

CENTRO FEDERAL DE EDUCAÇÃO TECNOLÓGICA DO PARANÁ

Curso de Pós-Graduação em Engenharia Elétrica e Informática Industrial

DISSERTAÇÃO

apresentada ao CEFET-PR

para obtenção do título de

MESTRE EM CIÊNCIAS

por

HUGO VIEIRA NETO

RECONHECIMENTO AUTOMÁTICO DE IMPRESSÕES DIGITAIS

UTILIZANDO WAVELETS E REDES NEURONAIS ARTIFICIAIS

Banca Examinadora:

Presidente e Orientador:

Prof. Dr. DÍBIO LEANDRO BORGES UFG

Co-orientadora:

Prof^a. Dr^a. LÚCIA VALÉRIA RAMOS DE ARRUDA CEFET-PR

Examinadores:

Prof. Dr. JOÃO FERNANDO MARAR UNESP-Bauru

Prof. Dr. JACQUES FACON PUC-PR

Curitiba, 30 de abril de 1998.

RESUMO:

Este trabalho propõe uma nova abordagem para o reconhecimento automático de impressões digitais. Diferentemente das técnicas tradicionais, baseadas na análise de peculiaridades relativas aos padrões de impressão digital e em algoritmos de busca dedicados especialmente a esse fim, é sugerida uma metodologia com base em ferramentas de representação de sinais e em modelos de classificação conexionistas.

Basicamente, a metodologia aqui proposta fundamenta-se em duas técnicas principais: o padrão de compressão *wavelet* para impressões digitais especificado pelo FBI como método de extração de características; e modelos de Redes Neurais Artificiais que utilizam técnicas lineares de treinamento como método de classificação dos padrões obtidos.

O principal objetivo do método sugerido é o desenvolvimento de um sistema de reconhecimento independente de técnicas específicas de análise e processamento de imagens de impressões digitais, procurando-se a obtenção de baixos índices de falsa aceitação e rejeição, sendo a prioridade para o índice de falsa aceitação.

As metodologias e resultados obtidos nos experimentos executados são apresentados, bem como suas respectivas análises e ainda algumas propostas para trabalhos futuros e melhorias. Alguns aspectos pertinentes ao *hardware* de aquisição de imagens e algumas técnicas de processamento digital de imagens também são apresentadas.

PALAVRAS-CHAVE:

Processamento Digital de Imagens / Reconhecimento de Padrões / Redes Neurais Artificiais

ÁREA/SUB-ÁREA DE CONHECIMENTO:

Sistemas de Computação

HUGO VIEIRA NETO

**RECONHECIMENTO AUTOMÁTICO DE IMPRESSÕES DIGITAIS
UTILIZANDO WAVELETS E REDES NEURONAIAS ARTIFICIAIS**

CURITIBA

1998

HUGO VIEIRA NETO

**RECONHECIMENTO AUTOMÁTICO DE IMPRESSÕES DIGITAIS
UTILIZANDO *WAVELETS* E REDES NEURONAIS ARTIFICIAIS**

Dissertação apresentada ao Curso de Pós-Graduação em Engenharia Elétrica e Informática Industrial do Centro Federal de Educação Tecnológica do Paraná como requisito parcial para obtenção do título de “Mestre em Ciências” - Área de Concentração: Informática Industrial.

Orientador: Prof. Dr. Dívio Leandro Borges
Co-orientadora: Profa. Dra. Lúcia Valéria Ramos de Arruda

CURITIBA

1998

There could be no honour in a sure success, but much might be wrested from a sure defeat.

Biggles McFly

Este trabalho é dedicado a todos aqueles que buscam o real desenvolvimento científico e tecnológico, abandonando a vaidade e os interesses pessoais em favor da evolução humana.

AGRADECIMENTOS

Primeiramente agradeço à minha família pelo apoio incondicional à realização deste trabalho. Em especial, agradeço à minha mãe, Beatriz, meu irmão, Rafael, e meus avós, Hugo, Maria da Luz e Alice, os quais têm participado ativamente de todos os momentos de minha vida. Agradeço também aos tios Luiz e Dirce e aos primos Luiz, João Fernandes, Sandra, Marco Aurélio e Cleunice por tudo que sempre fizeram por mim, apesar da distância, apenas física, que nos separa.

A Ogê e Fabiana Marques pelas muitas coisas que me ensinaram, talvez até sem ter tomado conhecimento, e por tudo que me ofereceram, inclusive uma grande e verdadeira amizade como poucas existentes. Igualmente agradeço ao amigo Ryon Cássio Braga pelos momentos de reflexão e estudo sobre a natureza da inteligência humana.

Aos colegas, professores e funcionários do Departamento de Informática e do Curso de Pós-Graduação em Informática Industrial do CEFET-PR, pelas preciosas amostras de suas impressões digitais, sem as quais não seria possível a realização desse trabalho, com agradecimentos especiais aos colegas Andreia Malucelli e José Carlos Ferreira da Rocha, por terem ajudado a trazer a grande maioria desses voluntários. Agradeço também ao colega Bruno Marchesi e ao professor Álvaro Luiz Stelle pelos materiais e pelas discussões sobre *wavelets*, e aos professores Dívio Leandro Borges e Lúcia Valéria Ramos de Arruda pelas constantes ajudas com redes neurais artificiais e revisões dos textos escritos.

Meus sinceros agradecimentos também ao professor João Fernando Marar (UNESP - Bauru) pelas críticas construtivas e sugestões para melhoria da qualidade do presente trabalho, quando da sua apresentação no IV Simpósio Brasileiro de Redes Neurais.

Finalmente agradeço às agências de fomento à pesquisa, CAPES e CNPq, pela bolsa de estudo concedida, sem a qual não seria possível o desenvolvimento deste trabalho.

SUMÁRIO

LISTA DE FIGURAS	vi
LISTA DE TABELAS.....	vii
RESUMO	ix
ABSTRACT	x
1 INTRODUÇÃO.....	1
2 CONSTRUÇÃO DO DISPOSITIVO DE AQUISIÇÃO.....	7
2.1 Conceitos Básicos	7
2.2 Dispositivo de Aquisição de Imagens.....	8
2.3 Base de Dados.....	12
3 PRÉ-PROCESSAMENTO E SEGMENTAÇÃO	15
3.1 Conceitos de Processamento Digital de Imagens.....	15
3.2 Pré-processamento das Imagens	21
3.3 Segmentação das Imagens.....	27
4 EXTRAÇÃO DE CARACTERÍSTICAS.....	35
4.1 Transformada <i>Wavelet</i>	36
4.2 Padrão WSQ de Compressão de Imagens de Impressões Digitais	43
4.3 Extração de Características das Imagens.....	46
5 CLASSIFICAÇÃO DOS PADRÕES DE IMPRESSÕES DIGITAIS.....	55
5.1 Perceptron Multicamadas.....	56
5.2 Redes com Funções Radiais de Base.....	60
5.3 Classificação dos Padrões.....	64

6 CONCLUSÕES	75
6.1 Sugestões para Trabalhos Futuros	77
APÊNDICE A	79
Primeira Base de Dados	79
Segunda Base de Dados	79
REFERÊNCIAS BIBLIOGRÁFICAS	82

LISTA DE FIGURAS

1	Figura 1.1 - Exemplos de minúcias	2
2	Figura 1.2 - Diagrama em blocos do sistema de reconhecimento de impressões digitais	6
3	Figura 2.1 - Dispositivo de aquisição de imagens de impressões digitais.....	9
4	Figura 2.2 - Aspecto final do dispositivo de aquisição.....	11
5	Figura 2.3 - Comparação das imagens obtidas	11
6	Figura 2.4 - Imagens obtidas	15
7	Figura 3.1 - Redimensionamento das imagens.....	22
8	Figura 3.2 - Aproveitamento das imagens.....	23
9	Figura 3.3 - Reconstrução das imagens após compactação pelo método WSQ	24
10	Figura 3.4 - Realce das imagens.....	25
11	Figura 3.5 - Equalização de histograma	26
12	Figura 3.6 - Diagrama em blocos da etapa de pré-processamento.....	27
13	Figura 3.7 - Limiarização das imagens	28
14	Figura 3.8 - Operações morfológicas	29
15	Figura 3.9 - Determinação da região de interesse.....	30
16	Figura 3.10 - Segmentação por máxima correlação.....	31
17	Figura 3.11 - Resultados obtidos com a segmentação	32
18	Figura 3.12 - Diagrama em blocos da etapa de segmentação	32
19	Figura 4.1 - Estrutura de sub-bandas	44
20	Figura 4.2 - <i>Wavelet</i> mãe originada pelo filtro de primeira geração	45
21	Figura 4.3 - Exemplos de condição de contorno	46
22	Figura 4.4 - <i>Wavelet</i> mãe originada pelo filtro <i>Symmlet8</i>	50
23	Figura 4.5 - <i>Wavelet</i> mãe originada pelo filtro H10.....	51
24	Figura 4.6 - <i>Wavelet</i> mãe originada pelo filtro S10	51
25	Figura 5.1 - Exemplo de RNA	55
26	Figura 5.2 - Erro médio quadrático durante o treinamento.....	65

LISTA DE TABELAS

1 Tabela 4.1 - Coeficientes do filtro de primeira geração	45
2 Tabela 4.2 - Coeficientes dos filtros associados às bases <i>wavelet</i> utilizadas	50
3 Tabela 4.3 - Resultados da análise estatística dos vetores de características obtidos com a base <i>Symmlet8</i>	52
4 Tabela 4.4 - Resultados da análise estatística dos vetores de características obtidos com a base do FBI para a primeira geração de codificadores WSQ ...	52
5 Tabela 4.5 - Resultados da análise estatística dos vetores de características obtidos com a base H10.....	52
6 Tabela 4.6 - Resultados da análise estatística dos vetores de características obtidos com a base S10.....	53
7 Tabela 4.7 - Resultados dos critérios de avaliação do desempenho das bases <i>wavelet</i> na extração de características	54
8 Tabela 5.1 - Erros médios quadráticos resultantes dos testes intermediários...	65
9 Tabela 5.2 - Índices de falsa rejeição entre membros das classes de treinamento.....	68
10 Tabela 5.3 - Índices de falsa aceitação entre membros das classes de treinamento.....	68
11 Tabela 5.4 - Índices de múltipla classificação entre membros das classes de treinamento.....	68
12 Tabela 5.5 - Índices de falsa aceitação entre membros das classes de teste (treinamento com 7 amostras).....	68
13 Tabela 5.6 - Índices de falsa aceitação entre membros das classes de teste (treinamento com 10 amostras).....	68
14 Tabela 5.7 - Índices de falsa rejeição entre membros das classes de treinamento - seleção progressiva.....	70
15 Tabela 5.8 - Índices de falsa aceitação entre membros das classes de treinamento - seleção progressiva.....	70

16 Tabela 5.9 - Índices de múltipla classificação entre membros das classes de treinamento - seleção progressiva.....	70
17 Tabela 5.10 - Índices de falsa aceitação entre membros das classes de teste (treinamento com 7 amostras) - seleção progressiva.....	70
18 Tabela 5.11 - Índices de falsa aceitação entre membros das classes de teste (treinamento com 10 amostras) - seleção progressiva.....	70
19 Tabela 5.12 - Índices de falsa rejeição entre membros das classes de treinamento - regressão global	71
20 Tabela 5.13 - Índices de falsa aceitação entre membros das classes de treinamento - regressão global	71
21 Tabela 5.14 - Índices de múltipla classificação entre membros das classes de treinamento - regressão global	71
22 Tabela 5.15 - Índices de falsa aceitação entre membros das classes de teste (treinamento com 7 amostras) - regressão global.....	72
23 Tabela 5.16 - Índices de falsa aceitação entre membros das classes de teste (treinamento com 10 amostras) - regressão global.....	72
24 Tabela 5.17 - Índices de falsa rejeição e falsa aceitação globais médios para 64 coeficientes	73
25 Tabela 5.18 - Índices de falsa rejeição e falsa aceitação globais médios para 32 coeficientes	73
26 Tabela 5.19 - Índices de falsa rejeição e falsa aceitação globais médios para 16 coeficientes	74
27 Tabela 6.1 - Tempo aproximado de processamento das diversas etapas do sistema.....	75

RESUMO

Este trabalho propõe uma nova abordagem para o reconhecimento automático de impressões digitais. Diferentemente das técnicas tradicionais, baseadas na análise de peculiaridades relativas aos padrões de impressão digital e em algoritmos de busca dedicados especialmente a esse fim, é sugerida uma metodologia com base em ferramentas de representação de sinais e em modelos de classificação conexionistas.

Basicamente, a metodologia aqui proposta fundamenta-se em duas técnicas principais: o padrão de compressão *wavelet* para impressões digitais especificado pelo FBI como método de extração de características; e modelos de Redes Neurais Artificiais que utilizam técnicas lineares de treinamento como método de classificação dos padrões obtidos.

O principal objetivo do método sugerido é o desenvolvimento de um sistema de reconhecimento independente de técnicas específicas de análise e processamento de imagens de impressões digitais, procurando-se a obtenção de baixos índices de falsa aceitação e rejeição, sendo a prioridade para o índice de falsa aceitação.

As metodologias e resultados obtidos nos experimentos executados são apresentados, bem como suas respectivas análises e ainda algumas propostas para trabalhos futuros e melhorias. Alguns aspectos pertinentes ao *hardware* de aquisição de imagens e algumas técnicas de processamento digital de imagens também são apresentadas.

ABSTRACT

This work proposes a new approach for automatic fingerprint recognition. Differently from traditional techniques, based on the analysis of peculiarities existing in fingerprint patterns and dedicated search algorithms for that purpose, it is suggested a methodology centred in signal representation tools and connectionist classification models.

Basically, the methodology proposed here supports itself in two main techniques: the FBI's Wavelet Scalar Quantization standard for fingerprint image compression as method for the feature extraction; and Artificial Neural Networks models that use linear training techniques as method for the classification of the resulting patterns.

The main objective of the suggested method is the development of a recognition system independent from specific fingerprint image analysis techniques, looking for low false acceptance and false rejection rates, with priority for the false acceptance rate.

The methodologies and results achieved in the executed experiments are presented, as well as their respective analysis and yet some proposals for future work and enhancements. Some aspects pertinent to the image acquisition hardware and some digital image processing techniques are also presented.

CAPÍTULO 1

INTRODUÇÃO

Uma das principais aplicações da área de reconhecimento automático de padrões encontra-se na verificação biométrica de identidade. Atualmente há um interesse crescente na tecnologia de verificação automática de identidade para as mais diversas aplicações, onde é necessária a identificação ou autenticação pessoal. Alguns exemplos que podem ser dados abrangem o acesso lógico a sistemas bancários, comércio eletrônico e correio eletrônico privativo na Internet. Esses são casos de aplicações em rede, onde o sistema deve certificar-se da identidade do usuário antes de proceder a transação. Outras aplicações residem no controle de acesso físico a ambientes restritos, sistemas de verificação de ponto de funcionários em empresas, entre outras possibilidades.

Apesar de existirem diversas maneiras de se identificar uma pessoa, o uso de verificação biométrica é altamente desejável. Os dados biométricos oferecem uma vantagem impossível de ser obtida com senhas e cartões, os quais podem ser esquecidos, perdidos ou, ainda pior, utilizados por pessoas não autorizadas. Existem diversas características biométricas que possibilitam a identificação. Entre elas podemos destacar os padrões de impressões digitais, padrões de retina, padrões de íris, padrões de voz, padrões genéticos, a forma das mãos, características faciais, características sanguíneas e características dos cabelos [Setlak 1996].

Dentre os métodos disponíveis, geralmente os mais bem aceitos e acessíveis são os que fazem uso de verificação de impressões digitais. A principal justificativa para esse fato é a unicidade das impressões digitais de cada indivíduo, característica conhecida e utilizada para a identificação pessoal perante a lei. Outra grande vantagem da verificação biométrica de identidade através de impressões digitais é o fato de serem invariantes com a idade e com o estado de saúde dos indivíduos [Baldi 1993]. Além disso, imagens de impressões digitais podem ser adquiridas por sistemas automáticos de maneira não invasiva, de forma relativamente simples e eficiente.

No entanto, o reconhecimento de impressões digitais constitui um problema clássico na área de reconhecimento de padrões, para o qual ainda não se considera ter sido encontrada uma solução completa em termos de confiabilidade e velocidade de resposta. Trata-se de um caso de classificação não-linear onde está presente um grande conjunto de padrões com características bastante específicas.

Algumas das abordagens tradicionais utilizadas para se tentar resolver esse problema fazem uso de características peculiares presentes nas imagens de impressões digitais que as tornam únicas. Trata-se de métodos semelhantes aos empregados por especialistas ao compararem visualmente um par de impressões digitais. Tais métodos têm como base a orientação dos sulcos e a localização das minúcias existentes nas imagens sob análise [Ratha 1995, Karu 1996].



Figura 1.1 - Exemplos de minúcias.

Minúcias são tipos específicos de descontinuidades nos padrões dos sulcos das impressões digitais, tais como bifurcações, terminações e ilhas, como ilustra a Figura 1.1. A quantidade de minúcias existentes numa impressão digital é da ordem de 50 a

150, porém tecnicamente a identificação de 10 delas é tida como suficiente para estabelecer a identidade do indivíduo [Baldi 1993]. A orientação dos padrões de sulcos, por outro lado, tem como principal objetivo classificar as impressões digitais dentro de cinco principais grandes grupos: 1) arco (*arch*); 2) arco em tenda (*tented arch*); 3) presilha externa (*left loop*); 4) presilha interna (*right loop*); e 5) verticilo (*whorl*) [NIST 1993], cujo detalhamento foge ao escopo deste trabalho, tendo em vista o objetivo de reconhecer padrões de impressões digitais e não de classificá-los em grupos característicos.

Infelizmente, abordagens baseadas na análise de minúcias são bastante sensíveis a ruído e deformações. Possíveis perturbações no processo de aquisição das imagens podem facilmente criar minúcias artificiais ou fazer minúcias existentes desaparecerem [Baldi 1993]. Essa particularidade estimulou o desenvolvimento de algoritmos especiais para melhoramento e purificação de imagens de impressões digitais [Hung 1993]. Outras desvantagens desse tipo de abordagem são a complexidade e o grande esforço computacional envolvido, por se tratar essencialmente de um problema de comparação de grafos montados a partir das coordenadas relativas das minúcias.

O objetivo deste trabalho é apresentar uma nova proposta para o reconhecimento de padrões de impressões digitais, independente da análise das minúcias presentes nas imagens. Tal abordagem visa a obtenção de um método de extração de características fundamentado na análise de sinais através de ferramentas matemáticas, que possa ser utilizado com algoritmos de Redes Neurais Artificiais (RNA), buscando simplificar e obter um alto desempenho para o sistema de reconhecimento.

Existem diversas razões para se acreditar na adequação de algoritmos de RNA para uma possível solução do problema de reconhecimento de impressões digitais. Tais algoritmos são capazes de realizar classificações não-lineares com eficiência, têm se mostrado altamente robustos e oferecem grande adaptabilidade às condições de diversos casos especiais [Rumelhart 1994]. Um aspecto particularmente importante é o fato das RNA serem treinadas através de exemplos, fato que possibilita a minimização do efeito das deformações e da incidência de ruído nas imagens, quando coletadas pelo dispositivo de aquisição, conforme mencionado anteriormente. Além disso, o conjunto das imagens de impressões digitais possui características estatísticas muito peculiares, onde o correspondente problema de reconhecimento é melhor delineado e confinado, quando comparado a outros problemas como o reconhecimento de caracteres manuscritos, para

o qual algoritmos de RNA foram aplicados com razoável sucesso [Le Cun 1990]. Por fim, as RNA podem ser treinadas de maneiras diferentes e ter seus parâmetros otimizados para atender aos requisitos particulares de aplicações específicas.

A aplicação de RNA para o caso específico de reconhecimento de impressões digitais é feita em [Baldi 1993]. Em [Blue 1994] é feita uma comparação entre classificadores estatísticos convencionais e classificadores por RNA com aplicação em imagens de impressões digitais, onde são apresentados alguns resultados promissores. Em [NIST 1992] também são apresentados resultados de experimentos com um sistema de classificação de impressões digitais utilizando RNA.

Os modelos de RNA contêm normalmente vários elementos básicos de processamento conectados entre si, exigindo comumente um grande poder de processamento para grandes quantidades de elementos, inviabilizando dessa maneira o uso direto de imagens como entrada. Portanto, existe a necessidade de se obter formas alternativas de representação dos padrões de entrada que preservem as características de interesse originais das imagens de maneira compacta e coerente.

A determinação de uma metodologia de extração de características relevantes de um padrão não é uma tarefa trivial. Para o caso particular de extração de características de imagens de impressões digitais, diversos métodos foram desenvolvidos, em sua maioria baseados nas características especiais dos padrões de impressão digital, tais como detecção de minúcias e orientação dos sulcos [Ratha 1995, Karu 1996], anteriormente citados. A transformada Karhunen-Loève do padrão de orientação dos sulcos de impressões digitais foi utilizada em [Blue 1994] para obter vetores de características, os quais foram usados como entrada para classificadores estatísticos e neuronais.

Uma dificuldade que surge nas abordagens tradicionais é o emprego de algoritmos especializados para extrair informações de orientação e minúcias [Ratha 1995, Karu 1996]. Por outro lado, o uso de transformadas matemáticas para mapear os padrões em um espaço de representação mais adequado é uma possibilidade conhecida. O uso da transformada *wavelet* discreta como técnica de extração de características foi considerado nesse trabalho por diversos motivos. Um deles foi o fato do FBI ter atingido resultados interessantes na compressão de imagens de impressões digitais, utilizando uma especificação particular de transformada *wavelet* [FBI 1993]. De fato, transformadas *wavelet* adequadas permitem muito boa reconstrução do sinal original,

enquanto propiciam boa localização tempo-frequência em múltiplas escalas [Graps 1995, Vaidyanathan 1995]. Além disso, a compressão de imagens através de transformadas *wavelet* apropriadas mostrou que é possível uma representação fiel das discontinuidades dos padrões de impressão digital (minúcias) [Sherlock 1996a, Sherlock 1996b]. Finalmente, algoritmos de transformadas *wavelet* discretas são mais facilmente implementados em processadores digitais de sinais ou *hardware* específico, pois consistem em uma técnica puramente matemática de análise de sinais, ao contrário dos algoritmos orientados a sulcos e minúcias, os quais são altamente específicos e muitas vezes fundamentados na comparação de grafos [Baldi 1993].

Com base no que foi exposto anteriormente, o objetivo principal desse trabalho consiste na obtenção de um sistema capaz de indicar se determinado padrão de impressão digital apresentado na sua entrada faz parte de alguma classe da base de dados utilizada para o seu treinamento. Sendo assim, é meta a obtenção de um método de reconhecimento de impressões digitais cujos índices de falsa aceitação e falsa rejeição sejam os mais baixos possíveis, apresentando ao mesmo tempo uma sensibilidade pequena aos problemas advindos de deformações e ruído nas imagens. Para isso é necessária a obtenção de uma representação adequada dos padrões de entrada, aliada a um compromisso entre precisão e generalização na etapa de classificação.

Para a viabilização desse desenvolvimento foi necessário dispor de várias imagens de impressões digitais, pois o treinamento das RNA exige que estejam disponíveis diversas versões da mesma impressão digital, tomadas em instantes e condições diferentes. Portanto, foi formada uma base de dados utilizando-se um dispositivo de aquisição de imagens de impressão digital especialmente construído para esse fim.

Depois de construída a base de dados de treinamento e validação, foi necessário o estudo de etapas de pré-processamento e segmentação (seleção da região de interesse nas imagens) para que as imagens pudessem ser convenientemente utilizadas nos processos de extração de características e classificação dos padrões obtidos. O diagrama em blocos simplificado do sistema de reconhecimento de impressões digitais é mostrado na Figura 1.2.

No Capítulo 2 são abordados os aspectos relativos à aquisição e digitalização das imagens utilizadas neste trabalho, abrangendo tópicos relacionados à especificação e construção do dispositivo de aquisição e algumas das técnicas existentes.

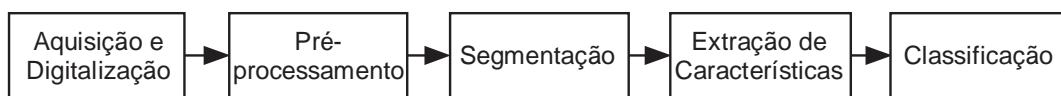


Figura 1.2 - Diagrama em blocos do sistema de reconhecimento de impressões digitais.

O Capítulo 3 descreve os detalhes pertinentes às etapas de pré-processamento e segmentação das imagens adquiridas. Para a realização dessas etapas foi utilizado como plataforma de *software* o aplicativo MatLab®* e sua *toolbox* de processamento de imagens. São enfocados nesse capítulo alguns aspectos e técnicas do processamento digital de imagens pertinentes ao trabalho desenvolvido.

A extração de características das imagens resultantes das etapas de pré-processamento e segmentação são discutidos e analisados no Capítulo 4. Aspectos teóricos e práticos da transformada *wavelet*, bem como do padrão de compressão de imagens de impressões digitais do *Federal Bureau of Investigation* (FBI), também são abordados nesse capítulo. Novamente o MatLab® foi utilizado como plataforma de *software*, contando com o auxílio da *toolbox* WaveLab [Buckheit 1995].

Quanto aos aspectos relativos à classificação e reconhecimento dos padrões obtidos na etapa de extração de características, estes são tratados no Capítulo 5. Nesse capítulo também é realizada uma análise dos modelos de RNA adequados à resolução do problema. A plataforma de *software* utilizada nesse estágio foi novamente o MatLab®, contando com o auxílio de rotinas de otimização de RNA que fazem uso de técnicas lineares de treinamento [Orr 1997].

Finalmente, o Capítulo 6 apresenta as conclusões a respeito do método proposto e o seu desempenho dentro dos experimentos realizados, bem como propostas para trabalhos futuros.

* MatLab é marca registrada de MathWorks, Inc.

CAPÍTULO 2

CONSTRUÇÃO DO DISPOSITIVO DE AQUISIÇÃO

Um sistema de visão artificial genérico possui como primeiro elemento a etapa de aquisição e digitalização de imagens. Esta etapa tem como função converter uma imagem em uma representação numérica adequada para o processamento digital, e compreende dois elementos principais: O primeiro consiste em um sensor ou dispositivo sensível a uma determinada faixa de energia no espectro eletromagnético (o espectro de luz visível, por exemplo), capaz de produzir em sua saída um sinal elétrico de amplitude proporcional ao nível de energia detectado. O segundo é constituído de um digitalizador (ou quantizador), responsável pela conversão do sinal elétrico analógico em informação digital, possível de ser interpretada por microprocessadores.

Alguns dos aspectos de projeto envolvidos nesta etapa que podem ser mencionados são: a escolha do tipo de sensor, o conjunto óptico a ser utilizado, as condições de iluminação, os requisitos de velocidade de aquisição, resolução e número de níveis de cinza da imagem digitalizada, entre outros.

Primeiramente serão abordados alguns aspectos teóricos quanto à aquisição e digitalização de imagens, para posteriormente serem apresentados detalhes sobre o método empregado neste trabalho.

2.1 Conceitos Básicos

Aquisição

Existem diversas alternativas para conversão de imagens em sinais elétricos utilizando dispositivos de estado sólido. Entre os principais sensores encontram-se as matrizes de fotodiodos, os dispositivos de injeção de carga (CID) e os dispositivos de carga acoplada (CCD), sendo este último o tipo mais comum de sensor utilizado atualmente [Schalkoff 1989].

A tecnologia CCD baseia-se no uso de pastilhas semicondutoras com um determinado número de elementos (recipientes) capazes de armazenar carga elétrica, que possuem um mecanismo de transferência entre eles. A quantidade de carga elétrica

armazenada nesses recipientes corresponde a valores analógicos, o que significa que erros de quantização são praticamente nulos. Dessa maneira, o CCD equivale a uma memória analógica, cujos dados são acessados serialmente através da transferência de carga entre os seus elementos de armazenamento.

A utilização de materiais fotossensíveis permite a construção de dispositivos CCD cujos recipientes de carga constituem também elementos sensíveis à luz visível. Assim, a carga armazenada em cada recipiente é proporcional à intensidade luminosa incidente sobre o mesmo. Posteriormente, a varredura seqüencial dos elementos pode gerar um sinal elétrico variante no tempo, responsável pelo mapeamento espacial da imagem. Em outras palavras, o valor do sinal elétrico obtido em determinado instante de tempo corresponde a um ponto específico da superfície fotossensível do CCD.

Normalmente, a formação de imagens envolve a utilização de sensores CCD de área, que consistem em matrizes bidimensionais de elementos fotossensíveis. Sensores de área são amplamente utilizados em câmeras de vídeo digitais e contribuem para o reduzido tamanho desses equipamentos [Schalkoff 1989].

Digitalização

O sinal analógico de vídeo obtido a partir do dispositivo sensor deve ser submetido a uma discretização espacial e em amplitude para assumir o formato adequado ao processamento digital. Denomina-se amostragem o processo de discretização espacial e quantização o processo de discretização em amplitude.

A amostragem converte a imagem analógica em uma matriz de M por N pontos, denominados *pixels* (elementos de imagem) e a quantização atribui a cada *pixel* um valor inteiro, na faixa de 0 a 2^n-1 . Do ponto de vista eletrônico, a digitalização consiste em uma conversão analógico/digital na qual o número de amostras por unidade de tempo indica a resolução espacial e o número de bits (n) do conversor analógico/digital utilizado determina o número de níveis de cinza resultantes na imagem digitalizada. Quanto maior o valor de n , maior o número de níveis de cinza presentes na imagem digitalizada.

2.2 Dispositivo de Aquisição de Imagens

Existem diversas maneiras de se adquirir imagens de impressões digitais, algumas delas já bastante consagradas. A técnica mais comumente utilizada, por ser a mais simples de

ser implementada, consiste na digitalização de imagens obtidas através da impressão com tinta em papel. A aquisição desse tipo de imagem pode ser feita utilizando-se câmeras de vídeo digitais ou *scanners* de documentos. No entanto, esse processo de obtenção de imagens se mostra lento e inconveniente, pois existe a possibilidade de ocorrência de erros durante a fase de impressão com tinta, tais como borrões e manchas, que normalmente serão detectados somente mais tarde, quando a imagem for digitalizada. Obviamente, o uso dessa técnica de aquisição de imagens inviabiliza o processamento em tempo real do sistema de reconhecimento ou classificação das impressões digitais.

Um outro método bastante utilizado em dispositivos comerciais de aquisição de imagens de impressões digitais, conhecido por Reflexão Interna Total Frustrada (*Frustrated Total Internal Reflection* - FTIR), mostra-se mais adequado ao processamento em tempo real. Essa técnica foi a utilizada para a construção do dispositivo de aquisição de imagens de impressões digitais utilizado nesse trabalho. Este método também é utilizado em [Baldi 1993].

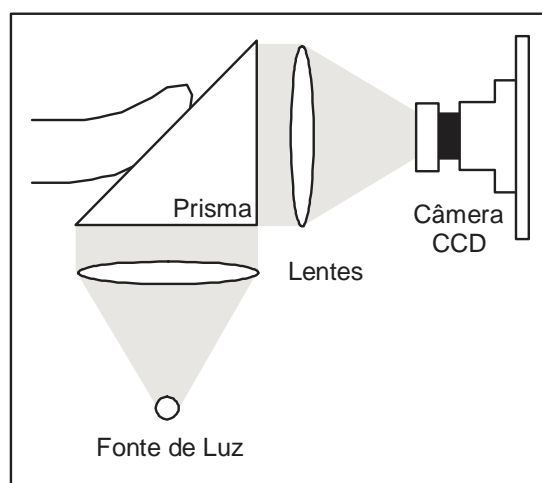


Figura 2.1 - Dispositivo de aquisição de imagens de impressões digitais.

O princípio de funcionamento da Reflexão Interna Total Frustrada é bastante simples de ser entendido, como ilustra a Figura 2.1. Um prisma é utilizado para projetar o padrão da impressão digital sobre uma câmera CCD, do seguinte modo: uma das faces ortogonais do prisma é iluminada por uma fonte de luz, cujos raios interagem com o padrão da pele ao refletir na face diagonal do prisma; a imagem correspondente é então

formada na outra face ortogonal do prisma e focalizada na câmera CCD através de uma lente.

O sinal proveniente da câmera CCD é posteriormente digitalizado por uma placa de captura de quadro instalada em um microcomputador. A placa de captura de quadro utilizada possui capacidade de fornecer imagens de 320 x 240 *pixels* com 32 níveis de cinza.

Como dispositivo sensor foi utilizada uma câmera CCD monocromática em miniatura, padrão NTSC de sinal de vídeo, do tipo usado em sistemas de circuito fechado de televisão. O tempo de aquisição de um quadro de vídeo para o padrão NTSC é de 33,3ms. Optou-se pelo conjunto de lentes fornecido pelo fabricante (Hunt Electronic Co., Ltd.) que apresentava a menor abertura angular ($f = 6\text{mm}$), objetivando a obtenção da menor distorção espacial possível nas imagens.

O conjunto óptico (prisma e lentes) adotado foi dimensionado de maneira a se obter imagens com resolução de 320 pontos por polegada a partir de uma área de exposição para as impressões digitais de 36 x 19mm. É importante ressaltar que o fato da impressão digital ser exposta na face diagonal e projetada numa das faces ortogonais do prisma causa uma compactação espacial de fator 0,707 na imagem resultante.

A fonte de luz utilizada no dispositivo foi um conjunto de três *leds* vermelhos retangulares montados lado a lado, de maneira que uma iluminação uniforme fosse obtida ao longo de uma das faces ortogonais do prisma. Dada a simplicidade do conjunto óptico utilizado no dispositivo de aquisição, não foi possível obter um foco perfeito na área inteira das imagens. Sendo assim foi necessária uma etapa de pré-processamento das imagens para compensar esse problema antes da sua entrada no algoritmo de segmentação, assunto que será abordado com maiores detalhes no Capítulo 3.

A montagem mecânica final do dispositivo de aquisição foi feita em poliestireno, buscando um *design* ergonômico dentro das possibilidades de construção, como mostrado na Figura 2.2. A janela de exposição da impressão digital foi recortada em forma de arco, com o objetivo de orientar o usuário quanto à posição correta de posicionamento do dedo. Essa janela de exposição também tem por objetivo limitar fisicamente a translação e rotação do dedo do usuário quando exposto, minimizando dessa forma a ocorrência de prováveis contratempos na fase de segmentação das imagens.

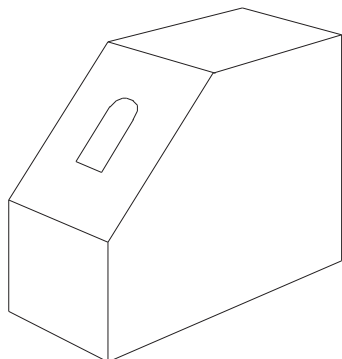


Figura 2.2 - Aspecto final do dispositivo de aquisição.

Na Figura 2.3 são apresentadas duas imagens da mesma impressão digital com a mesma resolução espacial (320 pontos por polegada), obtidas através de impressão com tinta e posterior digitalização através de *scanner* monocromático e através do dispositivo de aquisição construído. Pode-se observar a compactação espacial resultante na imagem obtida por Reflexão Interna Total Frustrada (Figura 2.3b) e também o espelhamento resultante na imagem obtida através de impressão com tinta e posterior digitalização (Figura 2.3a).



(a)



(b)

Figura 2.3 - Comparação das imagens obtidas: (a) pela digitalização de impressão com tinta em papel; (b) por Reflexão Interna Total Frustrada.

Ambos os métodos expostos possuem a desvantagem de possibilitar distorções nas imagens obtidas, incluindo o aparecimento de falsos detalhes e o desaparecimento de detalhes verdadeiros. Entretanto, o método por Reflexão Interna Total Frustrada torna mais difícil a possibilidade de fraude, uma vez que exige o contato físico entre o dedo do usuário e a face diagonal do prisma do dispositivo de aquisição. Outra vantagem do método utilizado é a possibilidade de aquisição das imagens em tempo real. Por outro lado, o método usado apresenta algumas desvantagens, como o fato das imagens obtidas serem dependentes da condição da pele (umidade e limpeza).

Recentemente, outras alternativas mais complexas para a aquisição de imagens de impressões digitais foram propostas, fazendo uso de diversos fenômenos físicos. Uma delas realiza uma varredura ultra-sônica do padrão de pele do dedo do usuário, valendo-se do tempo de reflexão das ondas para mapear a imagem. Outra técnica em desenvolvimento atualmente faz uso do mapeamento da intensidade do campo elétrico do dedo numa matriz de vários pontos para formar a imagem da impressão digital [Setlak 1996].

Todos os métodos de aquisição de imagens de impressões digitais apresentam vantagens e desvantagens particulares, porém todos são susceptíveis à incidência de sinais espúrios (ruído), distorções espaciais e efeitos de translação e rotação na imagem resultante. Para um desempenho adequado do algoritmo de classificação a ser empregado, todos esses fatores indesejáveis necessitam ser minimizados pelas etapas de pré-processamento e segmentação. Se isso não for completamente possível, os algoritmos de extração de características e classificação dos padrões devem ser robustos o suficiente para contornar tais problemas.

2.3 Base de Dados

Inicialmente foi montado um conjunto de 50 imagens de teste, constituído de 5 imagens distintas de cada uma das impressões digitais dos dez dedos das mãos de uma única pessoa. Esse primeiro conjunto teve como objetivo principal dar início aos experimentos com as etapas de pré-processamento e segmentação, avaliando o seu desempenho para diferentes áreas úteis dentro das imagens. Mais tarde, essas imagens serviram para avaliar a técnica de extração de características, utilizando ferramentas estatísticas e de RNA para esse fim. A Figura 2.4 apresenta algumas versões das imagens obtidas com os

diferentes dedos, através das quais pode-se ter uma noção comparativa do tamanho das suas áreas úteis.



(a)



(b)



(c)



(d)



(e)

Figura 2.4 - Imagens obtidas: (a) polegar; (b) indicador; (c) médio; (d) anular; (e) mínimo.

Adiantando o que será descrito em maiores detalhes no Capítulo 3, observou-se que as imagens dos dedos polegares se mostraram mais adequadas para serem utilizadas nos experimentos seguintes. Os maiores motivos para essa conclusão residem na maior área útil existente nas imagens obtidas com os polegares, como pode-se observar na Figura 2.4, e na menor possibilidade de ocorrência de efeitos de translação e rotação na fase de aquisição das imagens para esses dedos.

Sendo assim, passada a fase de testes iniciais montou-se uma segunda base de dados, contendo 1000 imagens correspondentes a 10 versões distintas das impressões digitais de ambos os dedos polegares das mãos de 50 pessoas diferentes. Procurou-se obter essas imagens a partir de pessoas de variadas faixas etárias e de ambos os sexos. Não foram dadas orientações aos voluntários quanto ao posicionamento dos polegares no dispositivo de aquisição, procurando adquirir as imagens do modo mais espontâneo possível. Dessa maneira, não houve preocupação em se obter imagens do núcleo principal das impressões digitais. Esse conjunto de dados foi então utilizado nos experimentos finais de reconhecimento, nos termos do objetivo principal deste trabalho.

CAPÍTULO 3

PRÉ-PROCESSAMENTO E SEGMENTAÇÃO

As imagens resultantes do processo de aquisição e digitalização podem apresentar diversas imperfeições, tais como a presença de *pixels* ruidosos, contraste ou brilho inadequado, foco não uniforme ao longo de sua área, entre outros problemas. A função da etapa de pré-processamento é a de aprimorar a qualidade das imagens para as etapas posteriores. As operações efetuadas nesse estágio são ditas de baixo nível porque trabalham diretamente com os valores de intensidade dos *pixels*, sem nenhum conhecimento prévio sobre quais deles pertencem à região de interesse da imagem ou ao seu fundo.

Por sua vez, a tarefa básica da etapa de segmentação é a de dividir as imagens em unidades significativas. Em outras palavras, consiste em detectar as regiões de interesse, correspondentes aos objetos nas imagens. Esta tarefa apesar de simples de se descrever, muitas vezes é das mais difíceis de se implementar. No caso das imagens de impressões digitais em questão neste trabalho, existe apenas um macro-objeto por imagem, fato que vem facilitar a sua segmentação. Além dessa vantagem, as condições de aquisição, tais como iluminação e localização da área de interesse nas imagens, são bem controladas nesse caso específico, colaborando também para uma maior facilidade na segmentação dessas imagens.

Antes de descrever os pormenores relativos às etapas de pré-processamento e segmentação do sistema de reconhecimento de impressões digitais aqui proposto e implementado na plataforma de *software* MatLab®, será feita uma breve revisão teórica dos fundamentos de processamento digital de imagens utilizados para atingir os resultados descritos na seqüência.

3.1 Conceitos de Processamento Digital de Imagens

Realce de imagens no domínio espacial

O principal objetivo das técnicas de realce é o de destacar detalhes finos na imagem. Alguns dos métodos de realce de imagens no domínio espacial são a filtragem passa-

altas básica e a passa-altas com ênfase em alta frequência. As técnicas de filtragem no domínio espacial são operações orientadas a vizinhança e atuam diretamente sobre a matriz de *pixels*, que nada mais é do que a própria imagem, utilizando o conceito de convolução com máscaras.

A operação de convolução unidimensional entre dois vetores A e B, denotada $A*B$, pode ser entendida como um somatório de produtos entre os valores de A e B, sendo que inicialmente o vetor B é espelhado e após cada soma de produtos é deslocado espacialmente de uma posição.

Este raciocínio pode ser expandido para o caso bidimensional, onde a imagem a ser processada é uma matriz bidimensional relativamente grande e corresponde ao conjunto A do caso unidimensional anterior, enquanto uma matriz de pequenas dimensões, denominada máscara ou janela, corresponde ao conjunto B. A máscara, após ter sido espelhada tanto na horizontal quanto na vertical, percorrerá todos os pontos da imagem deslocando-se ao longo de cada linha, da direita para a esquerda e de cima para baixo, até que tenha sido processado o último elemento da matriz imagem. Esses conceitos são melhor explicados a seguir:

Seja a sub-imagem com valores de níveis de cinza Z_1, \dots, Z_9 :

Z_1	Z_2	Z_3
Z_4	Z_5	Z_6
Z_7	Z_8	Z_9

E seja a máscara 3 x 3 de coeficientes genéricos W_1, \dots, W_9 :

W_1	W_2	W_3
W_4	W_5	W_6
W_7	W_8	W_9

A máscara acima percorrerá a imagem completa desde o seu canto superior esquerdo até seu canto inferior direito. A cada posição relativa da máscara sobre a

imagem, o *pixel* central da sub-imagem em questão será calculado em uma segunda matriz denominada imagem-destino, da seguinte forma:

$$Z = \sum_{i=1}^9 W_i \cdot Z_i \quad (3.1)$$

Operações de convolução com máscaras são amplamente utilizadas no processamento de imagens. Uma seleção apropriada dos coeficientes W_1, \dots, W_9 torna possível uma grande variedade de operações úteis, tais como redução de ruído, realce de detalhes, afinamento e detecção de características da imagem.

A vizinhança definida normalmente é a 8-vizinhança do *pixel* de referência, o que equivale a uma região 3 x 3 na qual o *pixel* central é o de referência. Nos casos em que a vizinhança é 1 x 1, o operador torna-se uma função de transformação ou de mapeamento. As técnicas de processamento de imagem pertencentes a este caso são freqüentemente denominadas técnicas ponto-a-ponto e serão abordadas mais adiante.

Filtro passa-altas

O formato da resposta ao impulso de um filtro passa-altas deve ser tal que a máscara correspondente apresente coeficientes positivos nas proximidades de seu centro e negativos ao se afastar. No caso de uma máscara 3 x 3, isto significa projetar uma máscara com *pixel* central positivo e todos seus oito vizinhos negativos [Gonzalez 1992]. Um exemplo de máscara com estas características é apresentado abaixo:

$$\frac{1}{9} \begin{bmatrix} -1 & -1 & -1 \\ -1 & 8 & -1 \\ -1 & -1 & -1 \end{bmatrix} \quad (3.2)$$

Pode-se notar que a soma algébrica dos coeficientes desta máscara é zero, significando que quando aplicada a regiões homogêneas de uma imagem, o resultado será zero ou um valor muito baixo, fato que é consistente com o princípio da filtragem passa-altas.

Filtro *high-boost*

A filtragem passa-altas também pode ser obtida subtraindo-se da imagem original uma versão submetida a um filtro passa-baixas. O filtro *high-boost* ou técnica da ênfase em

alta frequência nada mais é do que uma extensão desse conceito, na qual a imagem original é multiplicada por um fator de amplificação A .

A ênfase em alta frequência pode ser implementada utilizando-se a máscara apresentada a seguir:

$$\frac{1}{9} \begin{bmatrix} -1 & -1 & -1 \\ -1 & w & -1 \\ -1 & -1 & -1 \end{bmatrix}, \text{ onde } w = 9A - 1 \quad (3.3)$$

Quando $A = 1$, o filtro se comporta de forma idêntica a um passa-altas comum. Nos casos em que $A > 1$, parte da imagem original é adicionada ao resultado, restaurando parcialmente os componentes de baixa frequência. O resultado é uma imagem que se parece com a original, com um grau relativo de realce das bordas, dependente do valor de A . O processo genérico de subtração de uma imagem borrada da imagem original é conhecido na literatura como *unsharp masking* [Gonzalez 1992].

Equalização de histograma

O histograma de uma imagem consiste em um conjunto de números indicando a quantidade de *pixels* naquela imagem que apresentam um determinado tom de cinza. Esses valores são normalmente representados por um gráfico de barras que fornece para cada tom de cinza o número ou o percentual de *pixels* correspondentes na imagem. Através da visualização do histograma de uma imagem podemos ter uma indicação de sua qualidade quanto ao seu brilho médio e nível de contraste. Em outras palavras, há uma indicação se a imagem é predominantemente clara ou escura, e se possui alto ou baixo contraste.

As técnicas de modificação de histograma são conhecidas também como técnicas ponto-a-ponto, uma vez que o valor de tom de cinza de um certo *pixel* após o processamento depende apenas do seu valor original. Em contraste, nas técnicas de processamento orientadas a vizinhança, o valor resultante depende também dos *pixels* em torno do elemento de imagem original, como exposto anteriormente.

Diversas técnicas de modificação da distribuição dos *pixels* na escala de cinza podem ser implementadas a partir do conceito de transformação de intensidade. As

transformações de intensidade são funções que mapeiam os valores de tom de cinza de cada *pixel* da imagem original em um novo tom de cinza na imagem destino.

A equalização de histograma é uma técnica a partir da qual se procura redistribuir os valores de tons de cinza dos *pixels* em uma imagem, de modo a obter um histograma uniforme, no qual a quantidade relativa de *pixels* de qualquer nível de cinza é praticamente a mesma. Para tanto, utiliza-se uma função auxiliar denominada função de transformação. A forma mais usual de se equalizar um histograma é utilizar a função de distribuição acumulada da distribuição de probabilidades original, a qual pode ser expressa por:

$$s_k = \sum_{j=0}^k \frac{n_j}{n} = \sum_{j=0}^k p_r(r_j), \text{ para } 0 \leq r_k \leq 1 \quad (3.4)$$

onde k é a quantidade de níveis de cinza envolvidos, n_j é a quantidade de *pixels* com o nível de cinza j presentes na imagem e n é a quantidade total de *pixels* da imagem. Conforme a outra notação, $p_r(r_j)$ é a probabilidade de ocorrência do nível de cinza r_j , normalizado entre os valores 0 e 1.

O resultado final obtido pela equalização de histograma é uma melhoria do contraste geral da imagem original [Gonzalez 1992].

Limiarização

O princípio da limiarização consiste em separar as regiões de uma imagem quando esta apresenta duas classes (fundo e objeto). Devido ao fato da limiarização produzir uma imagem binária na saída, esse processo também é denominado binarização.

A forma mais simples de limiarização consiste na bipartição do histograma, convertendo os *pixels* cujo tom de cinza é maior ou igual a um certo valor de limiar (T) em brancos e os demais em pretos. No caso de níveis de cinza divididos basicamente em duas classes, onde o histograma apresenta dois picos e um vale, a limiarização é trivial. Para casos mais complexos foram desenvolvidas técnicas estatísticas para a determinação do valor de limiar.

A iluminação desempenha um papel significativo no processo de limiarização, uma vez que provoca alterações no histograma original da imagem, eventualmente eliminando uma região de vale entre dois picos, naturalmente propícia para a definição de um limiar global.

Operações morfológicas de dilatação e erosão binárias

A morfologia matemática está centrada no estudo da estrutura geométrica dos objetos presentes em uma imagem. Pode ser empregada em varias áreas de processamento e análise de imagens, para objetivos tão distintos quanto realce, filtragem, segmentação, detecção de bordas, esqueletização, afinamento, entre outros.

O principio básico da morfologia matemática consiste em extrair informações relativas à geometria e à topologia de um conjunto desconhecido (uma imagem), pela transformação através de outro conjunto completamente definido, chamado elemento estruturante. A base da morfologia matemática, portanto, é a teoria de conjuntos. O conjunto de todos os *pixels* pretos em uma imagem binária descreve completamente a imagem, uma vez que os demais pontos só podem ser brancos. Em imagens binárias, os conjuntos em questão são membros do espaço inteiro bidimensional Z^2 , onde cada elemento é um vetor de coordenadas (x,y) de cada *pixel* preto (por convenção) na imagem.

Algumas definições básicas para o entendimento das operações de morfologia matemática são dadas a seguir:

Sejam A e B conjuntos em Z^2 , cujos componentes são $a = (a_1, a_2)$ e $b = (b_1, b_2)$, respectivamente. A translação de A por $x = (x_1, x_2)$, denotada A_x , é dada por:

$$A_x = \{c \mid c = a + x, \text{ para } a \in A\} \quad (3.5)$$

A reflexão de B , denotada \tilde{B} , é definida como:

$$\tilde{B} = \{x \mid x = -b, \text{ para } b \in B\} \quad (3.6)$$

O complemento do conjunto A é:

$$A^c = \{x \mid x \notin A\} \quad (3.7)$$

Finalmente, a diferença entre dois conjuntos A e B , denotada $A - B$, é definida como:

$$A - B = \{x \mid x \in A, x \notin B\} = A \cap B^c \quad (3.8)$$

A seguir é exposta a operação de dilatação:

Sejam A e B conjuntos no espaço Z^2 e seja \emptyset o conjunto vazio. A dilatação de A por B , denotada $A \text{ dil } B$, é dada por:

$$A \text{ dil } B = \{x \mid \tilde{B}_x \cap A \neq \emptyset\} \quad (3.9)$$

O processo de dilatação consiste em obter a reflexão de B sobre sua origem e depois transladar esta reflexão de x . A dilatação de A por B é, então, o conjunto de

todos os x deslocamentos para os quais a interseção de \tilde{B}_x e A inclui pelo menos um elemento diferente de zero. Com base nesta interpretação, a equação anterior pode ser escrita como:

$$A \text{ dil } B = \left\{ x \mid [\tilde{B}_x \cap A] \subset A \right\} \quad (3.10)$$

O conjunto B é normalmente denominado elemento estruturante.

Os efeitos da dilatação são: aumento das partículas, preenchimento de pequenos buracos e conexão de grãos próximos [Facon 1996].

Por fim, é exposta a operação de erosão:

Sejam A e B conjuntos no espaço Z^2 . A erosão de A por B , denotada $A \text{ ero } B$, é definida como:

$$A \text{ ero } B = \left\{ x \mid B_x \subset A \right\} \quad (3.11)$$

Em outras palavras, significa dizer que a erosão de A por B resulta no conjunto de pontos x tais que B , transladado de x , está contido em A .

Os efeitos obtidos pela erosão são: diminuição de partículas, eliminação de grãos de tamanho inferior ao tamanho do elemento estruturante, aumento dos buracos e separação de grãos próximos [Facon 1996].

Pode ser provado que a dilatação e a erosão são operações duais entre si com respeito ao complemento e à reflexão [Gonzalez 1992]. Dessa forma:

$$(A \text{ ero } B)^c = A^c \text{ dil } \tilde{B} \quad (3.12)$$

3.2 Pré-processamento das Imagens

As imagens resultantes do processo de aquisição apresentaram alguns problemas que possivelmente prejudicariam o desempenho das etapas de extração de características e classificação dos padrões. Devido a esse fato, foi necessário o estudo de um conjunto de técnicas de processamento digital de imagens para o melhoramento da qualidade das imagens a serem utilizadas pelas etapas posteriores.

Problemas relativos à incidência de ruído e deformações espaciais mostraram-se presentes nas imagens, como já era esperado. Porém partiu-se do princípio que estes problemas deveriam ser contornados pelas técnicas de extração de características e classificação, evitando assim o uso de algoritmos específicos, conforme um dos objetivos propostos. Tal decisão procura a utilização mínima possível de técnicas de

processamento digital de imagens, as quais normalmente exigem grande esforço computacional.

Também ocorreram outros problemas, previamente mencionados nos capítulos anteriores, tais como a compactação espacial das imagens em uma de suas dimensões, resultante da projeção da face diagonal do prisma do dispositivo de aquisição em uma de suas faces ortogonais, e a falta de uniformidade do foco obtida ao longo da área das imagens, fato resultante da simplicidade do conjunto óptico utilizado.



(a)



(b)

Figura 3.1 - Redimensionamento das imagens: (a) imagem original; (b) imagem redimensionada por um fator de 1,414 na horizontal.

Inicialmente houve a intenção de compensar o efeito de compactação espacial, anteriormente mencionado, através de interpolação bilinear de colunas nas imagens. O

fator a ser utilizado no redimensionamento das imagens seria de 1,414 (o inverso do fator de compactação de 0,707), resultando em imagens de 452 x 240 *pixels*. O resultado de tal operação pode ser observado na Figura 3.1.

Mais tarde observou-se que a realização de tal operação de redimensionamento não trazia benefícios significativos à etapa de extração de características, resultando apenas em maior tempo de processamento para o sistema como um todo. Sendo assim, foi tomada a decisão de desprezar as laterais das imagens, onde não ocorre a presença de sinais relevantes, de modo a otimizar a área de processamento, ocupando apenas a região central de 240 x 240 *pixels* das imagens e reduzindo o processamento de *pixels* irrelevantes. O aproveitamento obtido pode ser avaliado no exemplo da Figura 3.2.



(a)



(b)

Figura 3.2 - Aproveitamento das imagens: (a) imagem original de 320 x 240 *pixels*; (b) região útil de 240 x 240 *pixels*, desprezando-se as laterais.

Ao contrário do problema de compactação espacial, a não uniformidade no foco da imagem mostrou ser de relevância para a etapa de extração de características. Tal fato foi constatado após a submissão das imagens originais ao algoritmo de compressão de imagens de impressões digitais especificado pelo FBI [Brislaw 1996, FBI 1993], quando foi possível perceber a baixa qualidade obtida na reconstrução das regiões das imagens onde o foco ficou comprometido, como pode-se observar na Figura 3.3. Sendo assim, conclui-se que quanto mais definida for a imagem, melhores serão os resultados obtidos nessa análise e, conseqüentemente, melhores os resultados obtidos pela extração de características.

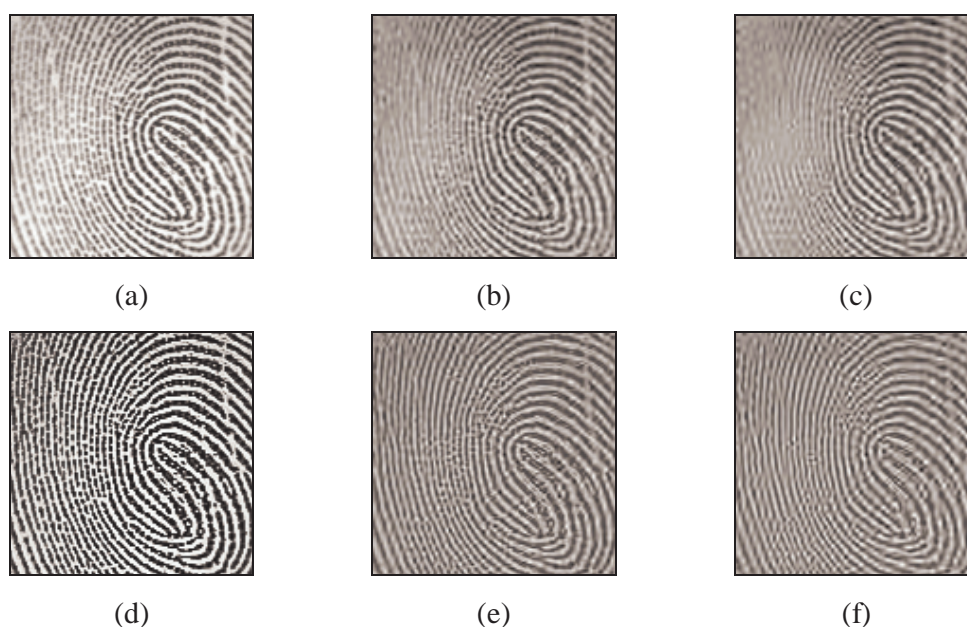


Figura 3.3 - Reconstrução das imagens após compactação pelo método WSQ: (a) imagem original; (b) reconstrução da imagem (a) após compactação de 1:10; (c) reconstrução da imagem (a) após compactação de 1:20; (d) imagem pré-processada; (e) reconstrução da imagem (d) após compactação de 1:10; (f) reconstrução da imagem (d) após compactação de 1:20.

A característica de falta de foco em determinadas áreas da imagem representa na realidade uma deficiência nas suas componentes de alta frequência, uma vez que estas são as responsáveis por detalhes finos, tais como as bordas presentes na imagem. Partindo-se dessa premissa, aplicou-se às imagens inicialmente um filtro passa-altas

comum, com o intuito de realçar as bordas presentes nas impressões digitais, utilizando para isso a máscara de convolução da equação (3.2).

Para que a perda de componentes de baixa frequência não fosse tão exagerada, optou-se pela utilização do filtro *high-boost*, utilizando-se o coeficiente central da máscara de convolução da equação (3.3) com valor $w=10$. Desse modo, as componentes de baixa frequência das imagens foram preservadas, sendo realizada uma ênfase nas altas frequências. Essa operação contribuiu significativamente para a melhoria do foco das imagens, como pode ser visto na Figura 3.4.

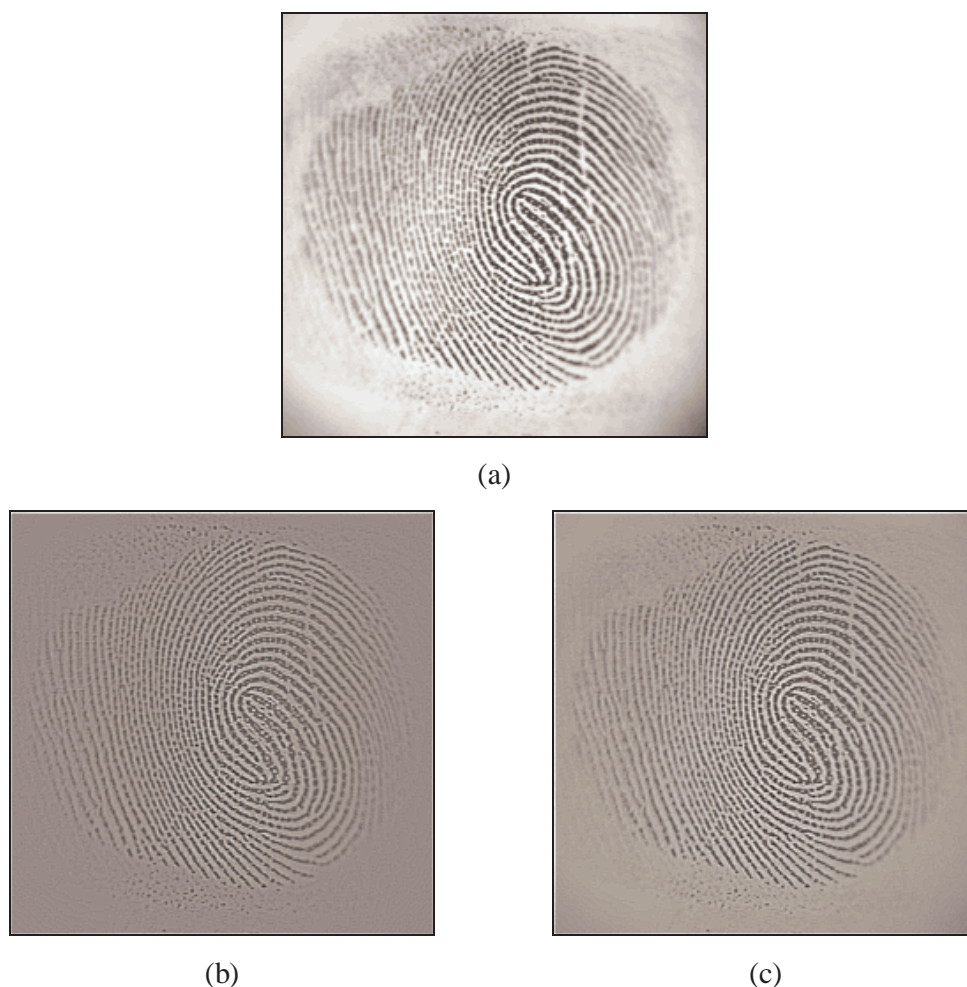


Figura 3.4 - Realce das imagens: (a) imagem original; (b) resultado do filtro passa altas aplicado à imagem (a); (c) resultado do filtro *high-boost* aplicado à imagem (a).

Pode-se reparar na Figura 3.4 que o contraste das imagens resultantes da filtragem passa-altas com ênfase em altas frequências foi seriamente comprometido, fato que motivou a aplicação de uma equalização de histograma para minimizar tal problema. Na realidade, a utilização de uma equalização de histograma posteriormente a uma filtragem passa-altas com ênfase em altas frequências é uma técnica comumente empregada em processamento digital de imagens [Gonzalez 1992]. Um exemplo de resultado obtido após a equalização de histograma pode ser visto na Figura 3.5.

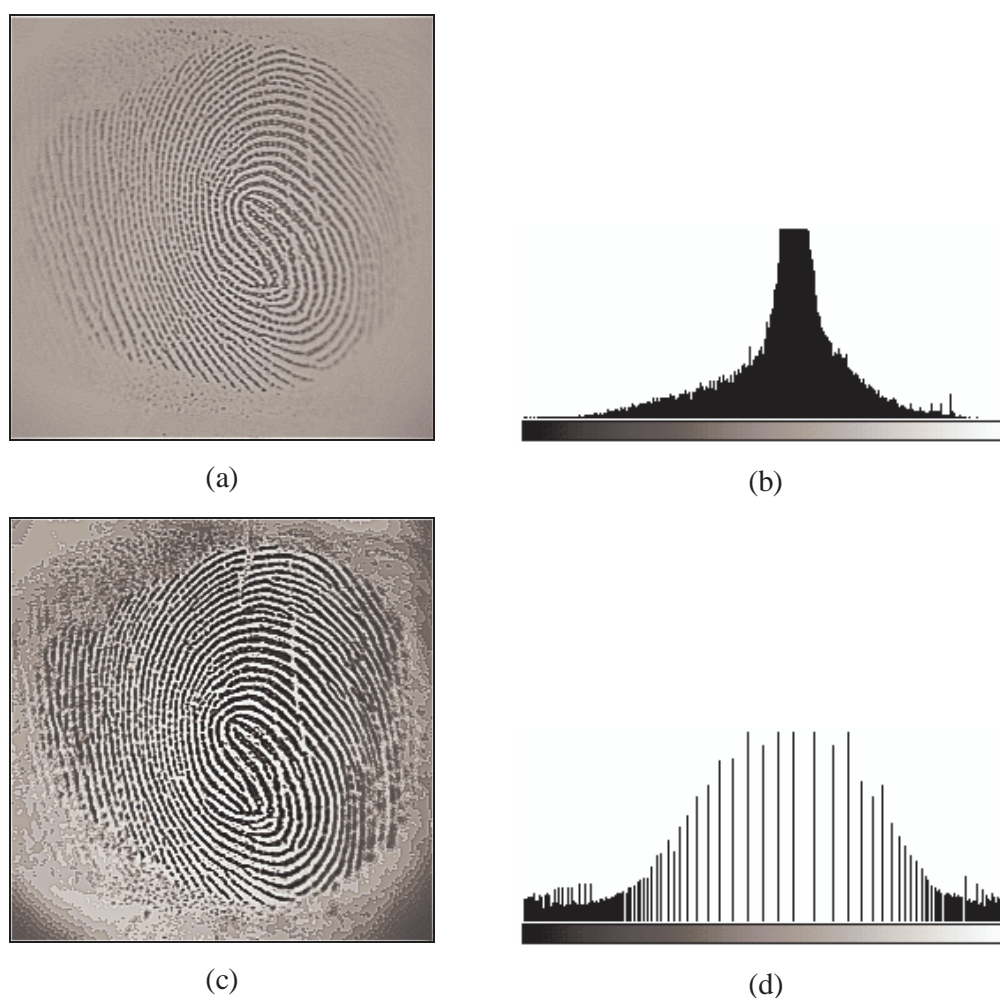


Figura 3.5 - Equalização de histograma: (a) imagem filtrada; (b) histograma da imagem filtrada; (c) imagem equalizada; (d) histograma da imagem equalizada.

O processo de equalização de histograma proporcionou uma distribuição mais uniforme dos níveis de cinza, resultando na esperada melhoria de contraste das imagens.

Com essa operação os resultados obtidos foram considerados adequados para a etapa seguinte de processamento, ou seja, a segmentação das imagens.

A Figura 3.6 ilustra as operações utilizadas na etapa de pré-processamento das imagens em forma de diagrama em blocos.

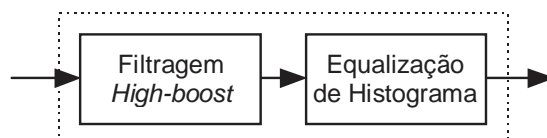


Figura 3.6 - Diagrama em blocos da etapa de pré-processamento.

Abaixo, o algoritmo de pré-processamento das imagens adquiridas é apresentado em pseudo-código:

1. Ler a imagem original adquirida (320 x 240 *pixels*)
2. Recortar a região central (240 x 240 *pixels*) da imagem original
3. Gerar a máscara de filtragem *high-boost* com coeficiente $A=10$
4. Convoluir a máscara de filtragem *high-boost* com a imagem recortada
5. Remapear valores da imagem filtrada para a faixa de valores de 0 a 255
6. Obter o histograma da imagem filtrada
7. Equalizar o histograma da imagem filtrada

3.3 Segmentação das Imagens

Uma vez realizada a adequação das imagens, torna-se necessária a determinação da região de interesse de cada uma delas, a ser submetida ao estágio de extração de características. A etapa de segmentação é exatamente a responsável pela determinação de tais regiões de interesse nas imagens.

Inicialmente decidiu-se realizar uma limiarização das imagens pré-processadas, com o objetivo principal de utilizar operações de morfologia matemática binária para a determinação da região de interesse das imagens. Graças às condições de iluminação bem controladas do dispositivo de aquisição das imagens, o processo de limiarização pode ser efetuado através da determinação de um valor fixo de limiar, evitando o emprego de técnicas estatísticas para a sua determinação.

Experimentos demonstraram que a melhor localização para o limiar seria próximo à extremidade inferior do histograma e optou-se pelo valor $T=32$. Essa escolha resultou em imagens cujos *pixels* pretos correspondem à região de melhor foco das impressões digitais, como pode ser visto no exemplo da Figura 3.7.

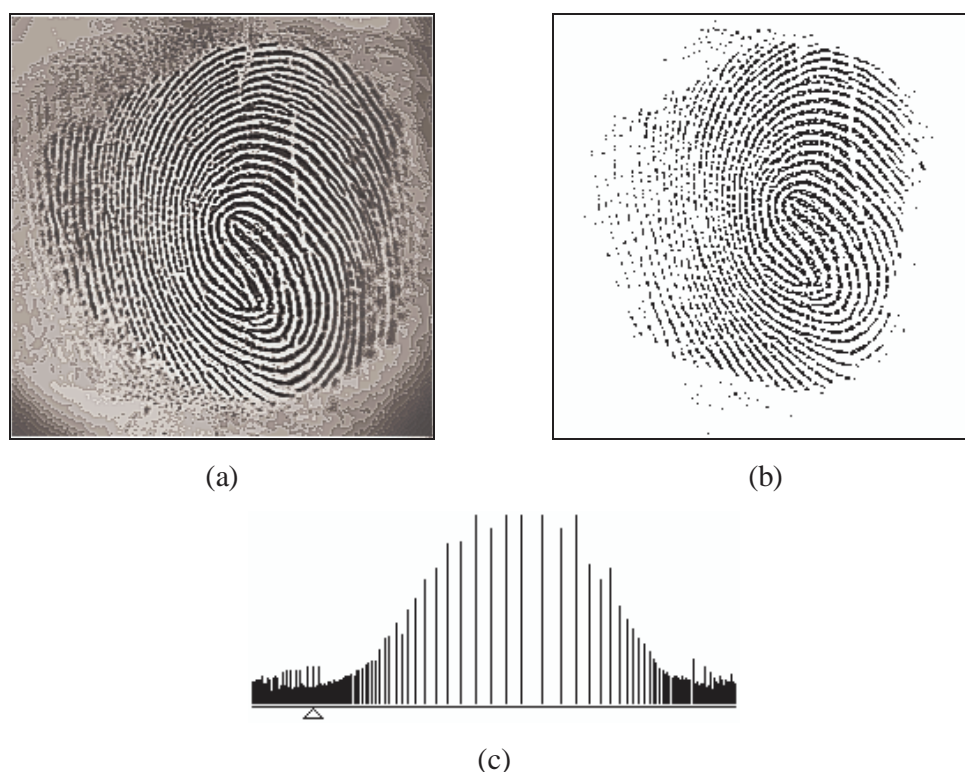


Figura 3.7 - Limiarização das imagens: (a) imagem pré-processada; (b) imagem limiarizada ($T=32$); (c) localização do limiar T no histograma.

A partir da imagem binarizada obtida, procurou-se então obter uma seqüência de operações morfológicas que resultassem em uma imagem contendo um único bloco conciso de *pixels* pretos, correspondente ao centro aproximado da região de interesse da imagem pré-processada correspondente. Para a execução dessa tarefa, utilizou-se primeiramente a operação de dilatação morfológica com cinco iterações, objetivando o fechamento dos buracos existentes entre os sulcos das impressões digitais. Em seguida foram efetuadas 30 iterações da operação de erosão morfológica para a eliminação de *pixels* pretos indesejáveis e também para redução do tamanho do bloco de *pixels* pretos resultante. Para todas as operações foi utilizado um elemento estruturante quadrado. Um exemplo do resultado final dessas operações pode ser observado na Figura 3.8.

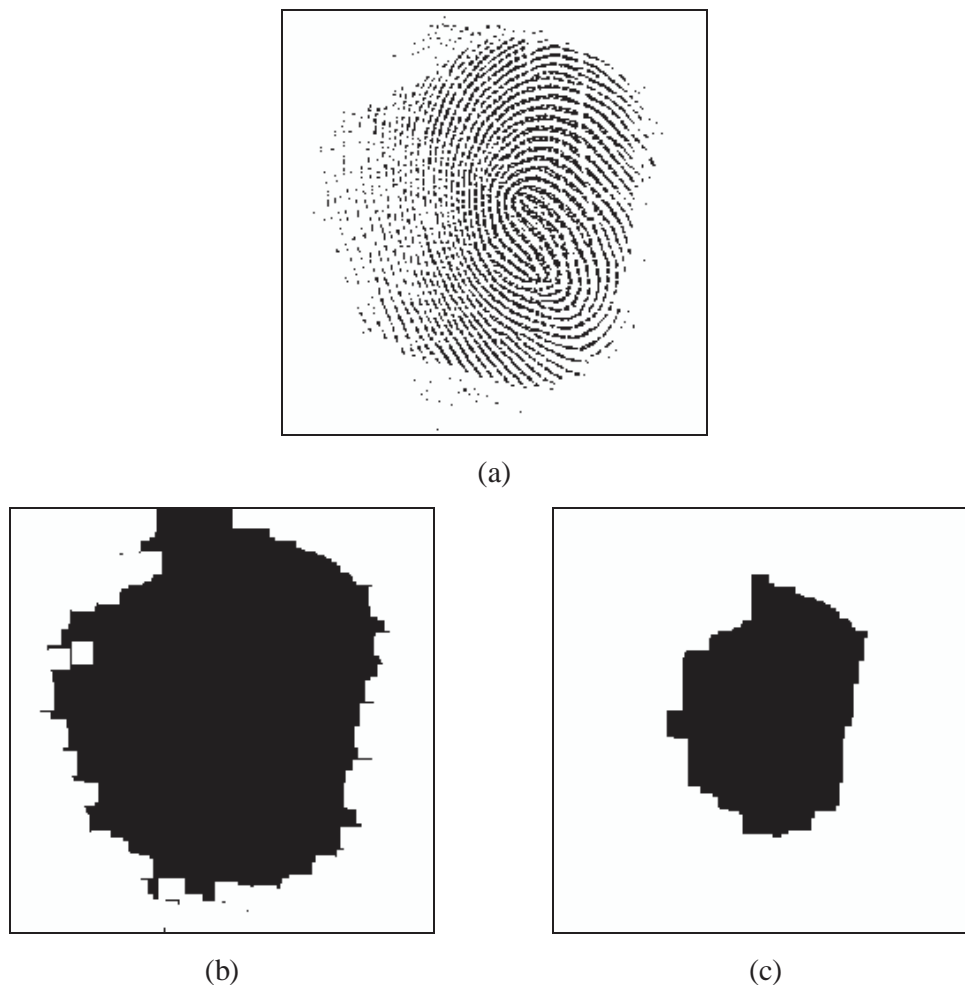


Figura 3.8 - Operações morfológicas: (a) imagem binarizada; (b) dilatação morfológica - 5 iterações; (c) erosão morfológica - 30 iterações.

O passo seguinte para a determinação da região de interesse foi varrer a imagem resultante das operações morfológicas, com o objetivo de obter as coordenadas dos vértices do menor retângulo envolvente do bloco de *pixels* pretos resultante. A partir dessas coordenadas foi possível determinar as coordenadas do centro da região de interesse, correspondente ao centro geométrico do retângulo obtido. Uma vez determinadas essas coordenadas centrais, foi possível recortar na imagem pré-processada correspondente um quadrado de 128 x 128 *pixels*, a ser futuramente submetido ao estágio de extração de características. Esse processo pode ser melhor entendido através do exemplo da Figura 3.9.

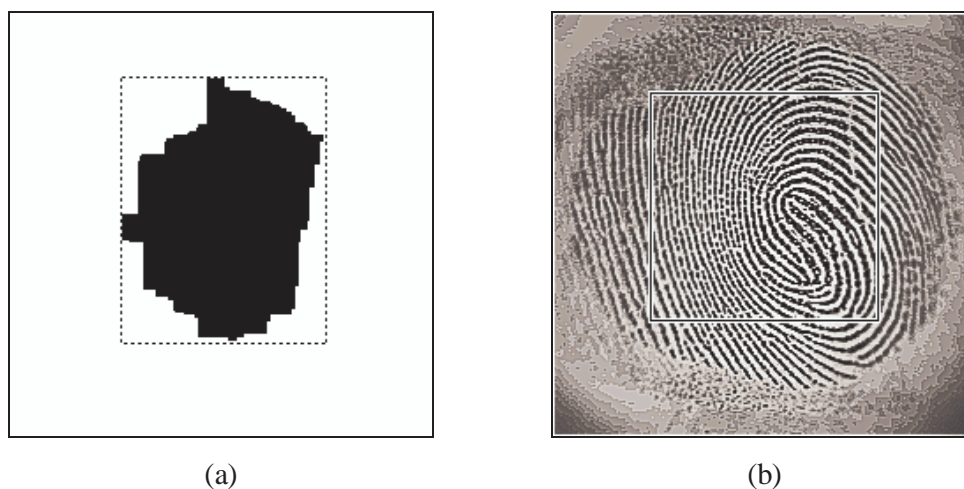


Figura 3.9 - Determinação da região de interesse: (a) imagem resultante das operações morfológicas, com o menor retângulo envolvente do bloco de *pixels* pretos assinalado; (b) região de interesse resultante de 128 x 128 *pixels* na imagem pré-processada correspondente.

O processo de segmentação das imagens de impressões digitais através do uso de morfologia matemática descrito anteriormente, quando aplicado a versões distintas da mesma impressão digital, apresenta resultados significativamente diferentes, devido principalmente a efeitos indesejados quanto ao posicionamento do dedo no dispositivo de aquisição. Esse fato não seria interessante para a etapa de extração de características, que também apresentaria resultados razoavelmente diferentes para versões diferentes da mesma impressão digital. Portanto, é desejável que as imagens resultantes do processo de segmentação de uma mesma impressão digital possuam a máxima semelhança possível.

Para contornar a dificuldade de obtenção de regiões de interesse semelhantes para imagens de uma mesma impressão digital, foi idealizada uma segunda técnica de segmentação para ser utilizada nas demais imagens de uma mesma classe, a partir do resultado obtido na segmentação da primeira imagem através de morfologia matemática. Optou-se pela utilização de uma sub-imagem central da região de interesse de 128 x 128 *pixels* obtida pela segmentação por morfologia matemática, que foi empregada como região de referência na determinação do máximo coeficiente de correlação com as demais imagens. O tamanho adotado para a região central de referência foi de 32 x 32 *pixels* e o motivo de tal escolha foi a necessidade de se utilizar uma matriz da menor

dimensão possível para a realização da correlação, pois infelizmente trata-se de uma operação que exige grande esforço computacional.

A segmentação por correlação aqui proposta baseia-se em um princípio simples. Realiza-se uma varredura da imagem a ser segmentada, deslizando-se a imagem de referência obtida na primeira imagem da classe e calculando-se os respectivos coeficientes de correlação correspondentes a cada posição, de maneira semelhante à utilizada em [Baldi 1993]. Como o coeficiente de correlação calculado em cada posição corresponde ao grau de semelhança entre a sub-imagem em questão e a própria imagem de referência, uma vez encontrada a posição de máxima semelhança, recorta-se a imagem de 128 x 128 *pixels* cujo centro geométrico esteja nessa posição. A Figura 3.10 ilustra esse processo para duas imagens distintas de uma mesma impressão digital.



Figura 3.10 - Segmentação por máxima correlação: (a) primeira imagem de uma classe de impressões digitais, segmentada por morfologia matemática, com as regiões de interesse e de referência assinaladas; (b) outra imagem da mesma classe de impressões digitais, segmentada por máxima correlação com a região de referência da primeira imagem, também com as regiões de interesse e de referência assinaladas.

Através da técnica de segmentação por máxima correlação foi possível minimizar problemas de translação que normalmente acontecem em versões distintas da mesma impressão digital, como ocorre no caso ilustrado na Figura 3.10. No entanto, esse método apresenta o inconveniente de ser bastante sensível a problemas de rotação. Essa

sensibilidade pode ser observada melhor quando utilizadas imagens de impressões digitais de dedos menores, cuja área útil é pequena e conseqüentemente mais suscetível a rotações no momento de aquisição. Como antecipado no capítulo anterior, os melhores resultados foram obtidos com polegares, devido à maior área útil das imagens e à limitação física de rotação imposta pela janela de exposição do dispositivo de aquisição de imagens. Um exemplo do resultado final dos métodos de segmentação, propostos para três imagens distintas de uma impressão digital de um mesmo polegar, pode ser observado na Figura 3.11. Um sub-conjunto de imagens pré-processadas e segmentadas da primeira e da segunda base de dados é dado no Apêndice A, com o objetivo de melhor ilustrar a diversidade de condições das imagens utilizadas nos experimentos realizados durante a elaboração deste trabalho.

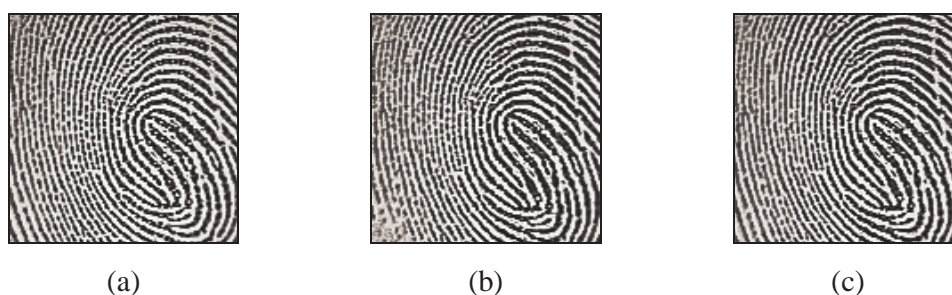


Figura 3.11 - Resultados obtidos com a segmentação: (a) imagem segmentada por morfologia matemática; (b) e (c) imagens distintas segmentadas por máxima correlação com a região de referência de (a).

O diagrama em blocos da etapa de segmentação, envolvendo as duas técnicas empregadas, está ilustrado na Figura 3.12.

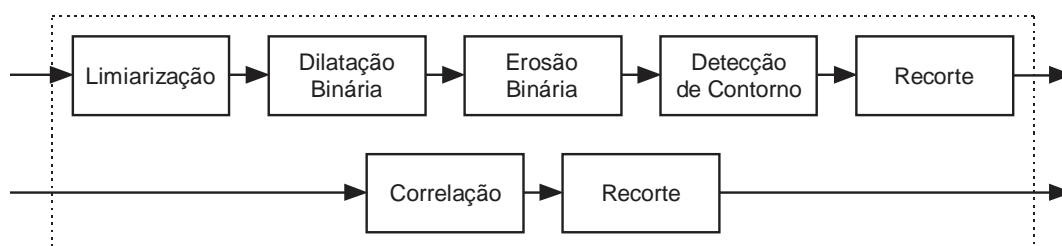


Figura 3.12 - Diagrama em blocos da etapa de segmentação.

A seguir, o algoritmo de segmentação das imagens pré-processadas é apresentado em pseudo-código:

1. Ler a imagem equalizada
2. Se for a primeira imagem da classe:
 - A. Limiarizar a imagem equalizada com limiar $T=32$
 - B. Efetuar 5 iterações de dilatação binária com elemento estruturante quadrado (3 x 3) na imagem limiarizada
 - C. Efetuar 25 iterações de erosão binária com elemento estruturante quadrado (3 x 3) na imagem dilatada
 - D. Varrer as colunas da imagem erodida, armazenando as coordenadas horizontais mínima (xmin) e máxima (xmax) dos pixels escuros existentes na imagem
 - E. Varrer as linhas da imagem erodida, armazenando as coordenadas verticais mínima (ymin) e máxima (ymax) dos pixels escuros existentes na imagem
 - F. Recortar a região de interesse (128 x 128 *pixels*) da imagem equalizada a partir do centro do retângulo resultante das coordenadas (xmin,ymin) e (xmax,ymax)
 - G. Armazenar a imagem da região de interesse
 - H. Recortar a região de referência central (32 x 32 *pixels*) da região de interesse da imagem
 - I. Armazenar a imagem da região de referência central
- Se for uma das demais imagens da classe:
 - A. Ler a imagem da região de referência da primeira imagem
 - B. Varrer a imagem equalizada, avaliando em cada ponto a correlação entre a região de referência da primeira imagem com a sub-imagem equalizada em questão
 - C. Determinar as coordenadas do ponto de máxima correlação avaliado
 - D. Recortar a região de interesse (128 x 128 *pixels*) da imagem equalizada a partir das coordenadas do ponto de máxima correlação avaliado
 - E. Armazenar a imagem da região de interesse

Concluída a segmentação coerente das imagens, estas estão prontas a serem submetidas ao método de extração de características proposto, que deverá gerar vetores que representem compacta e adequadamente os membros das diferentes classes de impressões digitais envolvidas no treinamento do sistema.

CAPÍTULO 4

EXTRAÇÃO DE CARACTERÍSTICAS

A obtenção de uma técnica coerente de extração de características relevantes de um padrão de imagem não constitui uma atividade das mais simples. O estágio de extração de características, também conhecido normalmente como etapa de representação e descrição dos padrões, tem como objetivo reduzir a dimensão dos padrões de entrada da etapa de classificação, tornando viável o seu processamento. Também é de suma importância que o método utilizado resulte em vetores de características capazes de diferenciar as diferentes classes de padrões envolvidas no processo, possibilitando ao mesmo tempo que padrões da mesma classe sejam reconhecidos como tal.

Diversos métodos foram desenvolvidos para o caso particular de extração de características de imagens de impressões digitais, em sua maioria baseados nas características especiais dos padrões de impressão digital, tais como detecção de minúcias e orientação dos sulcos [Ratha 1995, Karu 1996]. Em [Blue 1994], a transformada Karhunen-Loève do padrão de orientação dos sulcos de impressões digitais foi utilizada para se obter vetores de características. Entretanto, uma dificuldade que surge nessas abordagens tradicionais é o emprego de algoritmos especializados para extrair informações de orientação e minúcias.

Por outro lado, as transformadas matemáticas tornaram-se alguns dos métodos mais utilizados no processamento e análise de sinais e imagens. A principal razão para o seu uso é o fato da informação mais relevante do sinal poder ser comprimida, constituindo uma forma possível de representação e descrição para sistemas de classificação e reconhecimento de padrões. Essas transformações matemáticas expressam o sinal como uma combinação linear de funções básicas. No caso da transformada de Fourier, por exemplo, as funções básicas são termos em senos e cossenos, constituindo uma das formas de se modelar sinais [Stearns 1990].

Através da transformada *wavelet*, sinais são modelados como uma combinação linear de funções especiais, chamadas *wavelets*, que são dilatadas e transladadas para formar uma base de representação localizada no tempo (ou espaço) e na escala. Existe

uma infinidade de possibilidades de bases *wavelet*, as quais podem ser agrupadas em famílias. As classificações mais comuns são quanto à natureza da transformação (contínua ou discreta), quanto ao tipo de base gerada (ortonormal ou biortogonal), quanto à suavidade das funções (regularidade) e quanto à sua duração no tempo (suporte) [Vaidyanathan 1995]. Em [Vidakovic 1994] é dada uma excelente introdução à teoria de *wavelets*.

Para o caso de imagens de impressões digitais, as informações tidas como de maior relevância são constituídas pelas minúcias, que podem ser consideradas como descontinuidades no padrão da impressão digital. Portanto, é razoável esperar que funções básicas localizadas no tempo representem melhor esse tipo de sinal do que funções básicas com suporte não-compacto, como as funções de Fourier. O padrão WSQ (*Wavelet Scalar Quantization*) de compressão de imagens de impressões digitais estabelecido pelo FBI [Brislaw 1996, FBI 1993] utiliza a transformada *wavelet* como alicerce e constitui um forte indício da adequação do uso dessa técnica para padrões de impressões digitais. Uma discussão sobre o uso de diferentes bases *wavelet* para a compressão de imagens é feita em [Villasenor 1995]. Além desse fato, os algoritmos de transformadas *wavelet* discretas, por serem técnicas de processamento de sinais, são mais facilmente implementados em processadores digitais de sinais ou *hardware* específico do que os algoritmos resultantes das técnicas orientadas a sulcos e minúcias, que exigem técnicas específicas.

A técnica de extração de características aqui proposta segue algumas das especificações do FBI para o cálculo da transformada *wavelet* sobre as imagens de impressões digitais, a partir da qual são retiradas informações para compor vetores de características dos padrões. Assim como as etapas de pré-processamento e segmentação, a etapa de extração de características aqui descrita foi implementada para a plataforma de *software* MatLab, com o auxílio de funções da *toolbox* WaveLab [Buckheit 1995].

4.1 Transformada *Wavelet*

A transformada *wavelet* [Mallat 1989] permite a decomposição de sinais em componentes bem localizadas no tempo e na frequência (ou espaço e escala, para o caso bidimensional), fornecendo informações locais que não são possíveis de serem obtidas através de métodos tradicionais, como a transformada de Fourier [Vaidyanathan 1995].

A decomposição *wavelet* é realizada a partir de translações e dilatações de uma função conhecida por *wavelet* mãe. Versões contraídas da *wavelet* mãe (alta frequência) realizam uma análise temporal detalhada, enquanto versões dilatadas (baixa frequência) realizam uma análise detalhada em frequência.

Uma *wavelet* constitui na realidade um filtro passa-faixa, quando analisada do ponto de vista de processamento de sinais. Dessa forma, a transformada *wavelet* de um determinado sinal pode ser interpretada como a decomposição desse sinal em faixas de frequência com banda passante de largura relativa constante (Q constante) [Masters 1994], fato que possibilita a chamada análise em multiresolução [Gopinath 1993, Mallat 1989].

O modelo da transformada *wavelet* pode ser estendido para o caso bidimensional, aplicável a imagens, conforme demonstrado em [Mallat 1989].

Desenvolvimento teórico

A transformada *wavelet* de uma função $f(x)$ na sua forma mais geral é dada por

$$WT(a,b) = \left\langle f, \psi_{a,b} \right\rangle = |a|^{-1/2} \int_{-\infty}^{\infty} f(x) \psi\left(\frac{x-b}{a}\right) dx \quad (4.1)$$

onde $\langle _, _ \rangle$ representa o produto interno e $\psi_{a,b}(x) = \psi\left(\frac{x-b}{a}\right)$ representa a *wavelet*, com os parâmetros a e b representando dilatação e translação, respectivamente. Como $\psi(x)$ é escolhida de forma a possuir um espectro passa-faixa, o sinal no domínio de transformação nada mais é que uma versão filtrada do sinal no domínio do tempo, com banda passante dependente de a , e centrada em b .

Na forma discreta, a transformada *wavelet* é dada por

$$DWT(m,n) = c_{mn} = \left\langle f, \psi_{mn} \right\rangle \quad (4.2)$$

$$\text{onde } \psi_{mn}(x) = a_0^{-m/2} \psi(a_0^{-m} x - nb_0) = a_0^{-m/2} \psi(a_0^{-m} (x - na_0^m b_0)) \quad (4.3)$$

onde a_0 e b_0 são constantes e m e n inteiros que proporcionam dilatação e translação.

No caso discreto apenas amostras são calculadas no espaço bidimensional de transformação. O sinal no domínio do tempo é transformado em uma série de versões filtradas por diferentes filtros passa-faixa, cujas bandas passantes e taxas de amostragem diminuem com o aumento de m (dilatação de ψ). A dependência da taxa de amostragem

com a dilatação fica evidente a partir da equação (4.3), em consistência com a teoria da amostragem.

Uma questão importante que deve ser levantada é se existe uma transformada *wavelet* inversa que torne possível a reconstrução de $f(x)$ a partir das amostras do espaço de transformação $DWT(m,n)$. Se as funções ψ_{mn} forem ortogonais, pode-se reconstruir $f(x)$ de maneira direta, pois a ortogonalidade de $\psi(x)$ garante que a transformação seja linearmente independente, garantindo assim a existência da transformada inversa. Pode-se notar a similaridade com a transformada de Fourier, na qual as funções básicas (senos e cossenos) também são mutuamente ortogonais.

A partir do exposto acima, de agora em diante serão discutidas apenas as *wavelets* ortogonais para representação de sinais e ainda, serão consideradas as constantes $a_0=2$ e $b_0=1$, constituindo a forma mais simples e usual da transformada *wavelet*.

Se a função $f(x)$ for considerada como sendo uma combinação linear de versões transladadas de uma função passa-baixas $\phi(x)$ no nível de dilatação m , de acordo com a equação (4.3), essa versão dilatada de $\phi(x)$ será denotada como ϕ_{mn} e

$$f(x) = \sum_n c_{mn} \phi_{mn}(x) \quad (4.4)$$

onde $\phi(x)$ é conhecida por função de escala. Confirma-se que $c_{mn} = \langle f, \phi_{mn} \rangle$, aplicando a equação (4.4) na equação (4.2):

$$c_{mn} = \langle (\sum_k c_{mk} \phi_{mk}), \phi_{mn} \rangle = \sum_k c_{mk} \langle \phi_{mk}, \phi_{mn} \rangle = c_{mn} \quad (4.5)$$

se ϕ_{mn} e ϕ_{mk} forem ortonormais.

$$\langle \phi_{mn}, \phi_{mk} \rangle = \delta(n-k) \quad (4.6)$$

$$\text{onde } \delta(n) = \begin{cases} 1, n = 0 \\ 0, n \neq 0 \end{cases}$$

Também é possível concluir que

$$\langle f, f \rangle = \sum_n c_{mn}^2 \quad (4.7)$$

Deve-se notar a semelhança com a transformada de Fourier nas transformadas direta e inversa e a preservação da energia, de maneira análoga ao teorema de Parseval [Vaidyanathan 1995].

Sendo V_0 o espaço das funções cuja base é $\phi(x)$ na dilatação $m=0$, se $\phi(x) \in V_0$, o sinal de mais baixa frequência (mais dilatado) $\phi(x/2) \in V_1$, onde V_1 forma um sub-espaço

de V_0 . Continuando o processo, um conjunto de sub-espacos pode ser construído com sinais de frequência cada vez mais baixa e banda passante mais estreita, correspondendo aos sinais sucessivamente dilatados. Dessa forma, quanto maior o valor de m , mais restrito o sub-espaco correspondente.

Sinais contidos em V_1 também estarão em V_0 , mas não necessariamente a recíproca será verdadeira. Um sinal que pertence a V_0 mas não a V_m somente pode ser aproximado pela equação (4.4).

Grandes valores para m levam a aproximações mais grosseiras da função e com a diminuição de m obtém-se cada vez melhores aproximações do sinal, revelando maiores detalhes, até que $m=0$. Esse sinal, se estiver em V_0 , pode ser reconstruído perfeitamente. Essa técnica de aproximações sucessivas é conhecida como análise em multiresolução [Gopinath 1993, Mallat 1989].

A reconstrução de $f(x)$ no nível de dilatação $m-1$, denotada por f_{m-1} , é dado pela equação $f_{m-1} = \sum_n c_{(m-1)n} \phi_{(m-1)n}$, mas pode ser expressa também em termos da resolução mais baixa f_m e um termo de diferença:

$$f_{m-1} = \sum_n c_{mn} \phi_{mn} + \sum_k d_{mk} \psi_{mk} \quad (4.8)$$

A função $\psi(x)$ é conhecida como *wavelet* mãe e ψ_{mk} é obtida a partir de $\psi(x)$ conforme a equação (4.3).

Requisitando a ortonormalidade entre ϕ_{mn} e ψ_{mn} e também entre ψ_{mn} e ψ_{mk} , além da ortonormalidade entre ϕ_{mn} e ϕ_{mk} , já mencionada, temos:

$$\begin{aligned} \langle \phi_{mn}, \psi_{mk} \rangle &= 0 \\ \langle \psi_{mn}, \psi_{mk} \rangle &= \delta(n-k) \\ \langle \phi_{mn}, \phi_{mk} \rangle &= \delta(n-k) \end{aligned} \quad (4.9)$$

Dessa forma, o cálculo dos parâmetros c_{mn} e d_{mn} se torna simples:

$$\begin{aligned} c_{mn} &= \langle f_{m-1}, \phi_{mn} \rangle = \langle f_m, \phi_{mn} \rangle = \langle f, \phi_{mn} \rangle \\ d_{mn} &= \langle f_{m-1}, \psi_{mn} \rangle = \langle f_m, \psi_{mn} \rangle = \langle f, \psi_{mn} \rangle \end{aligned} \quad (4.10)$$

As equações acima resultam da ortogonalidade nas equações (4.4) e (4.8) e a última igualdade para c_{mn} corresponde à definição original da equação (4.2). De maneira análoga, d_{mn} pode ser definido considerando que f_{m-1} é a melhor aproximação de $f(x)$ em V_{m-1} .

Estendendo a equação (4.8), temos:

$$f_{m-1} = \sum_n c_{mn} \phi_{mn} + \sum_k d_{mk} \psi_{mk} + \sum_k d_{(m-1)k} \psi_{(m-1)k} + \dots + \sum_k d_{(m-i+1)k} \psi_{(m-i+1)k} \quad (4.11)$$

Portanto, o sinal f_0 é dado por um termo de baixa resolução f_m , determinado por c_{mn} e ϕ_{mn} , e por m termos de diferença, determinados por d_{jk} e ψ_{jk} , de modo a obter resoluções progressivamente mais altas. Em termos dos espaços gerados, partindo-se do espaço V_m e adicionando-se sucessivamente os espaços W_i , chega-se ao espaço V_0 . A base do espaço W_j é $\psi(x)$ no nível de dilatação j .

Se a função $f(x) \in V_m$, então $f(x) = f_{m-1} = f_m$ e $d_{mn} = 0$, de acordo com a equação (4.8) e portanto,

$$d_{mn} = \sum_k c_{(m+1)k} \langle \phi_{(m+1)k}, \psi_{mn} \rangle + \sum_k d_{(m+1)k} \langle \psi_{(m+1)k}, \psi_{mn} \rangle = 0 \quad (4.12)$$

e sendo assim, $\langle \phi_{(m+1)k}, \psi_{mn} \rangle = 0$ e $\langle \psi_{(m+1)k}, \psi_{mn} \rangle = 0$, e por extensão através da equação (4.11), temos:

$$\langle \phi_{(m+j)k}, \psi_{mn} \rangle = 0, j \geq 0 \quad (4.13)$$

$$\langle \psi_{(m+j)k}, \psi_{mn} \rangle = 0, j \neq 0$$

confirmando dessa maneira a ortogonalidade dos espaços W_i .

Considerando-se que $\phi(x) \in V_0 \subset V_1$ por definição, naturalmente $\phi(x)$ pode ser transformada como qualquer outra função e expressa no nível de transformação $m=-1$:

$$\phi(x) = \sqrt{2} \sum_k c_k \phi(2x - k) \quad (4.14)$$

onde

$$c_k = \sqrt{2} \langle \phi(x), \phi(2x - k) \rangle$$

De maneira análoga, $\psi(x)$ também pode ser transformada, pois $\psi(x) \in W_0 \subset V_{-1}$. Para manter as condições de ortogonalidade dadas acima, $\psi(x)$ pode ser dada em função dos mesmos coeficientes c_k :

$$\psi(x) = \sum_k \sqrt{2} (-1)^k c_{1-k} \phi(2x - k) \quad (4.15)$$

A equação acima garante a ortogonalidade entre $\phi(x)$ e $\psi(x-n)$.

Até o momento foram discutidas as condições para funções de escala [$\phi(x)$] e *wavelets* mães [$\psi(x)$] adequadas. Se uma função $\phi(x)$ bem comportada for encontrada, então a correspondente $\psi(x)$ também existe.

Decomposição e reconstrução

No nível de dilatação $m=0$, $c_{0n} = \langle f, \phi_{0n} \rangle$ e $f_0 = \sum_n c_{0n} \phi_{0n}$. No próximo nível, $m=1$, $c_{1k} = \langle f, \phi_{1k} \rangle = \langle f_0, \phi_{1k} \rangle = \sum_n c_{0n} \langle \phi_{0n}, \phi_{1k} \rangle$. O último produto interno pode ser expandido como:

$$\begin{aligned} \langle \phi_{0n}, \phi_{1k} \rangle &= 2^{-1/2} \int_{-\infty}^{\infty} \phi(x-n) \phi(x/2-k) dx \\ &= 2^{1/2} \int_{-\infty}^{\infty} \phi(2x-(n-2k)) \phi(x) dx = h(n-2k) \end{aligned} \quad (4.16)$$

onde $h(n-2k)$ são os coeficientes c_{n-2k} da equação (14). Logo:

$$c_{1k} = \sum_n c_{0n} h(n-2k) \quad (4.17)$$

O resultado acima constitui a base do algoritmo de transformação. Os coeficientes de transformação no nível de dilatação $m=1$ podem ser calculados pela transformação discreta dos coeficientes no nível de dilatação $m=0$, sem a necessidade de retornar ao sinal contínuo original. De modo análogo:

$$c_{1k} = \langle f, \psi_{1k} \rangle = \sum_n c_{0n} \langle \phi_{0n}, \psi_{1k} \rangle = \sum_n c_{0n} g(n-2k) \quad (4.18)$$

onde

$$g(n-2k) = \langle \phi_{0n}, \psi_{1k} \rangle = 2^{-1/2} \int_{-\infty}^{\infty} \phi(x-n) \psi(x/2-k) dx \quad (4.19)$$

Os resultados acima podem ser estendidos para os próximos níveis de dilatação mais altos, onde os mesmos $h(n)$ e $g(n)$ se aplicam, considerando a equação (4.16):

$$\langle \phi_{mn}, \phi_{(m+1)k} \rangle = \langle \phi_{0n}, \phi_{1k} \rangle = h(n-2k) \quad (4.20)$$

E, de maneira equivalente:

$$\langle \phi_{mn}, \psi_{(m+1)k} \rangle = \langle \phi_{0n}, \psi_{1k} \rangle = g(n-2k) \quad (4.21)$$

Portanto, o algoritmo para a transformada *wavelet* discreta é dado por:

$$c_{mk} = \sum_n c_{(m-1)n} h(n-2k) \quad (4.22)$$

$$d_{mk} = \sum_n c_{(m-1)n} g(n-2k)$$

Assim, o sinal contínuo original somente é utilizado no nível de dilatação mais baixo $m=0$ para se obter c_{0n} . A partir de então todas as transformações são realizadas nos coeficientes obtidos no nível de dilatação anterior, sendo os resultados idênticos aos que seriam obtidos ao se aplicar a transformada contínua à função original.

O algoritmo de reconstrução (transformada inversa) pode ser encontrado de maneira análoga, através da equação (4.8):

$$fm = \sum_k c_{(m+1)k} \phi_{(m+1)k} + \sum_k d_{(m+1)k} \psi_{(m+1)k}$$

e, de acordo com a equação (4.10), $c_{mn} = \langle f_m, \phi_{mn} \rangle$. Logo:

$$\begin{aligned} c_{mn} &= \sum_k c_{(m+1)k} \langle \phi_{mn}, \phi_{(m+1)k} \rangle + \sum_k d_{(m+1)k} \langle \phi_{mn}, \psi_{(m+1)k} \rangle = \\ &= \sum_k c_{(m+1)k} h(n-2k) + \sum_k d_{(m+1)k} g(n-2k) \end{aligned} \quad (4.23)$$

Dessa forma, os coeficientes c_{mn} podem ser obtidos a partir dos coeficientes $c_{(m+1)n}$ e $d_{(m+1)n}$ do próximo nível superior de dilatação ($m+1$), através da filtragem digital com as funções $h(n-2k)$ e $g(n-2k)$, respectivamente. É importante reparar que na transformada inversa os somatórios são função de k , enquanto que na transformada direta são função de n .

Para a reconstrução do sinal original, os coeficientes c e d do nível de dilatação m são combinados para encontrar os coeficientes $c_{(m-1)k}$ e então, juntamente com $d_{(m-1)k}$, encontram-se os coeficientes $c_{(m-2)k}$. O processo continua até a obtenção dos coeficientes c_{0k} . Sendo assim, o algoritmo de reconstrução requer os coeficientes c do último nível de dilatação (m), juntamente com os coeficientes d de todos os outros níveis mais baixos ($m-i$), sendo $0 < i < m$. As amostras do sinal no nível $m=0$ serão então idênticas às do sinal original.

Pode ser verificado que o número de coeficientes necessários para a reconstrução completa do sinal é N , o mesmo número de amostras em c_{0n} . Isso permanece sendo verdade, não importando se a transformação é levada até o último nível de dilatação possível, quando existe apenas um coeficiente c e um coeficiente d . Cada um dos coeficientes de transformação c_{mk} e d_{mk} é formado pela combinação linear das amostras de c_{0k} e gera um sistema linear de equações, as quais são linearmente independentes, uma vez que as *wavelets* são ortogonais. Na transformada inversa esse sistema é resolvido e, evidentemente, para N equações c_{0k} são necessários N coeficientes. Mais de N coeficientes causariam redundância de informação ou pior, informação contraditória, e menos de N não possibilitariam a solução.

De forma a reconstruir o sinal contínuo $f(x)$, a versão contínua da transformada *wavelet* é aplicada aos coeficientes c_{0k} . O sinal reconstruído dessa maneira é preciso se $f(x)$ faz parte do espaço V_0 , senão o sinal reconstruído será f_0 , a melhor aproximação

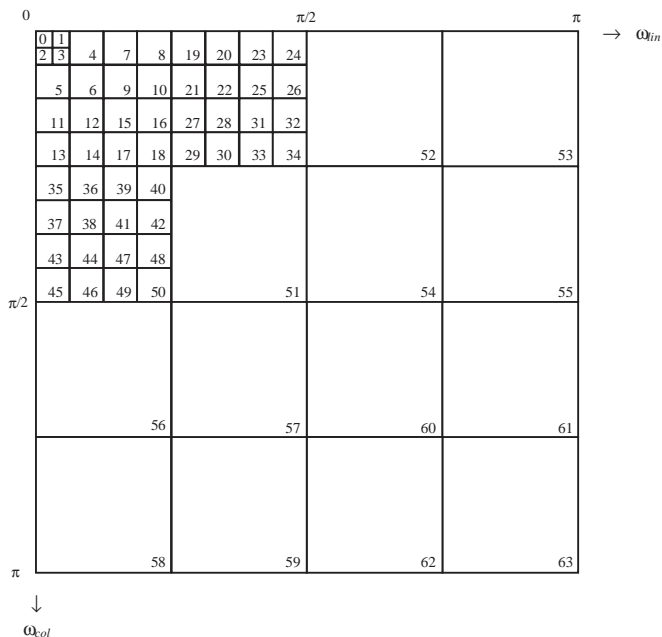
média quadrática através de $\phi(x)$ no nível $m=0$ de dilatação. Entretanto, o nível $m=0$ é uma escolha arbitrária e os coeficientes c_{-1k} poderiam ser calculados continuando-se a aplicar a transformada discreta inversa, assumindo-se que $f(x) \in V_0$ e estipulando $d_{0k}=0$. Dessa forma versões discretizadas do sinal contínuo original serão reconstruídas com resolução cada vez maior em termos de amostragem, mas sem conteúdo adicional de informação. Essas funções discretizadas convergem para a função contínua f_0 , com a redução na escala de tempo das correspondentes funções ϕ_{mm} , até que se tornem impulsos no limite. É interessante observar que os sinais pertencentes a V_0 não são limitados em frequência, pois as funções de suporte finito $\phi(x)$ possuem largura de faixa em frequência ilimitada.

4.2 Padrão WSQ de Compressão de Imagens de Impressões Digitais

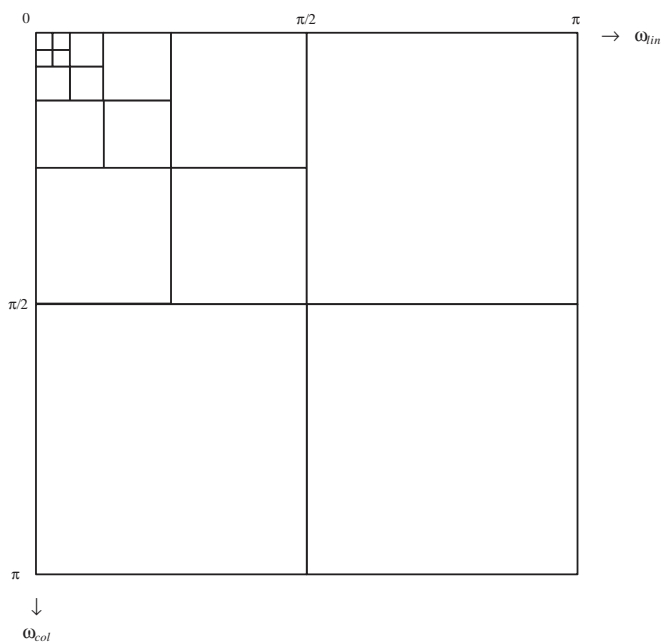
O FBI formulou um padrão para digitalização e compressão de imagens de impressões digitais em níveis de cinza. A compressão das imagens digitalizadas baseia-se na quantização escalar uniforme da decomposição em sub-bandas através da transformada *wavelet* discreta, denominada quantização escalar *wavelet* (*wavelet scalar quantization* - WSQ).

A padronização do método de compressão de imagens de impressões digitais [FBI 1993] prevê os detalhes pertinentes às etapas do sistema completo, incluindo os algoritmos da transformada *wavelet* discreta, quantização e compressão de dados, bem como as estruturas de dados a serem utilizadas no sistema. Serão abordados neste trabalho apenas os detalhes relativos especificamente à implementação da transformada *wavelet*, uma vez que as demais etapas do método WSQ não são de interesse direto para o sistema de reconhecimento de impressões digitais aqui proposto, pois tratam dos detalhes pertinentes ao algoritmo de compactação de dados propriamente dito.

O método WSQ foi originalmente especificado para uso em imagens de impressões digitais com resolução espacial de 500 pontos por polegada e 256 níveis de cinza, a serem decompostas em 64 sub-bandas de frequência espacial. Para essa decomposição, utiliza-se um banco de filtros de reconstrução perfeita, implementados bidimensionalmente com até cinco níveis de cascadeamento [Brislawn 1996]. A ilustração da estrutura dessa decomposição é dada na Figura 4.1a.



(a)



(b)

Figura 4.1 - Estrutura de sub-bandas: (a) método WSQ; (b) oitavas.

A estrutura de sub-bandas da Figura 4.1a não corresponde à estrutura de sub-bandas dividida em oitavas normalmente utilizada em decomposições *wavelet*, ilustrada na Figura 4.1b. São utilizadas bandas passantes de largura diferenciada, sendo que as baixas e médias frequências são particionadas em bandas passantes mais estreitas, de

modo a proporcionar uma alocação de bits mais flexível no processo de compressão em diferentes partes do espectro de frequência espacial [Brislaw 1996].

Quanto aos filtros digitais a serem utilizados para o cálculo da transformada *wavelet*, a especificação possibilita o uso de diferentes tipos de bancos de filtros de resposta finita ao impulso (*finite impulse response* - FIR) com até 32 coeficientes. Para a primeira geração de codificadores, foi sugerido o uso de um filtro que corresponde a uma família de *wavelets* biortogonais simétricas, cujos coeficientes aproximados são dados na Tabela 4.1.

h(-4)	h(-3)	h(-2)	h(-1)	h(0)	h(1)	h(2)	h(3)	h(4)
0,0378	-0,0238	-0,1106	0,3774	0,8527	0,3774	-0,1106	-0,0238	0,0378

Tabela 4.1 - Coeficientes do filtro de primeira geração.

O filtro representado pelos coeficientes dados na Tabela 4.1 origina a *wavelet* mãe que pode ser observada na Figura 4.2.

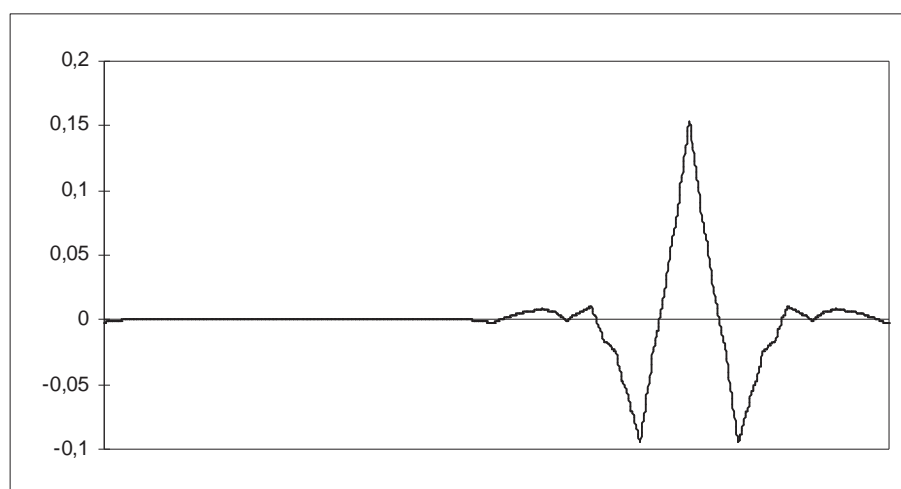


Figura 4.2 - Wavelet mãe originada pelo filtro de primeira geração.

Segundo [Brislaw 1996], um detalhe importante no uso de bancos de filtros para codificação de imagens é a solução adotada para as condições de contorno referentes às bordas das imagens. Enquanto pode-se simplesmente considerar as imagens como sendo periódicas e aplicar os filtros por convolução circular, resultados melhores são atingidos quando são impostas condições de contorno simétricas e aplicados os filtros às imagens simetricamente estendidas. A especificação do método WSQ prevê o

uso dessa solução para a condição de contorno a ser utilizada nas bordas das imagens, conhecida como transformação de extensão simétrica (*symmetric extension transform - SET*), que pode ser observada no exemplo da Figura 4.3.

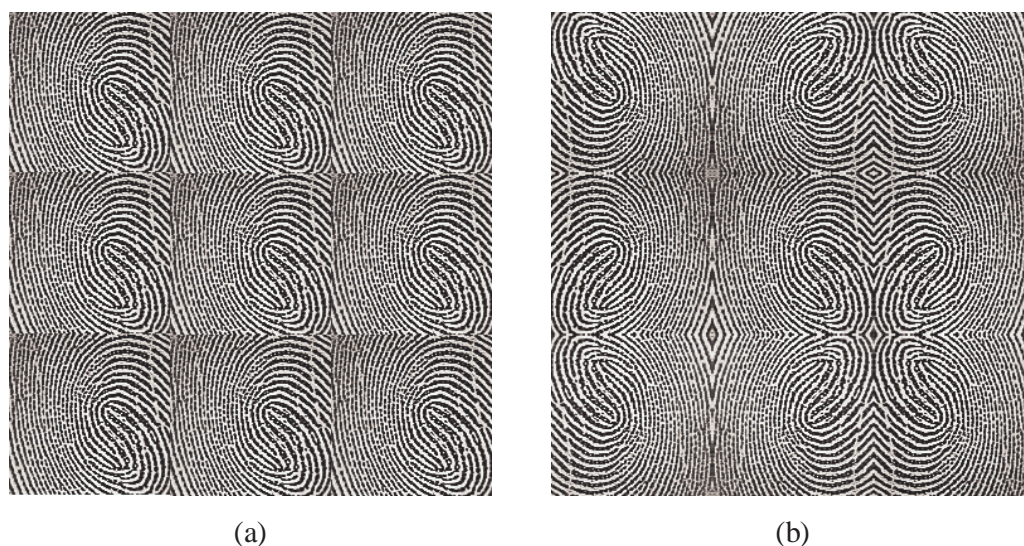


Figura 4.3 - Exemplos de condição de contorno: (a) periódica; (b) simétrica.

Finalmente, o último detalhe da especificação do FBI de relevância para o método proposto de extração de características, diz respeito à normalização das imagens a serem submetidas à transformada *wavelet*. Antes de ser decomposta, uma imagem deve ser normalizada segundo a fórmula a seguir:

$$I'(m,n) = \frac{I(m,n) - M}{R} \quad (4.24)$$

onde M é o valor médio da imagem e R é dado pela fórmula:

$$R = \frac{1}{128} \max(I_{max} - M, M - I_{min}) \quad (4.25)$$

I_{min} e I_{max} são, respectivamente, os valores de pixels mínimo e máximo na imagem $I(m,n)$. O efeito principal dessa normalização é proporcionar uma média aproximadamente nula nas sub-bandas de decomposição de mais baixa frequência [Brislaw 1996, FBI 1993].

4.3 Extração de Características das Imagens

Depois de submetidas às etapas de pré-processamento e segmentação, as imagens resultantes foram normalizadas segundo a equação (4.24). Posteriormente, as imagens

normalizadas foram submetidas à decomposição *wavelet* conforme a estrutura de sub-bandas definida pelo padrão WSQ, ilustrada na Figura 4.1.

Através de comparações entre decomposições realizadas em imagens normalizadas e não normalizadas, foi possível comprovar o efeito resultante da normalização nas sub-bandas de mais baixa frequência. Nas decomposições *wavelet* efetuadas em imagens não normalizadas, os coeficientes das sub-bandas mais baixas resultaram em valores muito superiores aos das demais sub-bandas, prejudicando a obtenção de uma certa uniformidade na ordem de grandeza dos coeficientes e confirmando a necessidade da normalização das imagens para melhores resultados.

Quanto à implementação da transformada *wavelet* em si, os filtros utilizados para esse fim seguiram as condições de contorno simétricas recomendadas pelo FBI. Uma justificativa especial para essa decisão é o fato da maioria dos problemas encontrados na validação dos algoritmos de compressão possuir relação direta com a correta implementação das condições de contorno simétricas dos filtros [Brislaw 1996]. Por outro lado, foram realizados experimentos com diferentes filtros, além do sugerido na primeira geração de codificadores WSQ. Maiores detalhes sobre a escolha do filtro e, conseqüentemente, da base *wavelet* a ser empregada no método de extração de características aqui proposto serão abordados mais adiante.

Para a construção de vetores de características a partir da decomposição *wavelet* havia o objetivo de se utilizar no máximo 64 coeficientes, uma vez que estes vetores determinam a dimensão do espaço de entrada das RNA a serem utilizadas e que dimensões muito elevadas poderiam resultar no comprometimento da velocidade de treinamento e resposta dos algoritmos.

Foram iniciadas então as tentativas de se escolher 64 entre os 16384 coeficientes resultantes da decomposição *wavelet* que representassem adequadamente as classes das impressões digitais. Convém lembrar que nessa etapa preliminar de experimentos foi utilizado apenas o primeiro conjunto de dados construído, conforme anteriormente descrito no Capítulo 2.

Numa primeira abordagem, baseada no conceito empregado em compressão de dados, foram selecionados e ordenados os 64 maiores coeficientes resultantes da transformada *wavelet* utilizada. Entretanto, não houve êxito nessa primeira tentativa, uma vez que os vetores de características resultantes desse método foram muito semelhantes, mesmo para imagens bastante distintas.

Surgiu então a idéia de se computar o vetor de características levando-se em consideração não somente o valor dos 64 maiores coeficientes da transformada, mas também suas respectivas coordenadas no plano de transformação. Idealizou-se uma equação simples, na forma de média ponderada entre os elementos envolvidos, dando-se maior peso ao valor dos coeficientes e pesos inferiores às suas coordenadas. Infelizmente, essa foi mais uma tentativa frustrada, resultando novamente em vetores muito semelhantes e não permitindo a adequada separação das imagens em classes coerentes.

Em seguida, foram realizadas as mesmas tentativas anteriores, porém utilizando-se os maiores coeficientes da análise *wavelet* por sub-banda, resultando em 64 coeficientes correspondentes a cada sub-banda específica da estrutura determinada pela especificação WSQ. Essas novas tentativas foram efetuadas inicialmente apenas com os coeficientes e posteriormente com inclusão das coordenadas dos coeficientes no cálculo dos vetores de características. Mais uma vez houve insucesso nessas tentativas, pelos mesmos motivos das tentativas anteriores, ou seja, incoerência na separação das diferentes classes.

Ainda antes de se chegar a um método que resultasse em vetores de características aproveitáveis, houve algumas tentativas com a transformada Karhunen-Loève aplicada sobre os coeficientes resultantes da decomposição *wavelet*, infelizmente também sem sucesso.

Após todas as tentativas anteriores frustradas, finalmente foi encontrada uma solução que acabou resultando em vetores de características que permitiram uma adequada separação de classes. Os vetores de características foram montados a partir dos 64 coeficientes *wavelet* resultantes das sub-bandas 0, 1, 2 e 3 da estrutura de decomposição especificada pelo FBI e posteriormente normalizados. O sucesso dessa escolha deve-se ao fato dessas sub-bandas conterem informação compactada sobre a imagem inteira, pois representam a análise *wavelet* mais profunda efetuada nas imagens.

Em experimentos posteriores, como será visto no capítulo seguinte, procurou-se reduzir o número de coeficientes dos vetores de características para 32 (sub-bandas 0 e 3) e 16 (sub-banda 0) com o intuito de verificar o impacto resultante na etapa de classificação.

Em seguida, o algoritmo de extração de características das imagens segmentadas é apresentado em pseudo-código:

1. Ler a imagem da região de interesse
2. Gerar coeficientes do filtro FIR para a decomposição *wavelet*
3. Construir a árvore de decomposição *wavelet* segundo o método WSQ
4. Normalizar a imagem da região de interesse segundo o método WSQ
5. Efetuar a decomposição *wavelet* da imagem da região de interesse, utilizando a árvore de decomposição e o filtro FIR especificados
6. Para as sub-bandas da estrutura de decomposição do método WSQ variando de 0 a 3, varrer seqüencialmente suas linhas e colunas, obtendo seus coeficientes e gerando o vetor de características da região de interesse
7. Normalizar o vetor de características da região de interesse

Cabe aqui esclarecer que o método de avaliação da separabilidade entre as classes resultantes das tentativas descritas anteriormente baseou-se na análise estatística dos seus respectivos vetores característicos representantes de cada classe. Foram calculadas as médias das correlações entre vetores da mesma classe (como medida de avaliação intra-classe) e as médias das correlações entre vetores de diferentes classes (como medida de avaliação extra-classe), com todas as combinações de pares de vetores possíveis. Essas avaliações foram feitas com todos os cinco membros de cada uma das dez classes envolvidas na primeira base de dados construída para esse fim. Nesse estudo, cada impressão digital correspondente a cada dedo das mãos foi considerada como uma classe específica, totalizando as dez classes mencionadas. Procurou-se observar se as correlações intra-classe eram superiores às correlações extra-classe, num indicativo de adequação do método para gerar classes coerentes e ao mesmo tempo separáveis. Os resultados desse estudo estatístico poderão ser observados mais adiante nas Tabelas 4.3, 4.4, 4.5 e 4.6.

A avaliação do desempenho da técnica de extração de características de imagens de impressões digitais a partir da transformada *wavelet* foi realizada com quatro bases *wavelet* distintas, de modo a comparar e selecionar a base de representação mais adequada. Os testes iniciais foram implementados adotando-se uma base *wavelet* de mínima assimetria e suporte compacto com momento de ordem 8 (*Symmlet8*), utilizada originalmente na implementação do método WSQ da *toolbox* WaveLab para o *software*

MatLab [Buckheit 1995]. Em seguida, utilizou-se a base sugerida pelo FBI para a primeira geração de codificadores WSQ, já abordada anteriormente. E finalmente, foram realizados experimentos com as bases *wavelet* H10 e S10, apresentadas em [Sherlock 1996a] e [Sherlock 1996b] para a compressão de imagens de impressões digitais. Os coeficientes aproximados dos filtros digitais associados a essas bases são listados na Tabela 4.2.

Base Wavelet	Coeficientes do Filtro Associado	
Symplet8	$h(-7) = 0,001890$	$h(1) = 0,777186$
	$h(-6) = -0,000303$	$h(2) = 0,481360$
	$h(-5) = -0,014952$	$h(3) = -0,061273$
	$h(-4) = 0,003809$	$h(4) = -0,143294$
	$h(-3) = 0,049137$	$h(5) = 0,007607$
	$h(-2) = -0,027219$	$h(6) = 0,031695$
	$h(-1) = -0,051946$	$h(7) = -0,000542$
	$h(0) = 0,364442$	$h(8) = -0,003382$
	FBI	$h(-4) = 0,037828$
$h(-3) = -0,023849$		$h(1) = 0,377403$
$h(-2) = -0,110624$		$h(2) = -0,110624$
$h(-1) = 0,377403$		$h(3) = -0,023849$
		$h(4) = 0,037828$
H10	$h(-4) = 0,002039$	$h(1) = 0,725212$
	$h(-3) = -0,002285$	$h(2) = 0,094588$
	$h(-2) = -0,112447$	$h(3) = -0,112447$
	$h(-1) = 0,094588$	$h(4) = -0,002285$
	$h(0) = 0,725212$	$h(5) = 0,002039$
S10	$h(-4) = 0,015089$	$h(1) = 0,815457$
	$h(-3) = 0,061312$	$h(2) = 0,067598$
	$h(-2) = -0,252348$	$h(3) = -0,252348$
	$h(-1) = 0,067598$	$h(4) = 0,061312$
	$h(0) = 0,815457$	$h(5) = 0,015089$

Tabela 4.2 - Coeficientes dos filtros associados às bases *wavelet* utilizadas.

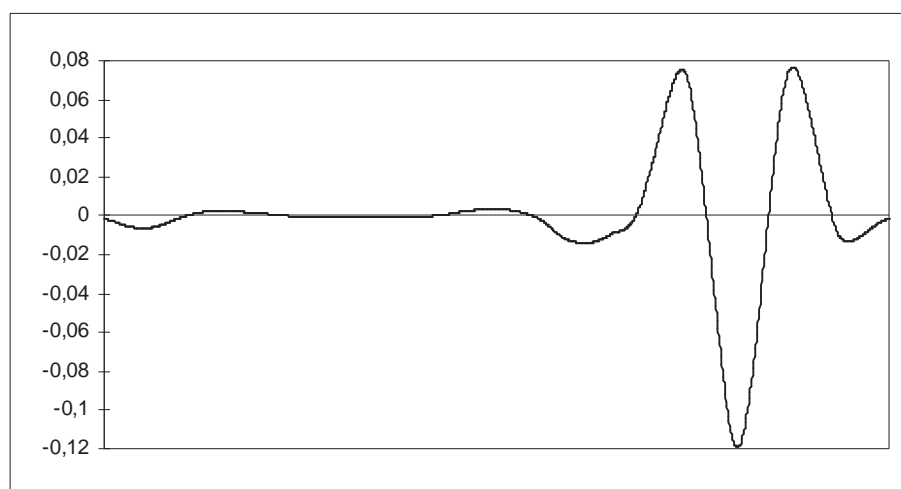


Figura 4.4 - Wavelet mãe originada pelo filtro *Symplet8*.

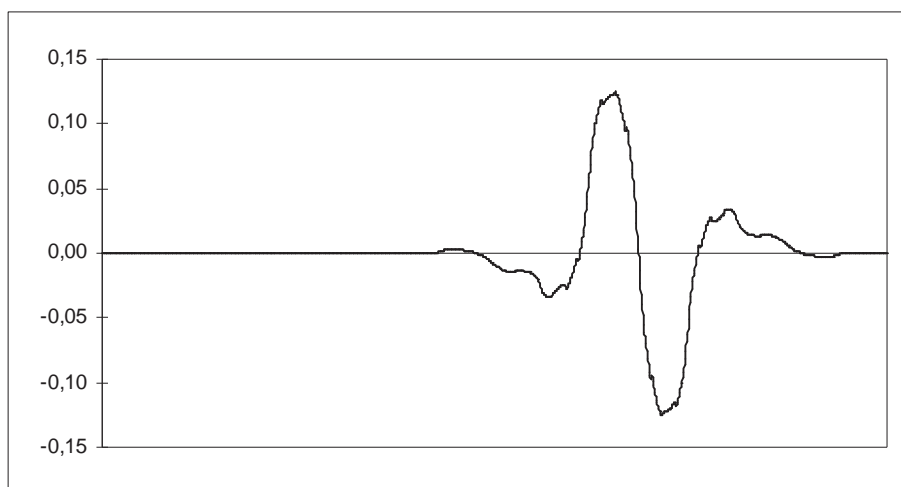


Figura 4.5 - Wavelet mãe originada pelo filtro H10.

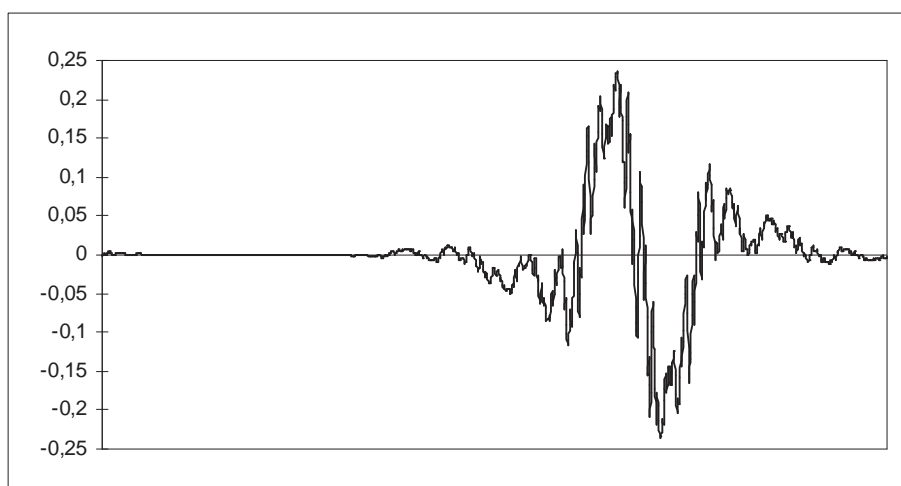


Figura 4.6 - Wavelet mãe originada pelo filtro S10.

As Figuras 4.4, 4.5 e 4.6 apresentam as *wavelets* mães originadas pelos filtros *Symmlet8*, H10 e S10, respectivamente. A *wavelet* mãe sugerida na especificação do FBI foi previamente apresentada na Figura 4.2.

Os resultados da análise estatística dos vetores de características obtidos para as dez classes de impressões digitais (f_1 a f_{10}) com as diferentes bases citadas anteriormente podem ser analisados nas Tabelas 4.3, 4.4, 4.5 e 4.6, correspondentes à base *wavelet Symmlet8*, à base do FBI para a primeira geração de codificadores WSQ, à base H10 e à base S10, respectivamente. Os valores em negrito apresentados nessas tabelas são correspondentes às avaliações correlacionais intra-classe, sendo os demais valores associados às avaliações extra-classe.

	f₁	f₂	f₃	f₄	f₅	f₆	f₇	f₈	f₉	f₁₀
f₁	0,8796	0,1075	0,1316	0,0600	0,3552	0,3791	0,2762	0,1537	0,1063	0,3815
f₂	0,1075	0,8064	0,4995	0,2847	0,1505	0,1024	0,3306	0,3790	0,2826	0,1967
f₃	0,1316	0,4995	0,8167	0,4315	0,2336	0,1670	0,3705	0,3885	0,4929	0,4099
f₄	0,0600	0,2847	0,4315	0,7866	0,1883	0,3942	0,3901	0,4562	0,5091	0,3802
f₅	0,3552	0,1505	0,2336	0,1883	0,7623	0,3713	0,2531	0,0952	0,0932	0,2450
f₆	0,3791	0,1024	0,1670	0,3942	0,3713	0,9692	0,7862	0,2425	0,4386	0,5553
f₇	0,2762	0,3306	0,3705	0,3901	0,2531	0,7862	0,9733	0,4760	0,6438	0,5984
f₈	0,1537	0,3790	0,3885	0,4562	0,0952	0,2425	0,4760	0,7739	0,5231	0,1746
f₉	0,1063	0,2826	0,4929	0,5091	0,0932	0,4386	0,6438	0,5231	0,8226	0,5358
f₁₀	0,3815	0,1967	0,4099	0,3802	0,2450	0,5553	0,5984	0,1746	0,5358	0,7053

Tabela 4.3 - Resultados da análise estatística dos vetores de características obtidos com a base *Symmlet8*.

	f₁	f₂	f₃	f₄	f₅	f₆	f₇	f₈	f₉	f₁₀
f₁	0,8712	0,0921	0,1250	0,0355	0,2948	0,3654	0,2689	0,1646	0,0894	0,3792
f₂	0,0921	0,8005	0,4888	0,2859	0,0947	0,0979	0,3308	0,4218	0,2844	0,1783
f₃	0,1250	0,4888	0,8001	0,4386	0,2236	0,1766	0,3886	0,3783	0,4933	0,4018
f₄	0,0355	0,2859	0,4386	0,7747	0,1638	0,3784	0,4122	0,4653	0,4947	0,3580
f₅	0,2948	0,0947	0,2236	0,1638	0,7367	0,3199	0,1853	0,0685	0,0422	0,2030
f₆	0,3654	0,0979	0,1766	0,3784	0,3199	0,9687	0,7732	0,2522	0,4436	0,5415
f₇	0,2689	0,3308	0,3886	0,4122	0,1853	0,7732	0,9719	0,4651	0,6577	0,6002
f₈	0,1646	0,4218	0,3783	0,4653	0,0685	0,2522	0,4651	0,7977	0,5359	0,1553
f₉	0,0894	0,2844	0,4933	0,4947	0,0422	0,4436	0,6577	0,5359	0,8254	0,5434
f₁₀	0,3792	0,1783	0,4018	0,3580	0,2030	0,5415	0,6002	0,1553	0,5434	0,7019

Tabela 4.4 - Resultados da análise estatística dos vetores de características obtidos com a base do FBI para a primeira geração de codificadores WSQ.

	f₁	f₂	f₃	f₄	f₅	f₆	f₇	f₈	f₉	f₁₀
f₁	0,8551	0,0929	0,1338	0,0876	0,3494	0,3628	0,2387	0,1434	0,0862	0,3580
f₂	0,0929	0,7891	0,4348	0,2722	0,1569	0,1001	0,3231	0,3660	0,2838	0,1723
f₃	0,1338	0,4348	0,7784	0,3876	0,2148	0,1370	0,3201	0,3318	0,4137	0,3939
f₄	0,0876	0,2722	0,3876	0,7775	0,1409	0,3504	0,3520	0,3899	0,4490	0,3623
f₅	0,3494	0,1569	0,2148	0,1409	0,7452	0,3408	0,2140	0,1136	0,0518	0,1862
f₆	0,3628	0,1001	0,1370	0,3504	0,3408	0,9620	0,7791	0,2225	0,3999	0,5372
f₇	0,2387	0,3231	0,3201	0,3520	0,2140	0,7791	0,9671	0,4597	0,6042	0,5612
f₈	0,1434	0,3660	0,3318	0,3899	0,1136	0,2225	0,4597	0,7535	0,4584	0,1489
f₉	0,0862	0,2838	0,4137	0,4490	0,0518	0,3999	0,6042	0,4584	0,8051	0,4945
f₁₀	0,3580	0,1723	0,3939	0,3623	0,1862	0,5372	0,5612	0,1489	0,4945	0,6936

Tabela 4.5 - Resultados da análise estatística dos vetores de características obtidos com a base H10.

	f_1	f_2	f_3	f_4	f_5	f_6	f_7	f_8	f_9	f_{10}
f_1	0,7090	0,1319	0,2101	0,1071	0,3104	0,3137	0,2262	0,0593	0,1258	0,2737
f_2	0,1319	0,6312	0,2889	0,2631	0,1828	0,1607	0,2561	0,2193	0,2483	0,1508
f_3	0,2101	0,2889	0,5526	0,1970	0,2129	0,0912	0,1506	0,1785	0,2011	0,2115
f_4	0,1071	0,2631	0,1970	0,6742	0,1698	0,2651	0,2963	0,3583	0,3029	0,2440
f_5	0,3104	0,1828	0,2129	0,1698	0,6640	0,2734	0,1529	0,1153	0,0355	0,1378
f_6	0,3137	0,1607	0,0912	0,2651	0,2734	0,8794	0,7133	0,2284	0,3586	0,4162
f_7	0,2262	0,2561	0,1506	0,2963	0,1529	0,7133	0,9087	0,3951	0,5101	0,3777
f_8	0,0593	0,2193	0,1785	0,3583	0,1153	0,2284	0,3951	0,6310	0,3055	0,1184
f_9	0,1258	0,2483	0,2011	0,3029	0,0355	0,3586	0,5101	0,3055	0,6796	0,3379
f_{10}	0,2737	0,1508	0,2115	0,2440	0,1378	0,4162	0,3777	0,1184	0,3379	0,4828

Tabela 4.6 - Resultados da análise estatística dos vetores de características obtidos com a base S10.

A partir da análise dos dados das Tabelas 4.3, 4.4, 4.5 e 4.6 verifica-se que os coeficientes de correlação entre membros da mesma classe foram sempre maiores que os coeficientes entre membros de classes diferentes, como era desejado. Pode-se observar também que em todas as situações o pior caso de similaridade entre classes ocorreu entre as classes f_6 e f_7 . Apesar disso, essas classes ainda foram distinguíveis entre si, quando submetidas à classificação por RNA na etapa seguinte dos experimentos.

Primeiramente, optou-se pelo uso da base *wavelet Symmlet8* para os demais testes a serem realizados com a segunda base de dados, montada apenas com imagens de impressões digitais de polegares. O principal motivo para a escolha da base *Symmlet8* havia sido o fato da análise estatística dos vetores resultantes dessa decomposição terem apresentado as maiores correlações intra-classe em relação às demais.

Por outro lado, adotando como critério de avaliação do desempenho das bases *wavelet* na extração de características a razão entre as correlações intra e extra-classe, conclui-se de que a base H10 seria mais adequada. Foram utilizados dois critérios de avaliação para se chegar a essa conclusão: a média das razões entre as correlações intra-classe e a média das correlações extra-classe, conforme a equação (4.26); e o mínimo entre as razões entre as correlações intra-classe e o máximo entre as correlações extra-classe, conforme a equação (4.27).

$$C_{med} = \text{med} \left(\frac{c_i}{\text{med}(c_e)} \right) \quad (4.26)$$

$$C_{minmax} = \min \left(\frac{c_i}{\max(c_e)} \right) \quad (4.27)$$

Os resultados obtidos após o emprego desses critérios de avaliação são apresentados na Tabela 4.7.

	Symmet8	FBI	H10	S10
C_{med}	2,6404	2,7422	2,8165	2,8753
C_{minmax}	1,1786	1,1694	1,2348	1,1600

Tabela 4.7 - Resultados dos critérios de avaliação do desempenho das bases *wavelet* na extração de características.

Através da análise dos resultados da Tabela 4.7, verifica-se que a base H10 apresenta o maior coeficiente C_{minmax} e o segundo maior coeficiente C_{med} em relação às outras bases *wavelet* utilizadas. Esses são fortes indicativos da sua melhor aplicabilidade ao algoritmo de extração de características em relação às demais, segundo os critérios estabelecidos.

Uma vez determinada a melhor base *wavelet* para a extração de características (H10), foi possível aplicar a técnica às imagens pré-processadas e segmentadas do segundo conjunto de dados e dar início aos experimentos da etapa de classificação utilizando algoritmos de redes neurais artificiais. Deve-se salientar aqui que os vetores de características obtidos para o primeiro conjunto de dados também foi aproveitado para testes com redes neurais, visando a confirmação da metodologia de avaliação estatística empregada para validação do método proposto de extração de características.

CAPÍTULO 5

CLASSIFICAÇÃO DOS PADRÕES DE IMPRESSÕES DIGITAIS

As RNA têm se mostrado altamente capazes de solucionar problemas de classificação não-linear em presença de distorções e ruído, como o caso de classificação de impressões digitais.

Muitas hipóteses simultâneas podem ser analisadas pelas RNA, utilizando para isso camadas paralelas compostas de diversos elementos computacionais (neurônios) conectados entre si através de pesos variáveis w_i , que podem ser excitadores (positivos) ou inibidores (negativos). Tais estruturas procuram atingir bons desempenhos através da densa interconexão entre seus elementos constituintes. O mais simples dos neurônios realiza uma soma ponderada de N entradas e passa o seu resultado por uma função de ativação não-linear. Cada neurônio é caracterizado por um nível de decisão (ou *offset*) θ , e pelo tipo de função de ativação [Lippmann 1987, Rumelhart 1994]. Um exemplo de representação gráfica de uma RNA pode ser observado na Figura 5.1.

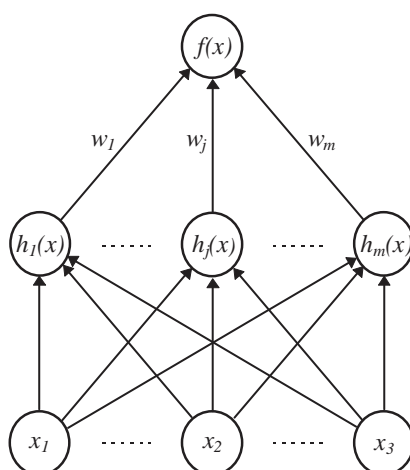


Figura 5.1 - Exemplo de RNA.

Os modelos conexionistas são especificados pela topologia da RNA, pela função de ativação dos seus neurônios e pelos algoritmos de aprendizado, os quais especificam um conjunto inicial de pesos e indicam como esses pesos devem ser adaptados durante o

uso para melhorar o seu desempenho. Após o estudo de algumas das topologias existentes, a que se mostra mais adequada para a solução do problema de classificação de padrões de impressões digitais é a *feed-forward*, em particular as redes perceptron multicamadas (*multi-layer perceptron*) e as redes com funções básicas radiais (*radial basis functions networks*), principalmente pelo fato de constituírem classificadores não-lineares com treinamento supervisionado.

5.1 Perceptrons Multicamadas

Perceptrons multicamadas contêm uma ou mais camadas de neurônios entre as suas camadas de entrada e saída. Tais camadas adicionais contêm neurônios escondidos que não são diretamente conectados aos neurônios de entrada e de saída. Os algoritmos de treinamento utilizados realizam uma técnica de otimização não-linear baseada em técnicas de gradiente descendente ou, em outras palavras, na retropropagação do erro existente na saída da RNA. Apesar de não se poder garantir que tais algoritmos de treinamento convergem sempre, estes têm se mostrado satisfatórios para muitos problemas de grande interesse.

Os perceptrons multicamadas são limitados pelas funções de ativação utilizadas nos seus neurônios. Se os mesmos fossem constituídos por funções lineares, uma RNA de camada única com pesos de conexão apropriados poderia reproduzir exatamente os resultados fornecidos por qualquer RNA multicamadas [Lippmann 1987]. As capacidades dos perceptrons que contêm uma, duas ou três camadas diferem quanto aos tipos de regiões de decisão formadas. Um perceptron de camada simples forma regiões de decisão separadas por um hiperplano, portanto sendo capaz de resolver apenas problemas linearmente separáveis.

Perceptrons de duas camadas podem formar qualquer região convexa, possivelmente aberta, no espaço gerado pelos vetores de entrada. Regiões convexas são formadas por regiões de hiperplanos, que por sua vez são formadas por cada neurônio na primeira camada da RNA. Cada neurônio na primeira camada se comporta como um perceptron simples e ativará sua saída apenas nos pontos localizados em um dos lados do hiperplano gerado pelos seus pesos e seu limiar. Os neurônios de saída proporcionam a região de decisão final, que é a interseção de todas as regiões delimitadas pelos hiperplanos formados pela primeira camada. Portanto, as regiões convexas têm no máximo tantos lados quantos forem os neurônios na primeira camada.

Um perceptron de três camadas pode formar regiões de decisão arbitrariamente complexas e consegue então separar classes entrelaçadas entre si. A região de decisão desejada pode ser dividida em pequenos hipercubos. Cada hipercubo requer $2N$ neurônios na primeira camada, um para cada lado do hipercubo, e um neurônio na segunda camada para realizar a composição das saídas da primeira camada. As saídas da segunda camada somente estarão ativas para entradas que se localizem no interior do seu respectivo hipercubo. Os hipercubos são associados às regiões de decisão adequadas pela conexão das saídas dos neurônios da segunda camada apenas ao neurônio de saída correspondente à região de decisão à qual aquele hipercubo pertence. Esse procedimento de construção pode ser generalizado para o uso de regiões convexas de formas arbitrárias em vez de hipercubos, e é capaz de gerar regiões não-convexas e descontínuas.

A análise que acaba de ser feita demonstra que não são necessárias mais que três camadas de neurônios para um perceptron, pois uma RNA desse tipo com três camadas já é capaz de gerar regiões de decisão arbitrariamente complexas. Também são dados indícios de como selecionar o número de neurônios a serem utilizados em perceptrons com três camadas. A quantidade na segunda camada deverá ser superior a um neurônio quando as classes forem desconexas ou entrelaçadas, sendo requerido um neurônio para cada região desconexa no pior caso. Na primeira camada, o número de neurônios deverá proporcionar pelo menos três lados para cada área convexa gerada pelos neurônios da segunda camada. Portanto, a quantidade de neurônios da primeira camada deverá ser pelo menos três vezes superior à quantidade da segunda camada [Lippmann 1987].

O que foi discutido até o momento abrange perceptrons multicamadas com uma saída quando as funções utilizadas são do tipo limitador. Um comportamento semelhante é apresentado por perceptrons multicamadas que possuem múltiplas saídas e fazem uso de funções do tipo sigmóide, para os quais o critério de decisão consiste em selecionar a saída de maior valor. Porém, seu comportamento é mais complexo devido às regiões de decisão serem formadas por curvas em vez de pequenos segmentos de retas e, sendo assim, sua análise torna-se mais complexa.

Entretanto, os perceptrons multicamadas que utilizam funções de ativação do tipo sigmóide, cuja função é diferenciável e expressa pela equação (5.1), podem ser treinados através do algoritmo *backpropagation*, que consiste em uma generalização do método dos mínimos quadrados, e utiliza uma técnica de busca por gradiente

descendente para minimizar uma função custo igual ao quadrado da diferença entre a saída atual da RNA e a saída desejada.

$$f(x) = \frac{1}{1 + e^{-(x-\theta)}} \quad (5.1)$$

Em função das entradas apresentadas, as respostas dos neurônios intermediários são calculadas através das equações (5.2) e (5.3), e as respostas dos neurônios de saída através da equação (5.4).

$$x'_j = f\left(\sum_{i=0}^{N_1-1} w_{ij} x_i - \theta_j\right), \quad 0 \leq j \leq N_1 - 1 \quad (5.2)$$

onde: x'_j corresponde às respostas dos neurônios da primeira camada, N_1 ao número de neurônios da primeira camada, w_{ij} aos pesos de conexão do i -ésimo neurônio da primeira camada ao j -ésimo neurônio de entrada, θ_j aos níveis de disparo do j -ésimo neurônio de entrada, x_i às entradas da RNA e N ao número de neurônios de entrada.

$$x''_k = f\left(\sum_{j=0}^{N_1-1} w'_{jk} x'_j - \theta'_k\right), \quad 0 \leq k \leq N_2 - 1 \quad (5.3)$$

onde: x''_k corresponde às respostas dos neurônios da segunda camada, N_2 ao número de neurônios da segunda camada, w'_{jk} aos pesos de conexão do j -ésimo neurônio da segunda camada ao k -ésimo neurônio da primeira camada e θ'_k aos níveis de disparo do k -ésimo neurônio da primeira camada.

$$y_l = f\left(\sum_{k=0}^{N_2-1} w''_{kl} x''_k - \theta''_l\right), \quad 0 \leq l \leq M - 1 \quad (5.4)$$

onde: y_l corresponde às respostas dos neurônios de saída, M ao número de neurônios de saída, w''_{kl} aos pesos de conexão do k -ésimo neurônio de saída ao l -ésimo neurônio da segunda camada e θ''_l aos níveis de disparo do l -ésimo neurônio da segunda camada.

Normalmente durante o treinamento, as saídas desejadas para todos os neurônios são tipicamente 0, a não ser que um determinado neurônio corresponda à classe à qual a entrada atual pertence, caso em que a saída desejada é 1. A RNA é treinada selecionando-se pequenos pesos e níveis de disparo ao acaso e então apresentando-se repetida e aleatoriamente todos os dados de treinamento. Os pesos são ajustados a cada tentativa usando informações adicionais que especificam a classe correta até que ocorra a convergência dos pesos e que a função custo seja reduzida a valores aceitáveis, conforme as equações (5.5) e (5.6). Um componente essencial do algoritmo é o método

iterativo que propaga termos de erro, necessários à adaptação dos pesos, a partir dos neurônios da camada de saída até os neurônios das camadas anteriores.

$$w_{ij}(t+1) = w_{ij}(t) + \alpha \delta_j x'_j \quad (5.5)$$

onde α é o fator de aprendizado ($0 < \alpha < 1$) e δ_j é um fator de erro para o neurônio j .

$$\begin{aligned} \delta_j &= y_j(1 - y_j)(d_j - y_j), \text{ se } j \text{ é um neurônio de saída} \\ \delta_j &= x'_j(1 - x'_j) \sum_k \delta_k w_{jk}, \text{ se } j \text{ é um neurônio intermediário} \end{aligned} \quad (5.6)$$

onde d_j é a saída desejada, y_j é a saída atual do neurônio j e k varia por todos os neurônios das camadas superiores à camada no neurônio j .

A convergência do algoritmo pode ser acelerada se um fator de momento for adicionado à equação (5.5):

$$w_{ij}(t+1) = w_{ij}(t) + \alpha \delta_j x'_j + \beta [w_{ij}(t) - w_{ij}(t-1)] \quad (5.7)$$

onde β é o fator de momento ($0 < \beta < 1$).

O procedimento de correção de erro é suficientemente simples em sua concepção. Durante o treinamento, as entradas são apresentadas à RNA e geram um conjunto de valores nos neurônios de saída, o qual é comparado com o conjunto de valores de saída desejado. Se houver coincidência, nenhuma alteração é feita na RNA. Caso contrário, algumas adaptações deverão ser feitas nos pesos das conexões. O problema consiste em determinar quais conexões na RNA inteira foram responsáveis pelo erro. A idéia principal consiste em se definir uma medida do desempenho geral da RNA e então encontrar um meio de otimizá-la. Pode-se penalizar um determinado neurônio proporcionalmente ao grau de interferência que o mesmo possui no erro do sistema. Ou seja, realizam-se mudanças em proporção à derivada do erro em relação aos pesos. Portanto, o problema de aprendizado resume-se a um problema de estimação de parâmetros [Rumelhart 1994].

Essa técnica tornou-se a mais popular para o treinamento das RNA e seu desempenho é geralmente bom, apesar do risco que uma procura por gradiente descendente oferece de encontrar mínimos locais em vez do mínimo global na função custo utilizada. Algumas sugestões para melhorar o desempenho e reduzir a ocorrência de mínimos locais incluem o uso de neurônios extras, a redução do termo de ganho usado para o ajuste dos pesos, e a inicialização de treinos com diferentes conjuntos de pesos aleatórios. Uma dificuldade inerente a esse algoritmo é a necessidade de apresentar todos os dados de treinamento várias vezes para que ocorra convergência

(em muitos casos, mais de 100 vezes). Apesar de terem sido propostos muitos algoritmos para acelerar a convergência, aparentemente as regiões de decisão complexas dos perceptrons multicamadas só se formam em poucas tentativas quando as classes são desconexas [Lippmann 1987].

É possível provar que qualquer função contínua de N variáveis pode ser calculada utilizando-se apenas somatórios lineares de funções não-lineares e continuamente crescentes de uma única variável. Sendo assim, pode-se interpretar que um perceptron de três camadas com $N(2N+1)$ neurônios utilizando funções não-lineares continuamente crescentes pode calcular qualquer função contínua de N variáveis. Portanto, é possível de ser criada qualquer função contínua com o grau de semelhança desejado em um classificador que utiliza um perceptron de três camadas [Lippmann 1987].

5.2 Redes com Funções Básicas Radiais

O algoritmo *backpropagation* para o perceptron multicamadas pode ser visto como a aplicação do método de otimização conhecido em estatística como aproximação estocástica. De maneira diferente, a abordagem dada pelas redes com funções básicas radiais é a de um problema de ajuste de curva em um espaço multidimensional. De acordo com esse ponto de vista, o aprendizado é equivalente a encontrar uma superfície no espaço multidimensional que proporciona a melhor aproximação dos dados de treinamento, com o critério de melhor aproximação sendo avaliado de algum modo estatístico. Dessa maneira, generalizar é equivalente a usar a superfície multidimensional encontrada para interpolar os dados de teste.

No contexto de RNA, os neurônios escondidos das camadas intermediárias fornecem um conjunto de funções que constituem uma base arbitrária para os vetores de entrada, quando esses são expandidos no espaço gerado por esses neurônios. Essas funções podem ser radiais, as quais foram primeiramente utilizadas na solução de problemas de interpolação real multivariada. Em aplicações de visão por computador, redes com funções básicas radiais foram utilizadas por [Borges 1994] para a classificação não-linear de padrões tridimensionais.

Nas redes com funções básicas radiais, a transformação do espaço de entrada para o espaço dos neurônios intermediários é não-linear, enquanto que a transformação do espaço dos neurônios intermediários para o espaço de saída é linear, pois pode-se

provar que um problema de classificação de padrões processado não-linearmente em um espaço cuja dimensão seja alta é mais provável de ser linearmente separável que em um espaço cuja dimensão seja baixa. Esse é o motivo de se usar dimensões altas para o espaço dos neurônios intermediários numa rede com funções básicas radiais [Haykin 1994].

No processo de aprendizado de uma rede com funções básicas radiais, os pesos lineares associados com os neurônios de saída tendem a se desenvolver em uma escala de tempo diferente daquelas das funções de ativação dos neurônios intermediários. Dessa maneira, as funções de ativação dos neurônios intermediários avançam lentamente, de acordo com alguma estratégia de otimização não-linear, e os pesos dos neurônios de saída ajustam-se rapidamente através de uma estratégia de otimização linear. Algo importante a ser notado nesse ponto é que as diferentes camadas de uma rede com funções básicas radiais desempenham tarefas diferentes, e portanto é razoável separar as otimizações das camadas intermediária e de saída utilizando técnicas diferentes e, talvez, operando em diferentes escalas de tempo. Existem diferentes estratégias de treinamento, as quais dependem de como os centros das funções básicas radiais são especificados. Essencialmente podem-se identificar três abordagens: 1) a seleção aleatória de centros fixos; 2) a seleção auto-organizada de centros; e 3) a seleção supervisionada de centros.

A abordagem mais simples consiste em assumir funções básicas radiais fixas, definindo as funções de ativação dos neurônios intermediários. Especificamente, a localização dos centros pode ser definida aleatoriamente no conjunto de dados de treinamento. Esse tipo de solução assume que os dados de treinamento estão distribuídos de maneira representativa para o problema específico em questão. Para as funções básicas radiais propriamente ditas, pode-se empregar funções isotrópicas de Gauss, cujos raios devem ser determinados de acordo com a extensão dos centros. A expressão matemática de uma função de Gauss é dada na equação (5.8). Os únicos parâmetros que são aprendidos no uso dessa técnica são os pesos lineares de conexão na camada de saída da RNA.

$$h(x) = \exp\left(-\frac{(x-c)^2}{r^2}\right) \quad (5.8)$$

onde c corresponde ao centro e r ao raio da função.

Uma rede com funções básicas radiais tradicional contém apenas uma camada de neurônios intermediários. As funções básicas, correspondentes aos neurônios intermediários, são alimentadas a partir do vetor de entrada e linearmente combinadas através de pesos, gerando a saída correspondente.

Se o método dos mínimos quadrados for utilizado para treinamento supervisionado de modelos neuronais lineares, temos:

$$f(\mathbf{x}) = \sum_{j=1}^m w_j h_j(\mathbf{x}) \quad (5.9)$$

onde \mathbf{x} é o vetor de entrada, $f(\mathbf{x})$ é a saída, w_j são os pesos das conexões, h_j são as funções radiais de cada neurônio intermediário e m é o número de neurônios intermediários da RNA.

Utilizando um conjunto de treinamento $\{(\mathbf{x}_i, y_i)\}$ com p elementos, o método dos mínimos quadrados minimizará a soma dos erros quadráticos relativos aos pesos do modelo:

$$S = \sum_{i=1}^p (y_i - f(\mathbf{x}_i))^2 \quad (5.10)$$

onde y_i é a saída desejada para o vetor de entrada \mathbf{x}_i .

Em outro método de aprendizado, mais complexo, as funções básicas radiais podem mover seus centros de maneira auto-organizada, enquanto os pesos lineares na camada de saída são calculados usando uma técnica de aprendizado supervisionada. Portanto a RNA passa por um processo híbrido de treinamento. O componente auto-organizado do processo serve para alocar os recursos da RNA de maneira significativa, colocando os centros das funções somente em regiões do espaço de entrada onde estão presentes os dados significativos.

Uma terceira técnica, ainda mais refinada, permite que os centros das funções básicas radiais e todos os outros parâmetros da RNA passem por um processo de treinamento supervisionado e a rede com funções básicas radiais assume a sua forma mais generalizada. Para o processo de correção do erro na saída, o algoritmo mais convenientemente utilizado é o que faz uso do mesmo processo de gradiente descendente usado para os perceptrons multicamadas, porém tal procedimento para uma rede com funções básicas radiais não envolve propagação retroativa do erro. Existem méritos práticos em permitir que os centros das funções movam-se quando é necessária

uma RNA com a mínima configuração possível, além de algumas demonstrações de desempenhos superiores apresentados por esse método [Haykin 1994].

Tanto as redes com funções básicas radiais quanto os perceptrons multicamadas são redes compostas de camadas de neurônios não-lineares, que constituem aproximadores universais. É possível portanto, encontrar redes com funções básicas radiais capazes de imitar precisamente o comportamento de um perceptron multicamadas e vice-versa. Entretanto, os dois tipos de RNA diferem em muitos aspectos. O mais importante deles talvez seja o fato dos perceptrons multicamadas gerarem aproximações globais para mapeamentos não-lineares entre entrada e saída, sendo capazes de generalizar em regiões do espaço de entrada onde poucos dados estão disponíveis. Por outro lado, as redes com funções básicas radiais que utilizam funções de ativação exponencialmente decrescentes (funções de Gauss) geram aproximações locais para mapeamentos não-lineares entre entrada e saída, resultando numa capacidade de aprendizado rápido e sensibilidade reduzida à ordem de apresentação dos dados de treinamento. No entanto, em muitos casos percebe-se que para que um mapeamento seja adequadamente representado, o número de funções básicas radiais requerido para a geração do espaço de entrada é consideravelmente grande.

Uma das grandes vantagens oferecidas pelas redes com funções básicas radiais é a possibilidade de treinamento em tempo real, pois basicamente utilizam-se de operações de álgebra linear (no caso mais simples de algoritmo de treinamento), ao contrário dos perceptrons multicamadas. Outro detalhe importante nesse tipo de arquitetura de RNA é a sua pouca capacidade de generalização, uma vantagem para a aplicação de classificação de impressões digitais, onde a classificação errônea e altamente indesejável. Para melhor controlar a capacidade de generalização de redes com funções básicas radiais existem métodos de regressão global e local (*ridge regression*) [Bishop 1991, Orr 1996] e para otimizar a quantidade de funções básicas existem técnicas de seleção de modelos (*model selection*) [Orr 1996], que são abordadas neste trabalho.

No caso de regressão local, um termo de penalização é adicionado à soma dos erros quadráticos, minimizando a seguinte função custo:

$$C = \sum_{i=1}^p (y_i - f(\mathbf{x}_i))^2 + \lambda \sum_{j=1}^m w_j^2 \quad (5.11)$$

onde λ é o termo de penalização global da RNA.

No entanto, por causa da natureza das funções básicas radiais, o caso de regressão local, com múltiplos termos de penalização, executa uma espécie de suavização local:

$$C = \sum_{i=1}^p (y_i - f(\mathbf{x}_i))^2 + \sum_{j=1}^m \lambda_j w_j^2 \quad (5.12)$$

onde λ_j são os termos de penalização locais da RNA.

Tanto o parâmetro de regressão global quanto os de regressão local são estimados de acordo com algum critério de seleção de modelo, minimizando dessa maneira os erros estimados da RNA.

As técnicas de seleção de modelos procuram estimar o erro da RNA para cada função, buscando minimizar a quantidade de funções básicas envolvidas. Neste trabalho foi utilizada a validação cruzada generalizada (*generalized cross-validation*) [Orr 1996] como critério para estimativa de erro. A variância estimada para o erro da RNA é dada por:

$$\sigma^2 = \frac{S}{p - m} \quad (5.13)$$

onde p é a quantidade de padrões utilizada no treinamento e m é a quantidade de pesos da RNA.

5.3. Classificação dos Padrões

Os primeiros experimentos realizados envolveram a utilização da primeira base de dados como entrada para uma rede perceptron multicamadas, com o intuito de confirmar os resultados estatísticos obtidos previamente. Utilizaram-se 64 neurônios de entrada (correspondentes à dimensão dos vetores de entrada), 16 neurônios escondidos e 10 neurônios de saída (correspondentes ao número de classes envolvidas).

Das cinco amostras disponíveis para cada classe de impressão digital do primeiro conjunto de dados, foram utilizadas quatro para o treinamento e apenas a amostra restante de cada classe para a validação. O objetivo principal desse primeiro experimento utilizando RNA foi a observação da convergência do algoritmo de treinamento e o comportamento final da RNA treinada relativo à classificação dos padrões.

A Figura 5.2 ilustra o comportamento do erro médio quadrático da saída da RNA durante o treinamento para fatores de aprendizado de 0,1; 0,2; 0,4 e 0,8. É possível observar a rápida convergência do algoritmo em todos os casos.

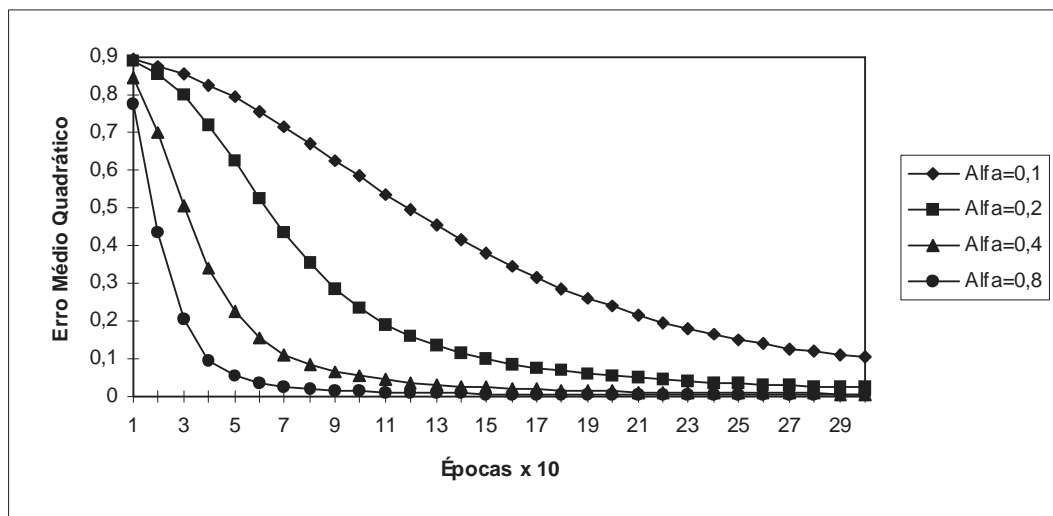


Figura 5.2 - Erro médio quadrático durante o treinamento.

O treinamento foi subdividido em estágios de 100 épocas, com a realização de testes intermediários para avaliar o erro médio quadrático resultante, utilizando-se das amostras de teste. A Tabela 5.1 apresenta os resultados intermediários obtidos ao final de cada teste intermediário.

Épocas	$\alpha = 0,1$	$\alpha = 0,2$	$\alpha = 0,4$	$\alpha = 0,8$
100	0,55763	0,25128	0,06299	0,01950
200	0,23605	0,07831	0,02305	0,00829
300	0,11036	0,04614	0,01411	0,00540
400	0,06677	0,03265	0,01037	0,00403
500	0,04690	0,02547	0,00824	0,00325

Tabela 5.1 - Erros médios quadráticos resultantes dos testes intermediários.

A técnica de treinamento empregada foi o algoritmo *backpropagation* padrão, sem nenhum fator de momento. O fator de aprendizado (α) foi variado em testes separados para avaliar a tendência de convergência para mínimos locais, que acabou não sendo observada, uma vez que em todos os casos a RNA classificou corretamente os padrões de teste e que os respectivos erros médios quadráticos resultantes foram bastante reduzidos [Vieira Neto 1997].

Dessa maneira, foi confirmada a adequada separabilidade entre as classes de vetores característicos obtidos na etapa de extração de características das imagens. No entanto, a base de dados utilizada pode ser considerada pouco significativa, face ao pequeno número de classes e amostras envolvidas. O perceptron multicamadas também

apresenta algumas desvantagens no uso para a aplicação proposta, pois sua grande capacidade de generalização acaba por classificar erroneamente padrões que não foram utilizados durante o treinamento. Em outras palavras, se um padrão de uma classe ainda desconhecida for apresentado à RNA, esta irá sempre procurar apresentar uma classificação em sua saída, mesmo que o padrão de entrada não seja exatamente correspondente à classe conhecida selecionada.

Os experimentos seguintes foram realizados a partir da segunda base de dados, contendo um número bem superior de classes e amostras de padrões, e utilizando-se da topologia de rede com funções básicas radiais. Foram separadas 90 classes para o treinamento supervisionado, sendo que 7 amostras de cada classe foram utilizadas para o treinamento e as 3 amostras restantes para a validação. Os resultados obtidos com as 90 classes envolvidas serviram para levantamento dos índices de falsa rejeição, falsa aceitação e múltipla classificação entre as classes utilizadas no treinamento.

Dessa forma, foi montada uma rede com funções básicas radiais com 64 neurônios de entrada (correspondente à dimensão dos vetores de entrada), 630 neurônios intermediários (correspondente ao número de padrões de treinamento) e 90 neurônios de saída (correspondente ao número de classes envolvidas). Os centros das funções básicas radiais foram posicionados sobre os vetores de treinamento, conforme sugerido em [Orr 1997]. O raio mínimo das funções foi determinado a partir da máxima distância entre os vetores de treinamento [Haykin 1994], conforme a equação (5.13):

$$r = \frac{d}{\sqrt{N}} \quad (5.14)$$

onde d corresponde à máxima distância euclidiana entre os vetores de treinamento e N é o número de vetores de treinamento.

As 10 classes restantes foram utilizadas para testes de falsa aceitação para classes não utilizadas durante o treinamento. Ainda nos testes de falsa aceitação, foram adicionadas as 3 amostras restantes das 90 classes ao conjunto de dados de treinamento.

Foram realizados experimentos com os raios das funções variando de r (0,2180) a $4r$ (0,8720) e também com limiares de decisão para a saída da RNA variando de 0,5 a 0,9, sendo o treinamento realizado pelo método dos mínimos quadrados.

A seguir, o algoritmo de classificação dos padrões de impressão digital é apresentado em pseudo-código:

1. Ler os 7 primeiros vetores de características das classes de impressão digital
2. Montar matriz de treinamento da rede, a partir dos 7 primeiros vetores de características
3. Gerar matriz de saídas desejadas da rede com funções básicas radiais
4. Ler os 3 vetores de características restantes das classes de impressão digital
5. Montar matriz de validação da rede, a partir dos 3 vetores de características restantes
6. Calcular o raio mínimo (r) para as funções básicas radiais, a partir da máxima distância euclidiana entre os vetores de características da matriz de treinamento
7. Para valores de raio das funções básicas radiais de r ; $2r$; $3r$; e $4r$:
 - A. Gerar a matriz de funções básicas radiais, a partir da matriz de treinamento e do raio, utilizando funções gaussianas
 - B. Calcular a matriz de pesos das conexões da rede com funções básicas radiais, a partir da matriz de treinamento da rede e da matriz de saídas desejadas
 - C. Calcular as saídas da rede, a partir da matriz de pesos e da matriz de validação da rede
 - D. Para limiares de 0,5; 0,7; e 0,9: calcular os índices de falsa rejeição, falsa aceitação e múltipla classificação, a partir da comparação entre as saídas obtidas e as saídas desejadas para a rede
8. Repetir o passo 7 utilizando a técnica de seleção progressiva, tendo como critério de estimativa de erro a validação cruzada generalizada
9. Repetir o passo 7 utilizando a técnica de regressão global, calculando o termo de penalização global a partir da matriz de funções básicas radiais
10. Repetir o passo 7 utilizando a técnica de regressão local, calculando os termos de penalização locais a partir da matriz de funções básicas radiais e do termo de penalização global
11. Repetir o algoritmo completo utilizando todos os vetores de características para treinar a rede com funções básicas radiais

As Tabelas 5.2, 5.3 e 5.4 mostram os resultados obtidos para os índices de falsa rejeição, falsa aceitação e múltipla classificação, respectivamente, entre as classes

utilizadas no treinamento. As Tabelas 5.5 e 5.6 mostram os resultados obtidos para os índices de falsa aceitação entre as classes não utilizadas no treinamento, para treinamento com 7 e 10 amostras, respectivamente.

	Limiar 0,5		Limiar 0,7		Limiar 0,9	
	Global	3 tentativas	Global	3 tentativas	Global	3 tentativas
r=0,2180	92,2%	85,6%	96,3%	93,3%	100%	100%
r=0,4360	46,3%	30,0%	69,6%	50,0%	90,4%	78,9%
r=0,6540	23,0%	7,8%	42,6%	18,9%	75,9%	54,4%
r=0,8720	17,0%	4,4%	33,7%	13,3%	64,4%	37,8%

Tabela 5.2 - Índices de falsa rejeição entre membros das classes de treinamento.

	Limiar 0,5		Limiar 0,7		Limiar 0,9	
	Global	3 tentativas	Global	3 tentativas	Global	3 tentativas
r=0,2180	0%	0%	0%	0%	0%	0%
r=0,4360	0%	0%	0%	0%	0%	0%
r=0,6540	0%	0%	0%	0%	0%	0%
r=0,8720	0%	0%	0%	0%	0%	0%

Tabela 5.3 - Índices de falsa aceitação entre membros das classes de treinamento.

	Limiar 0,5		Limiar 0,7		Limiar 0,9	
	Global	3 tentativas	Global	3 tentativas	Global	3 tentativas
r=0,2180	0%	0%	0%	0%	0%	0%
r=0,4360	0%	0%	0%	0%	0%	0%
r=0,6540	0%	0%	0%	0%	0%	0%
r=0,8720	0%	0%	0%	0%	0%	0%

Tabela 5.4 - Índices de múltipla classificação entre membros das classes de treinamento.

	Limiar 0,5	Limiar 0,7	Limiar 0,9
r=0,2180	0%	0%	0%
r=0,4360	0%	0%	0%
r=0,6540	0%	0%	0%
r=0,8720	1,0%	0%	0%

Tabela 5.5 - Índices de falsa aceitação entre membros das classes de teste (treinamento com 7 amostras).

	Limiar 0,5	Limiar 0,7	Limiar 0,9
r=0,2180	0%	0%	0%
r=0,4360	0%	0%	0%
r=0,6540	0%	0%	0%
r=0,8720	0%	0%	0%

Tabela 5.6 - Índices de falsa aceitação entre membros das classes de teste (treinamento com 10 amostras).

Na Tabela 5.2, o índice de falsa rejeição em três tentativas corresponde a um método utilizado para avaliação de sistemas comerciais de reconhecimento de impressões digitais, no qual o padrão somente é considerado como rejeitado após três tentativas frustradas de aceitação. A análise dos resultados apresentados na Tabela 5.2 comprova a pouca capacidade de generalização das redes com funções básicas radiais, pois de uma maneira geral os índices de falsa rejeição foram elevados. Observa-se que tais índices tendem a baixar com o aumento do raio das funções e com a diminuição do limiar de decisão da saída da RNA.

As Tabelas 5.3, 5.4, 5.5 e 5.6, por outro lado, apresentam resultados muito interessantes. Nas Tabelas 5.3 e 5.4 observa-se que não ocorreu falsa aceitação (classificação errônea), nem tampouco múltipla classificação entre as classes utilizadas no treinamento. Quando utilizados padrões não envolvidos no treinamento, houve apenas uma falsa aceitação para os 100 padrões, correspondente ao pior caso, ou seja, maior raio das funções utilizadas e menor limiar de decisão de saída da RNA, como observa-se na Tabela 5.5. Esse índice foi anulado com a utilização de todas as 10 amostras das 90 classes para o treinamento da RNA, como mostra a Tabela 5.6. Deve-se salientar que a utilização das 3 amostras restantes para o treinamento eleva o número de neurônios intermediários da RNA de 630 para 900.

O próximo experimento consistiu na tentativa de utilizar-se a técnica de seleção progressiva para reduzir a quantidade de funções básicas necessárias para a RNA, verificando a influência no seu desempenho geral. Houve redução de 630 para 21 neurônios intermediários para a RNA com raios iguais a r , para 563 neurônios intermediários para a RNA com raios iguais a $2r$, para 546 neurônios intermediários para a RNA com raios iguais a $3r$ e para 498 neurônios intermediários para a RNA com raios iguais a $4r$. O desempenho apresentado pelas RNA modificadas pelo processo de seleção progressiva, utilizando validação cruzada generalizada [Orr 1997] como critério de estimativa de erro, é mostrado nas Tabelas 5.7, 5.8, 5.9, 5.10 e 5.11.

Pela análise das Tabelas 5.7, 5.8, 5.9, 5.10 e 5.11, correspondentes às RNA modificadas pelo processo de seleção progressiva, observa-se que os índices de falsa aceitação e múltipla classificação foram preservados. Por outro lado, alguns dos índices de falsa rejeição sofreram um sensível aumento, fato compreensível tendo-se em vista a redução das funções básicas para a classificação dos padrões.

	Limiar 0,5		Limiar 0,7		Limiar 0,9	
	Global	3 tentativas	Global	3 tentativas	Global	3 tentativas
r=0,2180	96,3%	93,3%	98,9%	96,7%	100%	100%
r=0,4360	46,3%	30,0%	71,5%	51,1%	92,6%	82,2%
r=0,6540	23,0%	7,8%	43,3%	20,0%	77,4%	58,9%
r=0,8720	16,7%	4,4%	34,4%	13,3%	64,8%	37,8%

Tabela 5.7 - Índices de falsa rejeição entre membros das classes de treinamento - seleção progressiva.

	Limiar 0,5		Limiar 0,7		Limiar 0,9	
	Global	3 tentativas	Global	3 tentativas	Global	3 tentativas
r=0,2180	0%	0%	0%	0%	0%	0%
r=0,4360	0%	0%	0%	0%	0%	0%
r=0,6540	0%	0%	0%	0%	0%	0%
r=0,8720	0%	0%	0%	0%	0%	0%

Tabela 5.8 - Índices de falsa aceitação entre membros das classes de treinamento - seleção progressiva.

	Limiar 0,5		Limiar 0,7		Limiar 0,9	
	Global	3 tentativas	Global	3 tentativas	Global	3 tentativas
r=0,2180	0%	0%	0%	0%	0%	0%
r=0,4360	0%	0%	0%	0%	0%	0%
r=0,6540	0%	0%	0%	0%	0%	0%
r=0,8720	0%	0%	0%	0%	0%	0%

Tabela 5.9 - Índices de múltipla classificação entre membros das classes de treinamento - seleção progressiva.

	Limiar 0,5	Limiar 0,7	Limiar 0,9
r=0,2180	0%	0%	0%
r=0,4360	0%	0%	0%
r=0,6540	0%	0%	0%
r=0,8720	1,0%	0%	0%

Tabela 5.10 - Índices de falsa aceitação entre membros das classes de teste (treinamento com 7 amostras) - seleção progressiva.

	Limiar 0,5	Limiar 0,7	Limiar 0,9
r=0,2180	0%	0%	0%
r=0,4360	0%	0%	0%
r=0,6540	0%	0%	0%
r=0,8720	0%	0%	0%

Tabela 5.11 - Índices de falsa aceitação entre membros das classes de teste (treinamento com 10 amostras) - seleção progressiva.

Um terceiro experimento foi realizado, tendo por objetivo a obtenção de uma maior capacidade de generalização das RNA, visando a melhoria dos índices de falsa rejeição. Para isso, aplicou-se a técnica de regressão global, obtendo-se termos de

penalização iguais a $9,967 \cdot 10^{-4}$ para o caso de raios iguais a r ; 0,0191 para o caso de raios iguais a $2r$; 0,0270 para o caso de raios iguais a $3r$ e 0,0110 para o caso de raios iguais a $4r$. Os resultados obtidos com a técnica de regressão global são dados nas Tabelas 5.12, 5.13, 5.14, 5.15 e 5.16

Pela análise dos dados correspondentes aos resultados obtidos com a técnica de regressão global, observa-se que infelizmente a capacidade de generalização ocorreu de maneira inversa à desejada, aumentando a falsa aceitação (Tabela 5.16) em vez de reduzir a falsa rejeição (Tabela 5.12). Da mesma maneira que ocorreu no caso de seleção progressiva, houve um ligeiro aumento nos índices de falsa rejeição. Porém, os índices de falsa aceitação e múltipla classificação entre membros das classes de treinamento foram preservados.

	Limiar 0,5		Limiar 0,7		Limiar 0,9	
	Global	3 tentativas	Global	3 tentativas	Global	3 tentativas
r=0,2180	92,2%	85,6%	96,3%	93,3%	100%	100%
r=0,4360	46,3%	30,0%	70,7%	50,0%	91,1%	80,0%
r=0,6540	23,0%	7,8%	44,4%	21,1%	80,7%	63,3%
r=0,8720	17,8%	4,4%	36,2%	16,7%	71,1%	46,7%

Tabela 5.12 - Índices de falsa rejeição entre membros das classes de treinamento - regressão global.

	Limiar 0,5		Limiar 0,7		Limiar 0,9	
	Global	3 tentativas	Global	3 tentativas	Global	3 tentativas
r=0,2180	0%	0%	0%	0%	0%	0%
r=0,4360	0%	0%	0%	0%	0%	0%
r=0,6540	0%	0%	0%	0%	0%	0%
r=0,8720	0%	0%	0%	0%	0%	0%

Tabela 5.13 - Índices de falsa aceitação entre membros das classes de treinamento - regressão global.

	Limiar 0,5		Limiar 0,7		Limiar 0,9	
	Global	3 tentativas	Global	3 tentativas	Global	3 tentativas
r=0,2180	0%	0%	0%	0%	0%	0%
r=0,4360	0%	0%	0%	0%	0%	0%
r=0,6540	0%	0%	0%	0%	0%	0%
r=0,8720	0%	0%	0%	0%	0%	0%

Tabela 5.14 - Índices de múltipla classificação entre membros das classes de treinamento - regressão global.

	Limiar 0,5	Limiar 0,7	Limiar 0,9
r=0,2180	0%	0%	0%
r=0,4360	0%	0%	0%
r=0,6540	0%	0%	0%
r=0,8720	1,0%	0%	0%

Tabela 5.15 - Índices de falsa aceitação entre membros das classes de teste (treinamento com 7 amostras) - regressão global.

	Limiar 0,5	Limiar 0,7	Limiar 0,9
r=0,2180	0%	0%	0%
r=0,4360	0%	0%	0%
r=0,6540	1,0%	0%	0%
r=0,8720	1,0%	0%	0%

Tabela 5.16 - Índices de falsa aceitação entre membros das classes de teste (treinamento com 10 amostras) - regressão global.

O processo de treinamento através do método dos mínimos quadrados foi razoavelmente rápido, sendo da ordem de cinco minutos em ambiente MatLab®, rodando em uma estação de trabalho Sun SparcStation 20. Entretanto, os processos de seleção progressiva e regressão global mostraram-se bastante lentos, na ordem de dezenas de horas. O processo de regressão local, cujo tempo de execução é muito superior ao de regressão global, tornou-se inviável para aplicação em redes com funções básicas radiais das proporções em questão. Portanto, a técnica de regressão local foi empregada apenas nos experimentos seguintes, os quais foram realizados com redes de pequenas dimensões, em casos de reconhecimento individual de classes de impressões digitais.

Os experimentos finais foram realizados com redes treinadas com apenas uma classe de impressões digitais. Foram utilizadas 7 amostras da classe em questão para o treinamento e as 3 amostras restantes, juntamente com as demais amostras de todas as 99 classes restantes para a validação. Repetiu-se o experimento para todas as 100 classes de impressões digitais da segunda base de dados. A Tabela 5.17 apresenta o desempenho médio das RNA treinadas com classes individuais, utilizando-se vetores de características com 64 coeficientes. Visando observar o comportamento do sistema frente a uma redução na quantidade de coeficientes dos vetores de características, repetiu-se o experimento para casos com 32 e 16 coeficientes, cujos resultados médios são mostrados nas Tabelas 5.18 e 5.19, respectivamente.

São apresentados resultados referentes ao treinamento através do método dos mínimos quadrados (MQ), à aplicação da técnica de seleção progressiva (SP), à

aplicação da técnica de regressão global (RG) e à aplicação da técnica de regressão local (RL).

A partir dos dados fornecidos pelas Tabelas 5.17, 5.18 e 5.19, pode-se observar um aumento significativo nos índices de falsa rejeição e falsa aceitação com a diminuição da quantidade de coeficientes utilizados nos vetores de características. Além disso observa-se que toda tentativa de se diminuir os índices de falsa rejeição tem como efeito colateral um aumento dos índices de falsa aceitação. Verifica-se esse fato também quando do emprego das técnicas de regressão, as quais tendem a melhorar os índices de falsa aceitação, porém acabam por piorar os índices de falsa rejeição.

No capítulo seguinte são apresentadas conclusões a respeito dos resultados obtidos, além de sugestões para trabalhos futuros e melhoria do desempenho do sistema.

		Índice de Falsa Rejeição (%)				Índice de Falsa Aceitação (%)			
		r	2r	3r	4r	r	2r	3r	4r
Limiar 0,5	MQ	24,333	0,667	0	0	2,425	44,465	74,562	88,611
	SP	25,000	0,667	0	0	2,301	44,197	74,290	88,487
	RG	24,333	0,667	0	0	2,363	43,536	73,331	87,943
	RL	24,333	0,667	0	0	2,417	44,032	73,857	88,194
Limiar 0,7	MQ	45,667	5,667	0,333	0	0,265	19,463	51,570	72,126
	SP	48,333	5,667	0,333	0	0,267	19,337	51,242	71,811
	RG	46,000	6,333	0,333	0	0,255	18,703	49,932	70,551
	RL	45,667	5,667	0,333	0	0,263	19,098	50,746	71,180
Limiar 0,9	MQ	79,667	30,667	14,000	4,000	0,013	1,967	11,414	28,119
	SP	81,333	31,000	15,333	4,667	0,022	2,037	11,431	27,541
	RG	80,667	32,000	14,333	6,000	0,014	1,913	10,539	25,334
	RL	80,000	31,667	14,000	5,333	0,013	1,896	10,979	26,513

Tabela 5.17 - Índices de falsa rejeição e falsa aceitação globais médios para 64 coeficientes.

		Índice de Falsa Rejeição (%)				Índice de Falsa Aceitação (%)			
		r	2r	3r	4r	r	2r	3r	4r
Limiar 0,5	MQ	23,333	1,333	0	0	3,585	42,897	72,151	86,114
	SP	24,333	1,333	0	0	3,468	42,337	71,442	85,529
	RG	23,667	1,667	0	0	3,500	41,245	70,071	84,860
	RL	23,667	1,333	0	0	3,566	41,912	70,872	85,398
Limiar 0,7	MQ	40,000	6,333	0,667	0	0,627	20,459	49,830	69,805
	SP	42,000	6,667	0,667	0	0,614	19,998	49,005	68,505
	RG	41,333	6,667	1,000	0	0,586	19,428	47,116	66,957
	RL	40,667	6,333	1,000	0	0,622	19,800	48,173	67,986
Limiar 0,9	MQ	72,667	30,333	12,667	5,000	0,037	2,959	13,211	28,597
	SP	75,000	29,667	13,333	6,000	0,045	3,003	12,883	27,331
	RG	73,333	32,333	14,667	7,000	0,038	2,827	11,845	24,865
	RL	73,000	32,333	14,000	6,667	0,037	2,840	12,308	26,148

Tabela 5.18 - Índices de falsa rejeição e falsa aceitação globais médios para 32 coeficientes.

		Índice de Falsa Rejeição (%)				Índice de Falsa Aceitação (%)			
		r	2r	3r	4r	r	2r	3r	4r
Limiar 0,5	MQ	20,667	2,000	0	0	4,597	41,185	69,736	84,098
	SP	22,333	2,000	0	0	4,473	40,094	68,744	82,754
	RG	21,667	2,333	0	0	4,446	39,090	67,086	82,017
	RL	20,667	2,000	0	0	4,560	39,716	67,883	82,637
Limiar 0,7	MQ	39,667	6,333	1,333	0,333	1,085	20,421	48,239	67,620
	SP	41,667	6,667	1,000	0	1,097	19,685	46,994	65,927
	RG	39,667	7,333	2,000	0,667	1,026	18,917	44,593	63,956
	RL	39,333	6,667	1,667	0,667	1,074	19,242	45,676	64,939
Limiar 0,9	MQ	66,333	29,667	12,667	6,667	0,095	3,857	14,473	29,057
	SP	68,000	31,333	13,667	7,667	0,115	3,862	13,825	26,962
	RG	67,000	33,667	15,000	9,000	0,095	3,542	12,184	23,438
	RL	67,000	31,333	14,333	8,000	0,098	3,509	12,537	24,648

Tabela 5.19 - Índices de falsa rejeição e falsa aceitação globais médios para 16 coeficientes.

CAPÍTULO 6

CONCLUSÕES

Um sistema completo de reconhecimento de impressões digitais foi desenvolvido neste trabalho, envolvendo tanto a construção do *hardware* de aquisição das imagens, quanto o *software* associado às etapas de pré-processamento, segmentação, extração de características e classificação dos padrões.

Com respeito ao comportamento apresentado pelo dispositivo de aquisição de imagens construído, pode-se concluir que o mesmo funcionou a contento, apesar da sua grande simplicidade construtiva e funcional. As imagens adquiridas apresentaram características adequadas ao posterior processamento pelas demais etapas do sistema. Entretanto, foi necessária uma etapa de pré-processamento das imagens, com o objetivo de melhorar o foco resultante da aquisição.

A etapa de segmentação apresentou resultados muito bons, tendo em vista a diversidade das imagens coletadas, conforme pode ser observado no Apêndice A. Problemas de translação nas regiões de interesse das imagens foram plenamente contornados. Porém, o método de segmentação proposto mostrou-se bastante sensível a efeitos de rotação. Para tentar contornar esse problema, optou-se pela utilização de imagens de impressões digitais de polegares, cuja rotação no momento de aquisição é menos provável de ocorrer devido a restrições físicas impostas pelo *design* do dispositivo de aquisição.

Deve-se salientar também que o algoritmo de segmentação baseado na máxima correlação com a região de referência da primeira imagem segmentada torna-se fortemente dependente da qualidade dessa primeira imagem utilizada, fato que constitui uma desvantagem. Dessa maneira, se a primeira imagem segmentada por morfologia matemática for inadequada (com alto grau de rotação em relação às demais, por exemplo), acabará comprometendo a segmentação das imagens subseqüentes da mesma classe, mesmo que estas últimas sejam adequadas.

A etapa de extração de características a partir da transformada *wavelet* segundo a especificação do FBI parece oferecer uma representação coerente dos padrões, tendo em

vista os resultados estatísticos obtidos através de análise correlacional e os resultados preliminares com perceptrons multicamadas. Tais resultados indicam uma adequada separabilidade das classes de padrões geradas pelo método empregado. Experimentos de redução na quantidade de coeficientes dos vetores de características tiveram como consequência uma piora nos índices de falsa aceitação e falsa rejeição.

A transformada *wavelet* empregada no estágio de extração de características justifica a existência da etapa de segmentação das imagens baseada na máxima correlação das suas regiões centrais, uma vez que versões muito diferentes de imagens de uma mesma impressão digital, principalmente por efeitos de translação, resultam em vetores de características razoavelmente diferentes. É importante destacar que efeitos de rotação nas imagens também trariam consequências indesejadas para a etapa de extração de características, porém se houver um alto grau de rotação nas imagens o processo se tornará falho já na etapa de segmentação, devido à sensibilidade anteriormente comentada.

A segunda base de dados utilizada, com 100 classes de impressões digitais, apesar de possuir tamanho significativo ainda pode ser considerada muito pequena em relação ao número de impressões digitais existentes, da ordem de bilhões. No entanto, foi possível comprovar a separabilidade e coerência das classes geradas pelo estágio de extração de características a partir das imagens utilizadas.

A principal característica das redes com funções básicas radiais, que motiva especialmente o seu uso no sistema proposto, é a sua baixa capacidade de generalização. Tal fato contribui para a obtenção de baixos índices de falsa aceitação, mas por outro lado a pouca capacidade de generalização a partir dos exemplos apresentados durante o treinamento dificultam a obtenção de baixos índices de falsa rejeição para o sistema. Por outro lado, redes do tipo perceptron multicamadas generalizam demasiadamente, possibilitando a classificação errônea de padrões desconhecidos, ou seja, que não fizeram parte da base de dados utilizada para o treinamento.

Tentativas de melhoramento dos índices de falsa rejeição, tanto pela diminuição do limiar de decisão da saída da RNA, quanto pelo aumento do raio das funções básicas radiais, sempre resultaram em deterioração dos índices de falsa aceitação. As tentativas de emprego de técnicas de seleção progressiva e regressão global e local no estágio de classificação acabaram por não oferecer vantagens significativas no desempenho do sistema quanto aos índices de falsa rejeição. Entretanto, como a meta de funcionamento

do sistema é a identificação de indivíduos e não da classe a que suas impressões digitais pertencem, pode-se considerar os resultados deste trabalho como positivos tendo em vista os baixos índices de classificação errônea apresentados (falsa aceitação e múltipla classificação). Porém, para aplicações práticas, seria adequada a redução dos índices de falsa rejeição, como proposto a seguir nas sugestões para trabalhos futuros.

Como o trabalho aqui exposto constitui o início do desenvolvimento de um sistema mais robusto, consideramos a utilização de *wavelets* e redes neuronais artificiais promissora em relação ao problema de reconhecimento de impressões digitais. Ainda há muito a ser feito para a obtenção de resultados mais equilibrados entre a precisão e a generalização oferecidas pelas redes com funções básicas radiais.

A Tabela 6.1 apresenta os tempos de execução aproximados das rotinas envolvidas no sistema completo. Pode-se reparar que a etapa do sistema que mais consome tempo na execução é a segmentação. Estima-se uma redução de 2/3 no tempo de execução para uma implementação em linguagem C compilada, em relação à linguagem interpretada própria do MatLab®. É conveniente lembrar também que os estágios de pré-processamento, segmentação e extração de características podem ser implementados em processadores digitais de sinais (DSP), otimizando ainda mais o tempo de execução dos algoritmos em questão.

Pré-processamento	2,5 s
Segmentação por Morfologia Matemática	60 s
Segmentação por Máxima Correlação	150 s
Extração de Características	2,0 s
Treinamento por Mínimos Quadrados	3,0 s
Seleção Progressiva	0,1 s
Regressão Global	0,4 s
Regressão Local	10 s

Tabela 6.1 - Tempo aproximado de processamento das diversas etapas do sistema (estação de trabalho Sun SparcStation 20).

6.1 Sugestões para Trabalhos Futuros:

Apesar dos índices de falsa aceitação das RNA testadas neste trabalho terem se mostrado adequados, os resultados obtidos para os índices de falsa rejeição foram considerados elevados para aplicações práticas. Em trabalhos futuros sugere-se a implementação e teste de algoritmos de treinamento não-lineares para as redes com

funções básicas radiais. Algumas das técnicas que podem ser experimentadas são a seleção auto-organizada de centros [Moody 1989] e a seleção supervisionada de centros [Haykin 1994]. Ainda com respeito à topologia de RNA a ser utilizada, sugere-se também a verificação da adequação do emprego de funções básicas não-radiais orientadas [Saha 1991] e também de funções PPS-*wavelet* e PPS-radial [Marar 1997].

Quanto ao restante dos estágios envolvidos no sistema, sugere-se principalmente a pesquisa de um algoritmo de segmentação mais veloz e que preferencialmente fosse insensível a rotação. Técnicas automáticas de limiarização também podem ser testadas, verificando o seu efeito no desempenho final do sistema. Outra sugestão é a do estudo de outras alternativas de extração de características através do método WSQ, buscando representações melhores e mais compactas para os padrões.

A base de dados a ser utilizada em trabalhos futuros certamente deverá ser ampliada e para isso, também seria interessante a utilização de dispositivos comerciais de aquisição de impressões digitais, de modo a averiguar a adequação do método proposto para imagens de diferente qualidade e resolução espacial.

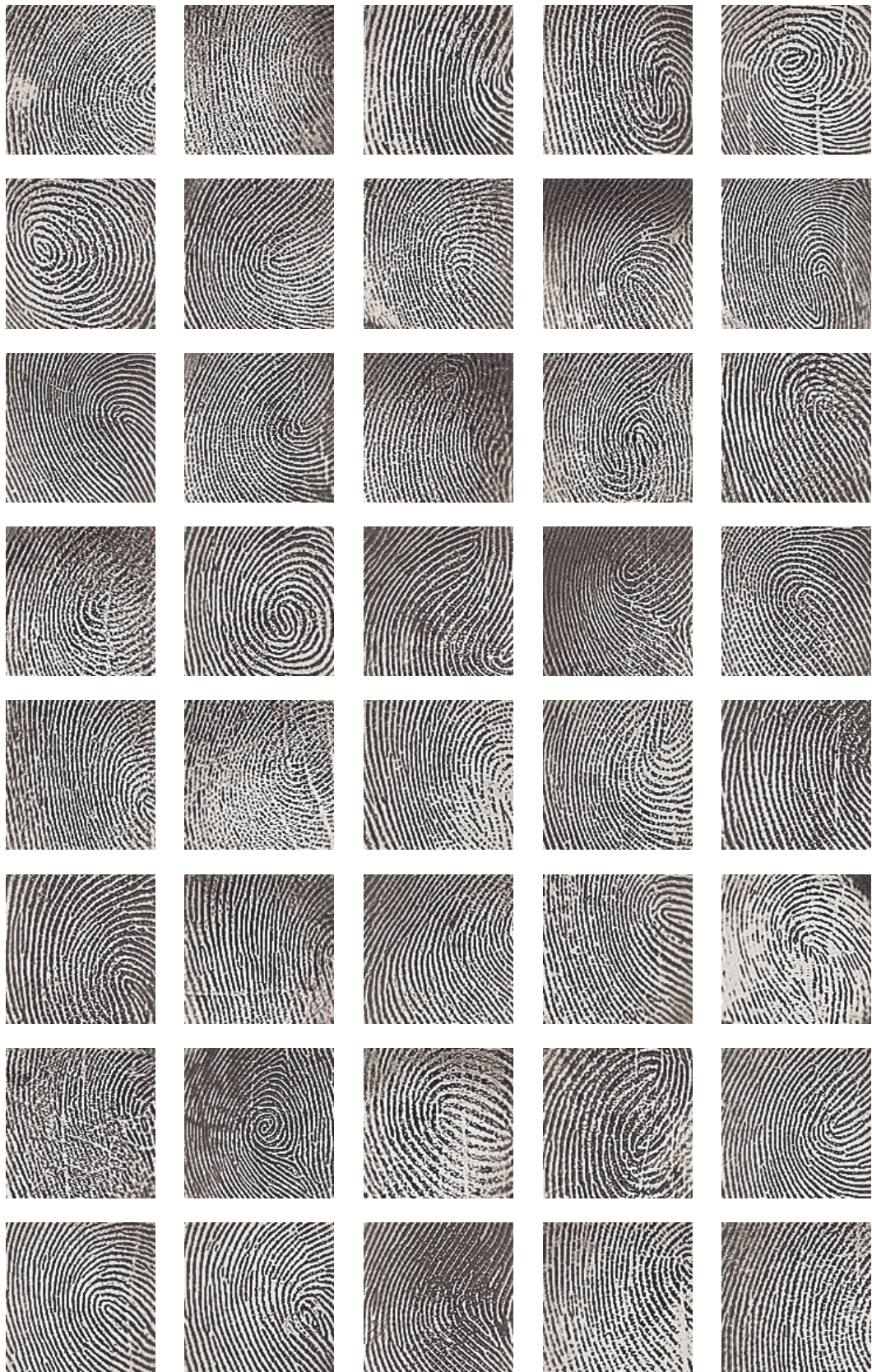
APÊNDICE A

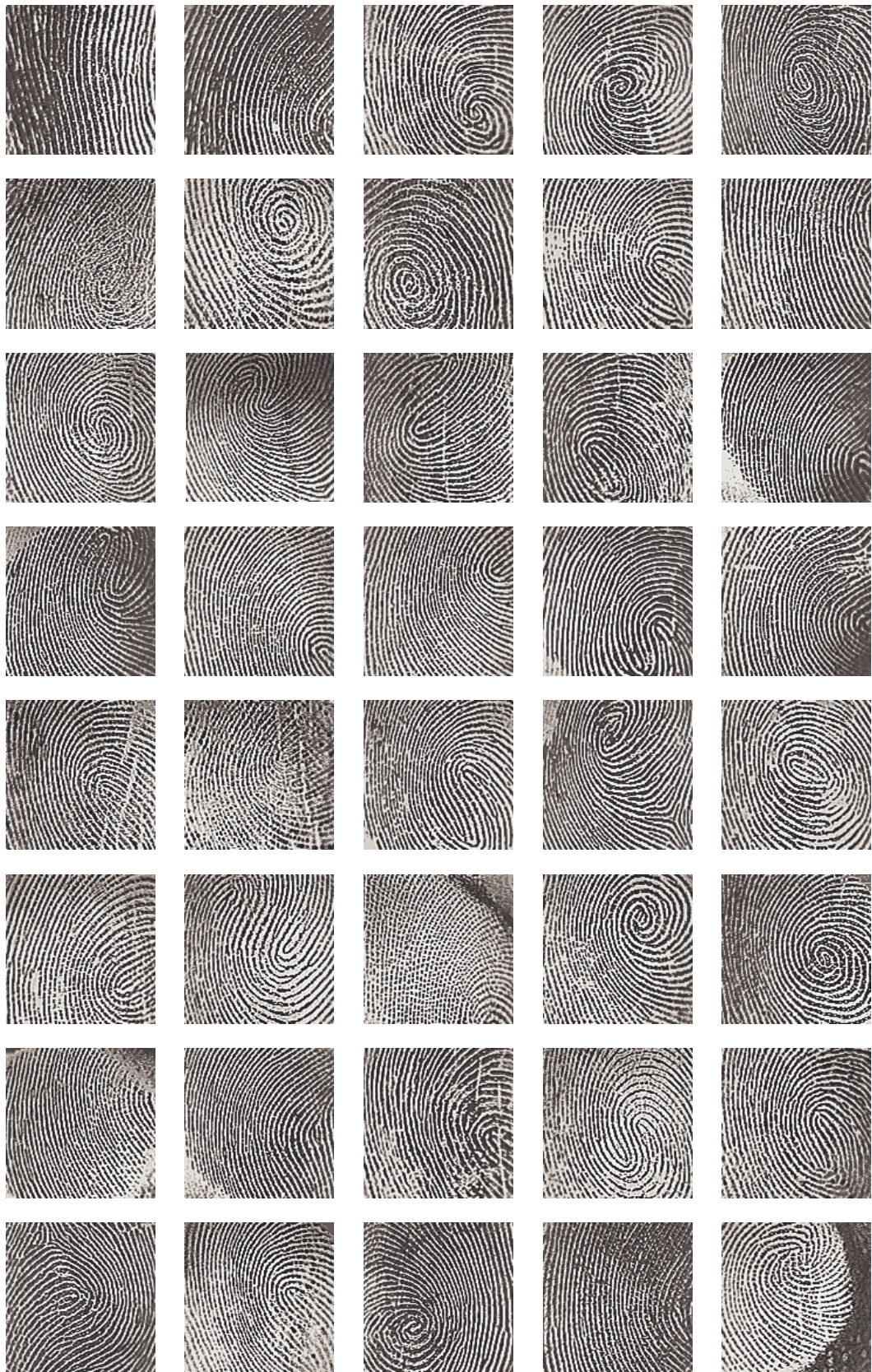
Primeira Base de Dados



Segunda Base de Dados







REFERÊNCIAS BIBLIOGRÁFICAS

- [Baldi 1993] BALDI, P., CHAUVIN, Y., Neural Networks for Fingerprint Recognition, **Neural Computation**, n. 5, pp. 402-418, 1993.
- [Bishop 1991] BISHOP, C., Improving the Generalization Properties of Radial Basis Function Neural Networks, **Neural Computation**, n. 3, pp. 579-588, 1991.
- [Blue 1994] BLUE, J. L. et al., Evaluation of Pattern Classifiers for Fingerprint and OCR Applications, **Pattern Recognition**, v. 27, n. 4, pp. 485-501, 1994.
- [Borges 1994] BORGES, D. L., ORR, M. J., FISHER, R. B., A Radial Basis Function Neural Network for Parts Identification of Three Dimensional Shapes, **VII SIBGRAPI**, 1994, pp. 77-84.
- [Brislawn 1996] BRISLAWN, C. M. et al., The FBI Compression Standard for Digitized Fingerprint Images, **SPIE**, v. 2847, August 1996.
- [Buckheit 1995] BUCKHEIT, J. et al., **WaveLab Reference Manual - version 0.7**, Stanford University, December 1995.
- [Facon 1996] FACON, J., **Morfologia Matemática: Teoria e Exemplos**, Editora Universitária Champagnat - PUC-PR, 1996.
- [FBI 1993] FEDERAL BUREAU OF INVESTIGATION, **Document IAFIS-IC-0110v2 - WSQ Gray-scale Fingerprint Image Compression Specification**, February 1993.

- [Fournier 1995] FOURNIER, A., Wavelets and their Applications in Computer Graphics, **SIGGRAPH'95 Course Notes**, 1995.
- [Gonzalez 1992] GONZALEZ, Rafael C., WOODS, Richard E., **Digital Image Processing - Third Edition**, Addison-Wesley, 1992.
- [Gopinath 1993] GOPINATH, R. A., BURRUS, C. S., Tutorial Overview of Filter Banks, Wavelets and Interrelations, **ISCAS'93**, 1993.
- [Graps 1995] GRAPS, A., An Introduction to Wavelets, **IEEE Computational Science & Engineering**, Summer 1995.
- [Haykin 1994] HAYKIN, Simon, **Neural Networks: A Comprehensive Foundation**, IEEE Press, 1994.
- [Hung 1993] HUNG, D. C. D., Enhancement and Feature Purification of Fingerprint Images, **Pattern Recognition**, v. 26, n. 11, pp. 1661-1671, 1993.
- [Karu 1996] KARU, K., JAIN, A. K., Fingerprint Classification, **Pattern Recognition**, v. 29, n. 3, pp. 389-404, 1996.
- [Lippmann 1987] LIPPMANN, R. P., An Introduction to Computing with Neural Nets, **IEEE Acoustics, Speech and Signal Processing Magazine**, pp. 4-22, April 1987.
- [Le Cun 1990] LE CUN, Y. et al., Handwritten Digit Recognition with a Back-propagation Network, **Neural Information Processing Systems**, v. 2, pp. 396-404, 1990.

- [Masters 1994] MASTERS, Timothy, **Signal and Image Processing with Neural Networks: A C++ Sourcebook**, John Wiley & Sons, 1994.
- [Mallat 1989] MALLAT, S. G., A Theory for Multiresolution Signal Decomposition: The Wavelet Representation, **IEEE Transactions on Pattern Analysis and Machine Intelligence**, vol. 11, n. 7, pp. 674-693, July 1989.
- [Marar 1997] MARAR, J. F., **Polinômios Potência de Sigmóides (PPS): Uma Nova Técnica para Aproximação de Funções, Construção de Wavenets e Suas Aplicações em Processamento de Imagens e Sinais**, Recife, 1997. Tese de Doutorado - Informática, Universidade Federal de Pernambuco.
- [Moody 1989] MOODY, J., DARKEN, C., Fast Learning in Networks of Locally-Tuned Processing Units, **Neural Computation**, vol. 1, pp. 281-294, 1989.
- [NIST 1992] NATIONAL INSTITUTE OF STANDARDS AND TECHNOLOGY, **Technical Report NISTIR 4880 - Massively Parallel Neural Network Fingerprint Classification System**, July 1992.
- [NIST 1993] NATIONAL INSTITUTE OF STANDARDS AND TECHNOLOGY, **Technical Report NISTIR 5163 - Comparative Performance of Classification Methods for Fingerprints**, April 1993.
- [Orr 1996] ORR, M. J. L., **Introduction to Radial Basis Functions Networks**, Centre for Cognitive Science - University of Edinburgh, April 1996.

- [Orr 1997] ORR, M. J. L., **MatLab Routines for Subset Selection and Ridge Regression in Linear Neural Networks**, Centre for Cognitive Science - University of Edinburgh, February 1997.
- [Ratha 1995] RATHA, N. K., CHEN, S., JAIN, A. K., Adaptive Flow Orientation-Based Feature Extraction in Fingerprint Images, **Pattern Recognition**, v. 28, n. 11, pp. 1657-1672, 1995.
- [Rumelhart 1994] RUMELHART, D. E., WIDROW, B., LEHR, M. A., The Basic Ideas in Neural Networks, **Communications of the ACM**, v. 37, n. 3, pp. 87-92, 1994.
- [Saha 1991] SAHA, A., CHRISTIAN, J., TANG, C. S., Oriented Non-Radial Basis Functions for Image Coding and Analysis, In: Lippmann, R., Moody, D., Touretzky, D. (eds.), **Advances in Neural Processing Systems 3**, Morgan Kaufmann, 1991, pp. 728-734.
- [Schalkoff 1989] SCHALKOFF, R.J., **Digital Image Processing and Computer Vision**, John Wiley & Sons, 1989.
- [Setlak 1996] SETLAK, D. et al., A Direct Fingerprint Reader, **CardTech / SecureTech Show**, 1996.
- [Sherlock 1996a] SHERLOCK, B. G., MONRO, D. M., Optimized Wavelets for Fingerprint Compression, **ICASSP'96**, v. 3, pp. 3445-3462, 1996.
- [Sherlock 1996b] SHERLOCK, B. G., MONRO, D. M., Psychovisually Tuned Wavelet Fingerprint Compression, **ICIP'96**, 1996.
- [Stearns 1990] STEARNS, Samuel D., HUSH, Don R., **Digital Signal Analysis**, Prentice-Hall International Editions, 1990.

- [Vaidyanathan 1995] VAIDYANATHAN, P. P., DJOKOVIC, I., Wavelet Transforms. In: CHEN, K. (ed.). **The Circuits and Filters Handbook**, IEEE Press, 1995, pp. 134-219.
- [Vidakovic 1994] VIDA KOVIC, B., MÜLLER, P., **Wavelets for Kids**, Duke University, 1994, unpublished. Disponível na Internet via FTP (ftp.isds.duke.edu/pub/Users/brani/papers/wav4kidsA.ps.Z)
- [Vieira Neto 1997] VIEIRA NETO, H., BORGES, D. L., Fingerprint Classification with Neural Networks, In: Borges, D. L., Martins, W. (eds.), **IVth Brazilian Symposium on Neural Networks**, IEEE Computer Society, 1997, pp. 66-72.
- [Villasenor 1995] VILLASENOR, J. D., BELZER, B., LIAO, J., Wavelet Filter Evaluation for Image Compression, **IEEE Transactions on Image Processing**, v. 4, n. 8, 1995.