; e

- Utilizar alguma técnica de extração de características locais da face juntamente com o algoritmo proposto nesta dissertação, para obter um algoritmo mais robusto em reconhecimento de face.

6.5 Considerações Finais deste Capítulo

Este capítulo apresentou as conclusões e contribuições deste trabalho e as sugestões para futuros trabalhos que poderão ser desenvolvidos a partir deste trabalho.

REFERÊNCIAS

[1] Zhao, W., Chellappa, R., Philips, P. J., e Rosenfeld, A. Face recognition: a literature survey. **ACM Computing Surveys**, p. 399-458, 2003.

[2] Barrett, W. A. A survey of face recognition algorithms and testing results. **Systems and Computers**, p. 301-305, 1998.

[3] Chellapa, R., Wilson, C. L., e Sirohey, S. Human and machine recognition of faces: a survey. **Proceedings of the IEEE**, p. 705-740, 1995.

[4] Ellis, H. D. Introduction to aspects of face processing: Ten questions in need of answers. **Aspects of Face Processing**, 1986.

[5] Duda, R. O. e Hart, P. E. **Pattern classification and scene analysis**. USA: Wiley-Interscience, 1973.

[6] Jain, A. K., Duin, R. P. W., e Mao, J. Statical pattern recognition: A review. **IEEE Transactions On Pattern Analysis abd Machine Intelligence**, 2000.

[7] Kohonen, T. **Self-Organizing Maps**. 3ª Edição. ed. [S.l.]: Springer, 2001.

[8] Sirocich, L. e Kirby, M. Low-dimensional procedure for the characterization of human face. **J. Opt. Soc. Am.**, p. 519-524, 1987.

[9] Turk, M. e Pentland, A. Eigenfaces for recognition. **J. Cogn. Neurosci.**, p. 72-86, 1991a.

[10] Turk, M. e Pentland, A. Face recognition using eigenfaces. **In Proceedings of IEEE Computer Vision and Pattern Recognition**, p. 585-590, 1991b.

[11] Kirby, M. e Sirovich, L. Application of karhunen-loeve procedure for the characterization of human faces. **IEEE Trans. Patt. Anal. Mach. Intell.**, p. 12, 1990.

[12] Pentland, A. e Choudhury, T. Face recognition for smart enviroments. **IEEE Computer Society Press**, p. 50-55, 2000.

[13] Sulaiman, N., Yee, Y. M., Woei, G. C., e Sharmini. Pattern recognition - face recognition technology. **Artificial Interlligence (SAK4602) - Lecture**, 2000.

[14] Bartlett, M., S., Lades, H., M., e Sejnowski, T. Independent component representation for face recognition. **In Proceedings, SPIE Syposium on Eletronic Imaging: Science and Technology**, p. 528-539, 1998.

[15] Belhumeur, P. N., Hespanha, J. P., e Kriegman, D. J. Eigenfaces vs. fisherfaces:

recognition using class specific linear projection. **IEEE Transactions On Pattern Analysis And Machine Intelligence**, p. 19(7), 1997.

[16] Swets, D. L. e Weng, J. Discriminant analysis and eigenfaces partition tree for face and object recognition from views. **In Proceedings, International Conference on Automatic FAce and Gesture Recognition**, p. 192-197, 1996.

[17] Swets, D. L. e Weng, J. Using discriminant eigenfaces for image retrieval. **IEEE Trans. Patt. Anal. Mach. Intell**, p. 18:831-836, 1996.

[18] Grob, M. **Visual computing:** the integration of computer graphics, visual perception and imaging. Berlin: Springer-Verlag, 1994.

[19] Fiena, A. **Reconhecimento de Faces via PCA: Análise de Desempenho**. Dissertação de Mestrado. Uberlândia. 2006.

[20] Bartleet M., Movellan J., e Sejnowski T. Face recognition by independent component. **IEEE Trans. on Neural Networks**, p. 13(6):1450-1464, 2002.

[21] Conttrell, G. W. e Fleming, M. Face recognition using unsupervised feature extraction. **In Proc. Int. Neural Network Conf.**, p. 1:322-325, 1990.

[22] Bourland, H. e Kamp, Y. Auto-association by multilayer perceptrons and singular value decomposition. **Biological Cybern.**, p. 59:291-294, 1988.

[23] Lawrence, S., Guiles, C. L., Tsoi, A. C., e Back, A. D. Face Recognition: A Convolutional Neural Network Approach. **IEEE Trans. Neural Netw.**, p. 8:98-113, 1997.

[24] Sato, Y., Yoda, I., e Sakaue, K. Automatic face classification by self-organization for face recognition. **Proceedings of the IEEE International Workshop on Analysis and Modeling of Faces and Gestures**, p. 162-172, 2003.

[25] Spies, H. e Ricketts, I. Face recognition in fourrier space. **Vision Interface 2000: Montreal**, p. 38-44, 2000.

[26] Podilchuk, C. e Zhang, X. Face recognition using dct-based feature vectors. **ICASSP96**, p. 2144-2147, 1996.

[27] Liposcak, Z. e Loncaric, S. Face recognition from profiles using morphological. **RATFG99**, 1999. 47-52.

[28] Kohonen, T. **Self-organization and associative memory**. New York: Springer-Verlag, 1988.

[29] Churchland, P. S. e Sejnowski, T. J. The computational brain. **Mit Press**, Cambrige, MA., 1992.

[30] Haykin, S. **Neural networks:** a comprehensive foundation. New York: Prentice Hall, 1998.

[31] Kohonen, T. Exploration of very large databases by self-organizing maps. **Proc. of the1997 IEEE Intl. Conf. on Neural Networks**, Houston, Texas, 1997. 1-6.

[32] Lampinen, J. e Oja, E. Clustering Properties of Hierarchical Self-Organizing Maps. **Journal of Mathematical Imaging and Vision**, v. 2, p. 261-272, 1992.

[33] Kohonen, T. Analysis of a simple self-organizing process. **Biological Cybernetics**, v. 44, n. 2, p. 135-140, 1982.

[34] Kohonen, T. Self-organizing formation of topologically correct feature maps. **Biologica lCybernetics**, v. 43, n. 1, p. 59-69, 1982.

[35] Bezdek J.C., Tsao E.C. e Pal, N.R. Fuzzy Kohonen clustering networks. **In Proc.IEEE Int. Conf on Fuzzy Systems**, p. 1035 - 1043 , 1992.

[36] Pedrini, H. e Schwartz, W. R. Análise de Imagens Digitais**:** Princípios, Algoritmos e Aplicações. 1ª. ed. [S.l.]: PIONEIRA THOMSON, 2008.

[37] Yang J. , Zhang, D., FrangI, A.F. Two-Dimensional PCA: A New Approach to Appearance-Based Face Representation and Recognition. **IEEE Transaction on Pattern Analysis and Machine Intelligence**, 26, n. 1, 2004. 131 - 137.

[38] Neagoe, V. e Ropot. Concurrent Selg-Organizing Maps for Pattern Classification. **Proccedings of the First IEEE Internationa Conference on Cognitive Informatic**, 2002. 304-312.

[39] Guimarães, T. S., Flôres, E. L., Carrijo, G. A., Veiga A. C. P. E Barcelos, C. A. Z. Two-Dimensional Principal Component Analysis and Concurrent Self-Organizing Maps for Face Classification. **International Conference on Systems, Signals and Image Processing**, Rio de Janeiro, n. 17, 2010.

[40] ATT-FACE DATABASE. Disponível em: <http://www.uk.research.att.com/facedatabase.html>.

[41] Vesanato J., Himberg J., AlhoniemI E. e Parhankagas J. SOM Toolbox for Matlab. Disponível em: <http://www.cis.hut.fi/projects/somtoolbox/>. Acesso em: jan. 2009.

[42] YALEA-FACE DATABASE, 1997. Disponível em: <http://cvc.yale.edu/projects/yalefaces/yalefaces.html>.

[43] Spacek, L. **Face Recognition Data**. Disponível em: <http://cswww.essex.ac.uk/mv/allfaces/index.html>. Acesso em: jan. 2010.

[44] Kohonen, T. An introduction to neural computing. **Neural Networks**, p. 3-16, 1988.

[45] Zhao, W., Chellappa, R., e Kris. Discriminant analysis of principal components for face recognition. **In PRoceedings, Internationa Conference on Automatica Face and Gesture Recognition**, 1998b. 336-341.

[46] Cottrell G. W., e Fleming M. K. Categorization of Faces Using Unsupervised Feature Extraction. **In Proc. Int. Neural Network Conf.**, p. 322-325, 1990.

[47] Bezdek, J.C., Keller, J., Krisnapuram, R., PAL, N.R. **Fuzzy Models and Algorithms for Pattern Recognition and Image Processing**. New York: Springer, v. 4, 1999.

[48] Kohonen, T. The Self-Organizing Map. **Proceedings IEEE**, 78, n. 9, 1990. 1464 - 1479.

[49] Chelapa R., Wilson, C. e Sirohey, S. Human and Machine Recognition of Faces: A survey. **Proceedings IEEE**, 83, 1995. 705 - 740.

[50] Sirovich, L. e Kirby M. Low-Dimensional Procedure for Characterization of Human Faces. **J. Optical Soc. Am.**, 4, 1987. 519 - 524.

[51] Turk, M. e Pentland Face recognition Using Eigenfaces. **Proc. IEEE Conf. on Computer Vision and Pattern Recognition**, 1991. 586-591.

[52] Turk, M. e Pentland d. Eigenfaces for Recognition. **J. Cognitive Neuroscience**, 3, n. 1, 1991. 71 - 86.

[53] M. Kirby E L. Sirovich. Application of the KL Procedure for the Characterization of Human Faces. **IEEE Trans. Pattern Analysis and Machine Interlligence**, 12, 1990. 103 - 108.

[54] Moataz M. e Wasfy B. A New Fast Facial Recognition Algorithm Applicable to Large Database, 2006.

[55] Neagoe, V. e Ropot. Concurrent Selg-Organizing Maps for Pattern Classification. **Proccedings of the First IEEE Internationa Conference on Cognitive Informatic**, 2002.