

# Reconhecimento de Caracteres Numéricos Manuscritos Usando a Transformada Wavelet

Suzete Élide Nóbrega Correia

Dissertação de Mestrado submetida à Coordenação dos Cursos de Pós-Graduação em Engenharia Elétrica da Universidade Federal da Paraíba - Campus II como parte dos requisitos necessários para obtenção do grau de Mestre no domínio da Engenharia Elétrica .

Área de Concentração: Processamento da Informação

João Marques de Carvalho, Ph.D.

Orientador

Campina Grande, Paraíba, Brasil

©Suzete Élide Nóbrega Correia, Julho de 2000



C824r      Correia, Suzete Elida Nobrega  
Reconhecimento de caracteres numericos manuscritos  
usando a transformada wavelet / Suzete Elida Nobrega  
Correia. - Campina Grande, 2000.  
70 f.

Dissertaca (Mestrado em Engenharia Eletrica) -  
Universidade Federal da Paraiba, Centro de Ciencias e  
Tecnologia.

1. Reconhecimento de Manuscritos 2. Transformada Wavelet  
3. Redes Neurais 4. Dissertacao I. Carvalho, Joao Marques  
de II. Universidade Federal da Paraiba - Campina Grande  
(PB)

CDU 681.3(043)


**RECONHECIMENTO DE CARACTERES NUMÉRICOS MANUSCRITOS  
USANDO A TRANSFORMADA WAVELET**

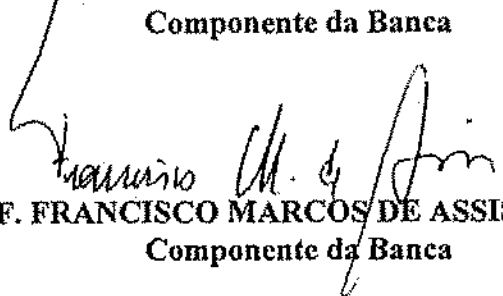
**SUZETE ÉLIDA NÓBREGA CORREIA**

Dissertação Aprovada em 04.07.2000

  
**PROF. JOÃO MARQUES DE CARVALHO, Ph.D., UFPB**  
Orientador

  
**PROF. BENEDITO GUIMARÃES AGUIAR NETO, Dr.-Ing., UFPB**  
Componente da Banca

  
**PROF. BENEMAR ALENCAR DE SOUZA, D.Sc., UFPB**  
Componente da Banca

  
**PROF. FRANCISCO MARCOS DE ASSIS, Dr., UFPB**  
Componente da Banca

CAMPINA GRANDE - PB  
Julho - 2000

## Dedicatória

Às pessoas mais importantes da minha vida: meus pais, Ari e Salete, meus irmãos, José Ary e Suzanne e meu noivo Michel.

## Agradecimentos

A Deus, por ter me dado o dom da vida e a capacidade de realizar esta dissertação.

Ao professor João Marques de Carvalho, por sua orientação, incentivo, colaboração e paciência durante todo o trabalho.

À amiga Luciana Veloso, pela grande contribuição através dos programas e artigos fornecidos.

Ao professor Francisco Marcos por ter propiciado o meu interesse por wavelets e à Madeiro, Marcos Moraes e Núbia, pelos estudos realizados.

Ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), pelo apoio financeiro.

À COPELE e seus funcionários, Ângela, Pedro e Eleonora, pela disponibilidade constante.

Por último, a todos os colegas de laboratório, Joseana, Vânia, Josemar, Robson, Rinaldo, Paulo Márcio, Vivian, Luiz Gonzaga, Leonardo e, em especial, aos amigos Izabel e Wamberto, dos quais sempre recebi demonstração de carinho, simpatia e incentivo.

## Resumo

O reconhecimento automático de caracteres numéricos manuscritos tem sido um tópico de pesquisa extensivamente estudado nas últimas duas décadas, devido ao seu valor teórico no reconhecimento de padrões e à enorme quantidade de aplicações, tais como, processar códigos postais em envelopes e numerais em cheques bancários. Embora, muita pesquisa já tenha sido desenvolvida, este continua sendo um problema de difícil solução devido às variações existentes na escrita manual. A transformada wavelet é uma nova ferramenta matemática desenvolvida recentemente, cujas propriedades de localização nos domínios do tempo e frequência, permitem uma representação das imagens dos numerais através de características globais e direcionais, simultaneamente. Nesta dissertação, um novo sistema de reconhecimento é proposto, formado por três etapas: pré-processamento, extração de características e classificação. O pré-processamento lida com a normalização em escala. A extração de características visa representar os numerais através dos coeficientes wavelets. Por fim, a classificação é feita usando uma rede neural multicamadas com agrupamento treinada com o algoritmo de retropropagação do erro. Experimentos realizados com as famílias ortogonais Haar, Daubechies, Coiflets e biortogonais Cohen-Daubechies-Feauveau utilizaram os caracteres da base de dados numérica do CENPARMI. Os resultados obtidos atestam o bom desempenho do método proposto.

## Abstract

The automatic recognition of handwritten numerical characters has been a research topic extensively studied in the last two decades because both of its theoretical value in pattern recognition and its potential for applications, such as in automatically processing postal ZIP codes from mail pieces and money amount from bankchecks. However, it still remains a difficult problem due to the large degree of variability the handwritten data may exhibit. The wavelet transform is a new tool developed in recent years, with the localization properties in both time and frequency domains, which simultaneously provides global and directional features about the numerals images. In this work, a novel approach for recognition of handwritten numerals is proposed, consisting of three stages: pre-processing, feature extraction and classification. Preprocessing deals with image normalization. Feature extraction aims to represent the normalized images by wavelets coefficients. Classification performs the final decision using a multilayer cluster neural network trained with the backpropagation algorithm. Experiments were realized with the wavelet families Haar, Daubechies, Coiflets and Cohen-Daubechies-Feauveau using the characters of the numerical database of CENPARMI. Results obtained show that the proposed method yields good performance.

# Lista de Símbolos e Abreviaturas

- DWT<sub>f</sub> - Transformada Wavelet Discreta (*Discrete Wavelet Transform*)  
CDF - Família wavelet Cohen-Daubechies-Feauveau  
RNMCA - Rede Neural Multicamadas com Agrupamento  
CENPARMI - Centro de Reconhecimento de Padrões e Máquinas Inteligentes (*Center for Pattern Recognition and Machine Intelligence*)  
PPI - *Pixels por Polegada* - (*Pixels per Inch*)  
LVQ - Algoritmo de Quantização Vetorial - (*Learning Vector Quantization*)  
MLP - *Perceptron* Multicamadas - (*Multilayer Perceptron*)  
DEE - Departamento de Engenharia Elétrica  
UFPb - Universidade Federal da Paraíba  
 $f(t)$  - Função ou sinal  
 $L^2(\mathbb{R})$  - Espaço das funções de quadrado integrável  
 $\psi(t)$  - Função wavelet mãe  
 $a$  - Fator de escalonamento  
 $b$  - Fator de translação  
 $\psi_{a,b}(t)$  - Wavelets escalonadas e transladadas pelos valores inteiros  $a$  e  $b$   
 $\psi_{j,k}(t)$  - Wavelet na escala  $j$  e translação  $k$   
 $\phi(t)$  - Função escalonamento modelo  
 $\phi_{j,k}(t)$  - Função escalonamento na escala  $j$  e translação  $k$   
 $j$  - Escala ou nível de resolução  
 $V_j$  - Subespaço de aproximação na resolução  $j$   
 $W_j$  - Complemento ortogonal de  $V_j$  em  $V_{j-1}$   
 $h(n)$  - Coeficientes do filtro de análise passa-baixas



$g(n)$  - Coeficientes do filtro de análise passa-altas  
 $c_j$  - Coeficientes de aproximação na resolução  $j$   
 $d_j$  - Coeficientes de detalhes na resolução  $j$   
 $N$  - Dimensão do sinal  
 $O(N)$  - Complexidade de ordem  $N$   
 $\tilde{\psi}(t)$  - Função wavelet dual  
 $\tilde{\phi}(t)$  - Função escalonamento dual  
 $\tilde{V}_j$  - Subespaço de aproximação dual na resolução  $j$   
 $\tilde{W}_j$  - Complemento ortogonal de  $\tilde{V}_j$  em  $V_{j-1}$   
 $\tilde{h}(n)$  - Coeficientes do filtro de síntese passa-baixas  
 $\tilde{g}(n)$  - Coeficientes do filtro de síntese passa-altas  
 $\phi(x, y)$  - Função escalonamento bi-dimensional  
 $\psi^1(x, y), \psi^2(x, y), \psi^3(x, y)$  - Wavelets bi-dimensionais  
 $A_j$  - Subimagem de coeficientes de aproximação na resolução  $j$   
 $D_j^1$  - Subimagem de coeficientes com detalhes horizontais na resolução  $j$   
 $D_j^2$  - Subimagem de coeficientes com detalhes verticais na resolução  $j$   
 $D_j^3$  - Subimagem de coeficientes com detalhes diagonais na resolução  $j$   
 $f(x, y)$  - Imagem digital  
 $f_s(x, y)$  - Imagem normalizada  
 $S_x$  - Fator de escala na direção  $x$   
 $S_y$  - Fator de escala na direção  $y$   
 $X, Y$  - Dimensões originais da imagem do caracter  
 $R_x, R_y$  - Dimensões normalizadas da imagem do caracter  
 $I_{j-1}$  - Imagem na resolução  $j - 1$   
 $H_k$  - Máscaras de Haar  
 $net_{pi}$  - Valor de ativação do padrão  $p$  no neurônio  $i$   
 $w_{ki}$  - Peso sináptico associado à conexão entre o neurônio  $k$  de uma camada e o neurônio  $i$  da camada adjacente  
 $i_{pk}$  - Valor de entrada no neurônio  $k$  do padrão  $p$   
 $O_{pj}$  - Valor de saída no neurônio  $j$  do padrão  $p$   
 $t_{pj}$  - Saída desejada no neurônio  $j$  do padrão  $p$

- $\varphi(.)$  - Função de ativação sigmóide
- $E_p$  - Função erro médio quadrático do padrão  $p$
- $\delta_{pj}$  - Gradiente local do padrão  $p$  no neurônio  $j$
- $\eta$  - Fator de aprendizagem
- $\alpha$  - Momento
- $K$  - Parâmetro de inclinação da função sigmóide
- $Z$  - Conjunto dos inteiros
- $\langle \cdot \rangle$  - Produto interno
- $\oplus$  - Soma direta
- $\otimes$  - Operador de convolução

# Lista de Figuras

1.1	Exemplo de variações existentes no numeral três com caracteres de diversos tamanhos, formatos e traçados. . . . .	2
2.1	Arquitetura do sistema de reconhecimento de manuscritos. . . . .	5
3.1	Ondas e wavelets. . . . .	18
3.2	A wavelet e suas versões escalonadas: (a) $a < 1$ , (b) $a = 1$ , (c) $a > 1$ . . . . .	20
3.3	Decomposição wavelet representada pelos subespaços de aproximação $\{V_j\}_{j \in \mathbb{Z}}$ e detalhes $\{W_j\}_{j \in \mathbb{Z}}$ . . . . .	23
3.4	Algoritmos de decomposição e reconstrução wavelet uni-dimensionais. . . . .	25
3.5	Algoritmo de decomposição wavelet bi-dimensional para um nível de resolução. A imagem $A_{j-1}$ é decoposta nas subimagens $A_j$ (características gerais), $D_j^1$ (detalhes horizontais), $D_j^2$ (detalhes verticais) e $D_j^3$ (detalhes diagonais). . . . .	28
3.6	Decomposição wavelet da imagem de um quadrado em dois níveis de resolução usando as wavelets de Haar: (a) Imagem original de dimensão $128 \times 128$ pixels, (b) Imagem com os coeficientes wavelets de dimensão $128 \times 128$ pixels. Os pixels pretos e brancos correspondem, respectivamente, a valores nulos e não-nulos. . . . .	29
3.7	Disposição das subimagens de coeficientes wavelets $A_j$ , $D_j^1$ , $D_j^2$ e $D_j^3$ para dois níveis de resolução ( $j = 1, 2$ ). . . . .	30

3.8	Algoritmo de reconstrução wavelet bi-dimensional para um nível de resolução. A imagem $A_{(j-1)}$ é reconstruída a partir das subimagens $A_j$ (características gerais), $D_j^1$ (detalhes horizontais), $D_j^2$ (detalhes verticais) e $D_j^3$ (detalhes diagonais). . . . .	31
3.9	Decomposição de Fourier da imagem de um quadrado: (a) Imagem original de dimensão $128 \times 128$ pixels, (b) Imagem com os coeficientes de Fourier de dimensão $128 \times 128$ pixels. . . . .	32
4.1	Exemplo de uma imagem normalizada do numeral 8: (a) Caracter original com $29 \times 18$ pixels, (b) Caracter normalizado com $16 \times 16$ pixels. . . . .	36
4.2	Máscaras de Haar: (a) características gerais, (b) características verticais, (c) características horizontais, (d) características diagonais. . . . .	38
4.3	Extração de características usando as wavelets biortogonais Cohen-Daubechies-Feauveau 3/7 em três níveis de resolução. . . . .	39
4.4	Arquitetura de uma rede neural com três camadas. . . . .	40
4.5	Arquitetura da rede neural multicamadas com agrupamento (RNMCA) para um nível de resolução. . . . .	45
5.1	Amostras de numerais representativos da base de dados do CENPARMI. . . . .	48
5.2	Amostras de numerais confusos da base de dados do CENPARMI. . . . .	48
5.3	Influência do comprimento do filtro no sistema de reconhecimento de numerais manuscritos usando as wavelets de Daubechies. . . . .	58

# Lista de Tabelas

3.1	Classificação das famílias wavelets. . . . .	33
5.1	Distribuição dos numerais por classe na base de dados do CENPARMI. . . . .	47
5.2	Configurações testadas. . . . .	51
5.3	Resultados obtidos para a configuração 1. . . . .	52
5.4	Resultados obtidos para a configuração 2. . . . .	53
5.5	Resultados obtidos para a configuração 3. . . . .	54
5.6	Resultados obtidos para a configuração 4. . . . .	55
5.7	Comparação entre os resultados obtidos por Lee et al. e o sistema implementado. . . . .	57
5.8	Taxas médias obtidas com as famílias wavelets Daubechies (Daub), Coiflets (Coif) e Cohen-Daubechies-Feauveau (CDF) testadas. . . . .	57
5.9	Matriz de confusão da família CDF 2/2. . . . .	59
5.10	Matriz de confusão da família CDF 2/4. . . . .	59
5.11	Matriz de confusão da família CDF 3/3. . . . .	60
5.12	Matriz de confusão da família CDF 3/7. . . . .	60
5.13	Tempo de processamento para cada etapa do sistema usando as wavelets CDF 3/7. . . . .	61
5.14	Resultados obtidos por diversos sistemas que utilizaram a base de dados do CENPARMI. . . . .	61
5.15	Resultados obtidos sem a etapa de extração de características. . . . .	62

# Índice

<b>1</b>	<b>Introdução</b>	<b>1</b>
1.1	Motivação . . . . .	1
1.2	Objetivos da dissertação . . . . .	2
1.3	Organização do texto . . . . .	2
<b>2</b>	<b>Reconhecimento de Caracteres Numéricos Manuscritos</b>	<b>4</b>
2.1	Modelo do sistema de reconhecimento . . . . .	4
2.1.1	Aquisição da imagem . . . . .	5
2.1.2	Pré-processamento . . . . .	6
2.1.3	Extração de características . . . . .	6
2.1.4	Classificação . . . . .	6
2.2	Técnicas de extração de características . . . . .	7
2.2.1	Características globais . . . . .	8
2.2.2	Distribuição dos pontos . . . . .	9
2.2.3	Características topológicas ou geométricas . . . . .	10
2.2.4	Transformada wavelet como extrator de características . . . . .	11
2.3	Tipos de classificadores . . . . .	12
2.4	Sistemas de reconhecimento: revisão bibliográfica . . . . .	13
2.5	Conclusão . . . . .	16
<b>3</b>	<b>Transformada Wavelet</b>	<b>17</b>
3.1	Propriedades da transformada wavelet . . . . .	18
3.2	Conceitos básicos . . . . .	19

3.3	Wavelets e análise multiresolucional . . . . .	21
3.3.1	Análise multiresolucional . . . . .	21
3.3.2	Algoritmos de decomposição e reconstrução . . . . .	23
3.3.3	Transformada Wavelet Aplicada a Imagens . . . . .	27
3.3.4	Wavelet <i>versus</i> Fourier . . . . .	29
3.4	Famílias wavelets . . . . .	31
3.5	Conclusão . . . . .	32
<b>4</b>	<b>Sistema Implementado</b>	<b>34</b>
4.1	Normalização em escala . . . . .	34
4.2	Extração de características . . . . .	36
4.3	Redes neurais multicamadas com retropropagação do erro . . . . .	40
4.3.1	Rede neural multicamadas com agrupamento (RNMCA) . . . . .	44
4.4	Conclusão . . . . .	45
<b>5</b>	<b>Resultados Experimentais</b>	<b>46</b>
5.1	Base de dados . . . . .	46
5.2	Parâmetros para avaliação de desempenho . . . . .	47
5.2.1	Regra de Rejeição . . . . .	49
5.3	Testes Efetuados e Análise de Desempenho . . . . .	50
5.3.1	Influência do número de níveis de resolução . . . . .	50
5.3.2	Escolha da melhor wavelet . . . . .	53
5.3.3	Análise Comparativa . . . . .	56
5.4	Conclusão . . . . .	56
<b>6</b>	<b>Conclusão</b>	<b>63</b>
6.1	Contribuições . . . . .	64
6.2	Perspectivas de Trabalhos Futuros . . . . .	64

# Capítulo 1

## Introdução

### 1.1 Motivação

O reconhecimento de caracteres manuscritos é um exemplo típico do reconhecimento automático de padrões cuja finalidade é desenvolver nos computadores a habilidade de interpretar o significado da escrita manual. Esse é um tema de grande interesse na comunidade científica por mais de duas décadas, devido a enorme quantidade de aplicações encontradas, dentre as quais, podem ser citadas, a automatização do processo de leitura de códigos postais, de cheques bancários e de formulários [1].

Apesar de numerosos métodos já terem sido propostos, o reconhecimento automático de manuscritos continua sendo um problema complexo e não totalmente resolvido, uma vez que existem inúmeras possibilidades para se representar um mesmo caracter na escrita manual [2]. Normalmente, os caracteres manuscritos podem apresentar diversas formas e tamanhos quando escritos por pessoas diferentes ou pela mesma pessoa, bem como, podem possuir aparência ambígua ou estarem ilegíveis. Muitos são os fatores que contribuem para esta diversidade de estilos, dentre eles incluem-se o tipo de papel e instrumento de escrita utilizados, habilidade, origem, nível de educação, profissão, idade e estado emocional do escritor [3, 2].

No intuito de dominar as distorções e variações de formas apresentadas pela escrita à mão, tenta-se aperfeiçoar os métodos existentes e desenvolver novas técnicas que possibilitem a implementação de sistemas eficientes e robustos. Nesse contexto, a



transformada wavelet surge como uma ferramenta promissora devido à sua propriedade multiresolucional que permite analisar tanto o formato geral do caracter quanto as suas particularidades, capazes de distingui-lo de outros caracteres semelhantes [4].



Figura 1.1: Exemplo de variações existentes no numeral três com caracteres de diversos tamanhos, formatos e traçados.

O uso da transformada wavelet em sistemas de reconhecimento de caracteres é recente. Apenas um pequeno número de trabalhos foi até agora publicado sobre este tema, [4, 5], dos quais, o proposto por Lee et al. [5] é tomado como base nesta dissertação.

## 1.2 Objetivos da dissertação

O objetivo principal deste trabalho é apresentar o sistema de reconhecimento desenvolvido no decorrer das atividades de pesquisa, que utiliza a análise multiresolucional da transformada wavelet para extrair características representativas da imagem dos caracteres, e conseqüentemente obter bons índices de reconhecimento. As técnicas empregadas em cada etapa do sistema, que vão desde a aquisição até o reconhecimento propriamente dito, também serão descritas. Por fim, os resultados obtidos são apresentados e avaliados segundo alguns parâmetros de desempenho, como reconhecimento, erro, rejeição e confiabilidade do sistema.

## 1.3 Organização do texto

A organização do texto desta dissertação é feita como se segue:

O capítulo 2 contém uma descrição de cada etapa que compõem um sistema de reconhecimento de caracteres numéricos manuscritos. As principais técnicas de extração de características e os tipos de classificadores são apresentados. Logo a seguir, são revistos alguns dos métodos de reconhecimento existentes na literatura.

O capítulo 3 é dedicado a apresentação da transformada wavelet, a qual é empregada no processo de extração de características dessa dissertação. Inicialmente, são apresentados os conceitos básicos sobre as wavelets e sobre a análise multiresolucional. Os algoritmos para decomposição e reconstrução para os casos de uma e duas dimensões são descritos em seguida, bem como, um breve comentário sobre as famílias wavelets empregadas.

O capítulo 4 refere-se a descrição do sistema implementado durante as atividades de pesquisa. Os algoritmos usados para normalização em escala na etapa de pré-processamento, para a representação dos numerais na etapa de extração de características e para a decisão final na etapa de classificação são explicados. Uma introdução sobre redes neurais multicamadas e sobre o algoritmo de treinamento de retropropagação do erro também é feita.

O capítulo 5 apresenta os resultados experimentais obtidos. A base de dados numérica escolhida para os testes é descrita, como também, os parâmetros usados para avaliação de desempenho dos sistemas de reconhecimento. O capítulo é concluído com a análise dos resultados e comparação do sistema implementado com outros sistemas desenvolvidos.

O capítulo 6 contém a conclusão do trabalho e suas principais contribuições. Propostas de trabalhos futuros também são sugeridas no final do capítulo.

## Capítulo 2

# Reconhecimento de Caracteres Numéricos Manuscritos

O principal objetivo deste capítulo é descrever o processo geral de reconhecimento de caracteres numéricos manuscritos. As técnicas de extração de características serão abordadas, como também os tipos de classificadores existentes. Uma introdução ao uso da transformada wavelet como extrator de características será feita e, por fim, uma revisão bibliográfica dos principais sistemas de reconhecimento de numerais será apresentada.

### 2.1 Modelo do sistema de reconhecimento

O reconhecimento de caracteres numéricos manuscritos pode ser visto como uma operação típica de classificação de imagens digitais, o qual é composto por uma seqüência de quatro procedimentos: aquisição da imagem, pré-processamento, extração de características e classificação [6, 7, 8, 9]. As etapas de aquisição e pré-processamento estão relacionadas com a preparação das imagens (qualidade visual), já as etapas de extração de características e classificação dizem respeito ao processo propriamente dito de interpretação ou reconhecimento de padrões. A Figura 2.1 ilustra a arquitetura típica de um sistema de reconhecimento de manuscritos.

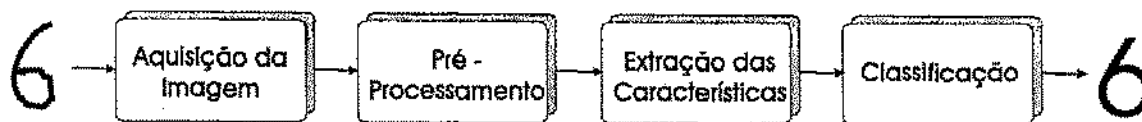


Figura 2.1: Arquitetura do sistema de reconhecimento de manuscritos.

### 2.1.1 Aquisição da imagem

Numa abordagem simplificada, o termo *imagem* se refere a uma função contínua da intensidade de luz refletida dos objetos, denotada por  $F(x, y)$ , cujo valor ou amplitude nas coordenadas espaciais  $(x, y)$  fornecem a intensidade (brilho) da imagem naquele ponto. Para que essa função possa ser processada convenientemente em um computador, é necessário convertê-la para uma forma digital, isto é, uma aproximação discreta da função contínua  $F(x, y)$ . Mais especificamente, uma imagem digital é uma matriz de amostras de  $F$  igualmente espaçadas. O valor de cada elemento equivale ao nível de cinza na mesma posição da imagem contínua. Estas amostras são denominadas elementos da imagem, elementos de quadro, ou simplesmente *pixels* (*picture elements*) [10].

A aquisição da imagem é a fase inicial em qualquer sistema de reconhecimento e tem como finalidade representar a imagem do caracter na sua forma digitalizada. Dependendo da relação entre o reconhecimento e aquisição os sistemas podem ser classificados em *on-line* ou *off-line*.

Nos sistemas *on-line*, o computador reconhece o caracter à medida que a escrita está sendo realizada, através de dispositivos acoplados ao computador, como por exemplo, uma mesa digitalizadora ou caneta eletrônica. Nos sistemas *off-line*, os caracteres são escritos em um papel com caneta convencional e depois digitalizados através de um *scanner* ótico, em que o reconhecimento é desempenhado após o término da escrita.

Em geral, os sistemas de reconhecimento *on-line* apresentam melhores desempenhos que os sistemas *off-line* [11, 12, 8]. Informações dinâmicas da escrita como pressão, velocidade, aceleração, movimento e posição da caneta são fatores adicionais existentes nos dados *on-line* que permitem detectar com certa facilidade a ocorrência de traçados sobrepostos [11]. O interesse em se trabalhar com o reconhecimento *off-line* está rela-

cionado principalmente ao fato desta abordagem preservar ao máximo a naturalidade da escrita, não requerendo dispositivos especiais, nem tampouco a presença do autor durante a fase de aquisição, pressupostos estes, que, são necessários para se fazer o reconhecimento *on-line* [12].

### 2.1.2 Pré-processamento

A etapa de pré-processamento deve melhorar a qualidade da imagem digitalizada, no intuito de obter uma maior probabilidade de êxito no reconhecimento da imagem. Os objetivos são corrigir defeitos e imperfeições, que surgiram durante o processo de digitalização, bem como reduzir as variações presentes nos caracteres manuscritos e tornar o conjunto de dados o mais consistente possível [8, 13]. Quanto melhor for a qualidade da imagem após o pré-processamento, maior será a possibilidade de sucesso no reconhecimento. Nessa etapa operam os algoritmos de redução de ruído, realce de imagem e normalização de caracteres.

### 2.1.3 Extração de características

A extração de características busca representar a imagem do caracter por um conjunto de atributos relevantes que são os responsáveis diretos pelo reconhecimento. Este conjunto de características precisa possuir o máximo de informação sobre o caracter a ser reconhecido. Boas características precisam satisfazer os seguintes requerimentos: (i) pequena invariância intraclasse: os caracteres da mesma classe precisam ter valores numéricos próximos; (ii) grande separação interclasse: os caracteres de classes distintas devem possuir valores numéricos bastante diferentes [7].

### 2.1.4 Classificação

A tarefa do classificador é identificar a classe a qual o caracter analisado pertence, com base nas características extraídas sobre o mesmo. Dessa forma, a escolha do classificador é dependente do tipo de características extraídas. Para o caso de caracteres numéricos, normalmente são adotadas dez classes, uma para cada numeral.

## 2.2 Técnicas de extração de características

A eficiência de um sistema de reconhecimento é dependente de cada uma das etapas que fazem parte do processo. Entretanto, a seleção de características representativas é o fator mais importante para a obtenção de boas taxas de reconhecimento, uma vez que o estágio seguinte baseia-se apenas nestas características para a tomada de decisão[7]. Dessa forma, é essencial que estas características sejam independentes das variações e distorções apresentados pelos caracteres.

Numerosas técnicas de extração de características têm sido desenvolvidas para capturar informações capazes de representar os caracteres manuscritos, variando de acordo com os atributos extraídos e o modo de extração destes atributos [3, 7]. O objetivo desta seção é apresentar as principais técnicas de extração em uso no reconhecimento de caracteres. Suen [3] dividiu essas técnicas em três categorias: características globais, distribuição dos pontos e características topológicas ou geométricas. Essas categorias foram analisadas e comparadas segundo os critérios de sensibilidade às distorções e facilidade de implementação. Tais critérios são descritos a seguir:

1. **Sensibilidade às distorções nas imagens dos caracteres** - os caracteres podem sofrer diversos tipos de alterações, causadas por:
  - (a) Ruído - elemento não pertencente a imagem original que introduz descontinuidades nos segmentos de linha, lacunas, segmentos espúrios, pontos isolados, etc;
  - (b) Distorção - inclui variações locais, cantos arredondados e variações de proporção no tamanho;
  - (c) Variações de estilo - consiste no uso de diferentes formas para representar um mesmo caracter, bem como no uso de traços e inclinações para personalizar a escrita;
  - (d) Translação - movimento do caracter inteiro ou de seus componentes em relação a uma determinada posição de referência;
  - (e) Rotação - mudança na orientação do caracter.

O ruído é geralmente introduzido pelo processo de digitalização da imagem do caracter, bem como pela superfície ou instrumento utilizados na escrita. As distorções e variações de estilos são principalmente produzidas pelo escritor. As translações e rotações podem ser resultado tanto do escritor como da parte mecânica do sistema (digitalizador, superfície e instrumento de escrita). Geralmente, esses fatores devem ser atenuados ou eliminados durante o pré-processamento da imagem.

2. **Implementação prática da técnica** - esta pode ser avaliada pelos seguintes fatores:

- (a) Facilidade na implementação de máscaras - avalia a facilidade em se construir automaticamente máscaras que correspondam a cada classe de caracteres, com a finalidade de obter diferentes características usadas no método em questão;
- (b) Velocidade - rapidez com que os caracteres são reconhecidos;
- (c) Complexidade - determina o grau de dificuldade na implementação de um determinado método;
- (d) Independência - diz respeito ao fato de uma única técnica ser capaz de atingir determinado desempenho sem ser necessário o uso de técnicas suplementares.

### 2.2.1 Características globais

Essa categoria inclui técnicas que utilizam as características extraídas de todos os pontos pertencentes a um retângulo, o qual circunscreve o caracter em questão. Essas características representam a imagem como um todo, e não refletem propriedades locais, geométricas ou topológicas de uma região da imagem. A seguir, são apresentados os métodos que extraem características globais:

1. **Similaridade e correlação**: esta técnica toma como características os estados (branco ou preto) de cada um dos pontos pertencentes à imagem. Tal método mede a similaridade entre a imagem a ser reconhecida e um padrão de referência,

através da combinação e correlação entre os pontos ou grupo de pontos da imagem e do padrão de referência.

2. **Transformadas globais e expansão em séries:** estes métodos decompõem a imagem em uma combinação linear de funções de base. Foram criados com a finalidade de extrair características que fossem invariantes a alguma deformação global da imagem, tais como rotação e translação. Os mais explorados têm sido a transformada e série de Fourier, a transformada de Walsh, a transformada de Hadamard, bem como a transformada de Karhunen-Loeve.

Características globais podem ser facilmente extraídas, embora sofram algumas limitações. A técnica de similaridade e correlação é capaz apenas de reconhecer caracteres de mesmo tamanho e rotação. Portanto, muitas vezes se torna ineficiente para o caso de manuscritos [14]. Além disso, muitas comparações são necessárias para uma mesma classe de numeral, causando um alto custo computacional, uma vez que cada caracter de entrada deve ser comparado com todos os padrões de referência [7].

Transformações e expansões em série possuem grande facilidade de implementação e alta sensibilidade às distorções e variações de estilos. Por exemplo, manchas ou ruídos podem deslocar os pontos das imagens de sua posição de alinhamento, afetando seriamente as características obtidas por estas técnicas [8]. A propriedade de invariância à rotação dos descritores de Fourier [15, 7] pode criar dificuldades na classificação de caracteres que são similares na forma e cuja diferença pode ser atribuída a uma rotação ou translação. Como exemplo, técnicas que utilizam transformações e expansões em série, geralmente, confundem-se na distinção entre as classes 2 e 5, 6 e 9, e 4 e 7 [15].

### 2.2.2 Distribuição dos pontos

As técnicas desse grupo extraem atributos a partir da distribuição estatística dos pontos que formam a imagem do caracter, produzindo um conjunto de características de dimensão reduzida. Os métodos mais empregados são citados a seguir:

1. **Divisão em zonas:** o retângulo que contém o caracter é dividido em várias regiões, sobrepostas e não sobrepostas, denominadas zonas. As características



usadas para reconhecer o caracter refletem as densidades de pontos nessas regiões.

2. **Momentos estatísticos:** os momentos estatísticos dos *pixels* pretos em relação a um ponto de referência escolhido no caracter, tal como, o centro de gravidade ou uma outra coordenada do sistema, são usados como características do caracter.
3. **Características loci:** para cada *pixel* branco do fundo da imagem, um conjunto de vetores verticais e horizontais são gerados. O conjunto de características usado nesse caso, é dado pelo número de interseções que esses vetores fazem com os contornos que formam o caracter.
4. **Distâncias e cruzamentos:** a característica de cruzamento é obtida do número de vezes que o caracter é cortado por segmentos de linha especificados em determinadas direções. As distâncias entre os pontos que formam o caracter e pontos especificados na imagem (por exemplo, os pontos que determinam o limite do retângulo que contém o caracter), formam um outro conjunto de características.

Uma considerável tolerância às distorções e às pequenas variações de estilos são observados nas três últimas técnicas descritas, principalmente nas técnicas de características loci, distâncias e cruzamentos. Nesses grupos de técnicas, existe uma relativa dificuldade na implementação das máscaras, devido ao grande número de variações possíveis dessas características e variações no formato dos caracteres. Por outro lado, essas técnicas provêm alta velocidade de processamento.

### 2.2.3 Características topológicas ou geométricas

Essa categoria é baseada na extração de características que descrevem aspectos importantes da geometria e topologia do desenho do caracter. As técnicas incluídas nessa categoria podem representar propriedades globais ou locais do caracter, tais como:

1. **Segmentos de reta e linhas curvas em várias direções:** neste caso, são extraídos traços verticais, horizontais e diagonais, bem como convexidades e concavidades apresentados pela geometria do caracter.

2. **Pontos finais, interseções de linhas e cavidades:** estas características são representativas da topologia do caracter, e incluem os pontos finais, interseções e o número de buracos que o caracter possui.

Devido à complexidade para extrair as características geométricas e topológicas e a grande variação de estilos dos caracteres manuscritos, torna-se bastante difícil gerar máscaras que extraíam essas características. Entretanto, uma vez implementada essas técnicas, permitem ao sistema uma grande velocidade de processamento. Tais técnicas possuem altas tolerâncias às distorções e variações de estilos, e podem aceitar um certo grau de translação e rotação.

#### 2.2.4 Transformada wavelet como extrator de características

Diversos sistemas de reconhecimento de manuscritos baseados nas técnicas anteriormente descritas foram desenvolvidos nos últimos anos, os quais servem hoje como referenciais para a base de novas pesquisas realizadas nesse campo. Em muitos aspectos, suas qualidades e deficiências têm influenciado sobremaneira as pesquisas atuais, que buscam corrigir as discrepâncias encontradas em algumas implementações. Uma das últimas técnicas que surgiram, para a extração de características relevantes, usa a transformada wavelet, e foi escolhida para representar eficientemente os numerais no sistema implementado neste trabalho.

A transformada wavelet é uma nova ferramenta desenvolvida por cientistas na última década e que representa fielmente sinais não-estacionários, tais como sinais de voz e imagens. Algumas aplicações no processamento de imagens incluem filtragem espacial, compressão, análise de textura, detecção de bordas e extração de características [16, 17, 18, 19].

Como a transformada de Fourier, a transformada wavelet é usada na decomposição de sinais através de um conjunto de funções de base, nesse caso, denominadas wavelets. Essas funções representam o sinal em diferentes escalas e com diferentes resoluções. Por exemplo, quando o sinal é analisado por uma grande "janela" é possível identificar suas características gerais. Similarmente, se o sinal for analisado por uma "janela" pequena, particularidades ou detalhes, podem ser observados.

A decomposição wavelet da imagem do caracter é capaz de extrair características gerais e características direcionais, como os segmentos de linhas que compõem o caracter nas direções vertical, horizontal e diagonais. Esses três detalhes direcionais contêm os componentes de alta frequência da imagem original, enquanto, a informação global do caracter, os componentes de baixa frequência. Uma vez que os detalhes da imagem do caracter em diferentes resoluções geralmente caracterizam diferentes estruturas físicas do mesmo, os coeficientes obtidos a partir dessa transformada podem ser muito úteis no reconhecimento de caracteres numéricos manuscritos [5].

Por ser uma técnica recente, poucos trabalhos que utilizam wavelets foram desenvolvidos para o reconhecimento de manuscritos [4, 5]. Devido à propriedade de extração de atributos globais e locais, essa técnica não se enquadra em nenhuma das categorias propostas por Suen [3]. Apesar da implementação com máscaras ser uma tarefa complicada neste caso, os coeficientes wavelet podem ser facilmente obtidos através de um algoritmo em árvore composto por filtros passa-baixas e passa-altas [18, 19]. O sistema de reconhecimento desenvolvido por Lee et al. [5] mostrou-se robusto em relação aos vários estilos de escritas e tamanho dos caracteres.

## 2.3 Tipos de classificadores

Nos sistemas de reconhecimento de caracteres, o classificador escolhido deve ser adequado ao tipo e ao formato das características extraídas [7]. Normalmente, têm-se utilizado três categorias de classificadores: classificadores estruturais, classificadores estatísticos e redes neurais artificiais [13, 12, 7].

Os classificadores estruturais são também conhecidos como classificadores sintáticos, e são indicados para os sistemas que utilizam extratores de características geométricas ou topológicas [2]. Tal forma de classificação tem sido empregada em uma grande quantidade de sistemas de reconhecimento de padrões. Entretanto, é um método muito simples e pode não apresentar resultados satisfatórios [14]. Geralmente, a arquitetura desses classificadores é baseada na decisão em árvore.

Nos classificadores estatísticos, as características são da forma de n-uplas ou vetores. O objetivo desses classificadores consiste em estimar a probabilidade do caracter

analisado pertencer a uma das possíveis classes. As técnicas usadas para a classificação são regras de decisão, funções discriminantes e outros procedimentos estatísticos [12].

As redes neurais artificiais muitas vezes são consideradas como um tipo particular de classificador estatístico, uma vez que as características também são da forma de n-uplas ou vetores, e existe uma equivalência entre certos modelos de redes neurais com técnicas estatísticas fundamentais. Mas, como as redes neurais possuem propriedades peculiares, tais como a capacidade de generalização, abstração e aprendizagem a partir de exemplos, podem ser consideradas como uma categoria independente de classificadores [12].

Atualmente, os principais métodos de classificação desenvolvidos são baseados em redes neurais [20, 21, 14]. Pesquisas nessa área indicam que tal técnica produz melhores resultados do que os outros métodos de classificação, quando o problema em estudo é o reconhecimento de manuscritos. A vantagem em se usar redes neurais é que tal ferramenta pode encontrar limites de decisão não-lineares entre diferentes classes, e desse modo oferece uma maneira prática de solucionar problemas de classificação de padrões complexos [14]. Além de classificador, as redes neurais também podem ser usadas para a extração de características [12].

Muitos sistemas, utilizam uma combinação de mais de um tipo de classificadores no reconhecimento, no intuito de aumentar o desempenho. Nesse caso, os classificadores de múltipla decisão podem utilizar um ou outro esquema de classificação, em múltiplos estágios ou como classificadores paralelos, em que o resultado de cada classificação individual é utilizado para decidir a classificação final. A vantagem desses sistemas é que o erro específico produzido por um classificador, pode ser contrabalanceado pelos resultados dos outros classificadores [7].

## 2.4 Sistemas de reconhecimento: revisão bibliográfica

Nesta seção, alguns dos trabalhos mais significativos, publicados nos últimos anos, são revistos.

Shridhar et al. [22, 23] propuseram um algoritmo de reconhecimento estrutural que utiliza características derivadas originalmente do contorno esquerdo e direito da

imagem do caracter numérico. Alta velocidade de processamento e boa precisão foram obtidos, entretanto, esse algoritmo não rejeita amostras confusas ou ilegíveis.

Em um artigo interessante, Lam e Suen [24] descrevem um sistema de reconhecimento que consiste de uma combinação sequencial de um classificador estrutural rápido e um algoritmo de relaxação robusto, que determina a classe a qual o numeral pertence através de um critério de menor distância. A classificação é baseada na configuração de um conjunto de primitivas derivadas da imagem do numeral, que contém informação sobre convexidades e segmentos de linha.

Duas abordagens foram empregadas no trabalho desenvolvido por Veloso [8, 25]. A primeira abordagem é formada por um algoritmo de reconhecimento estrutural, que utiliza as seguintes características para a tomada de decisão: número e posição das cavidades do numeral, sequências de cruzamento e número de interseções. A segunda abordagem utiliza máscaras de Kirsch na extração de características direcionais, e redes neurais na classificação. Uma análise comparativa entre os dois algoritmos propostos foi realizada, e o algoritmo baseado em redes neurais forneceu melhores desempenhos do que o algoritmo estrutural.

Suen et al. [2] propuseram quatro métodos para o reconhecimento automático dos caracteres numéricos. Todos eles utilizam características topológicas e geométricas sobre o formato dos numerais, e classificadores estruturais para a identificação do numeral. Vários testes foram realizados, comparando a eficiência de cada sistema isolado e a combinação dos mesmos. Bons resultados foram obtidos, e a taxa de reconhecimento variou entre 86,05 a 93,90 %.

Em outro trabalho, Shridhar e Badrelein [15] desenvolveram um sistema de decisão múltipla para o reconhecimento de numerais manuscritos, que consiste de dois estágios. No primeiro, quinze descritores de Fourier são usados para descrever o formato dos caracteres. Uma medida de distância estatística foi adotada para medir a similaridade entre os vetores de características e vetores de referência. Devido à propriedade de invariância à rotação apresentada pelos descritores de Fourier, houve problemas no reconhecimento entre as classes de numerais 2 e 5, 6 e 9, e ocasionalmente, 4 e 7. Como solução, um segundo estágio foi adicionado usando um classificador baseado em descritores topológicos.

Recentemente, Knerr et al. [21] desenvolveram um classificador baseado em rede neural com uma única camada de treinamento, e que pode ser aplicado eficientemente ao reconhecimento de caracteres numéricos manuscritos. Nesse método, a decomposição é realizada por um procedimento denominado de STEPNET, que divide o problema do reconhecimento de caracteres em sub-problemas, os quais podem ser resolvidos por separadores lineares.

Cho et al. [20] descrevem um classificador de múltipla decisão, em que redes MLP (*Multilayer Perceptron*) são treinadas independentemente com características distintas: matriz normalizada, gradientes de Kirsch e características da seqüência do contorno. A classificação do caracter é obtida levando em conta a classificação fornecida por cada rede individualmente, utilizando um esquema concensual para decidir a classificação final. O método de fusão utilizado considera os diferentes desempenhos de cada rede na combinação do resultado final de classificação, o qual é baseado na noção de lógica *fuzzy*.

O sistema apresentado por Lee [26] é composto por dois estágios. No primeiro, cinco vetores de características são obtidos, um deles é formado pela imagem original do caracter comprimida e os quatro restantes, contém informações direcionais extraídas através da convolução da imagem do caracter com as máscaras de gradientes de Kirsch. No segundo estágio, cada um dos cinco vetores são apresentados a uma rede neural multicamadas com agrupamento. Este tipo de rede neural classifica numerais similares de forma bastante eficiente.

Kapogiannopoulos e Papadakis [4] desenvolveram um sistema que usa a transformada wavelet biortogonal para a extração de características dos contornos dos numerais. Tal método aplica a wavelet Cohen-Daubechies-Feauveau (CDF) 3/9 em um vetor que contém as coordenadas dos pontos que fazem parte do contorno. Os detalhes e a aproximação global obtida nos cinco primeiros níveis de resolução são usados como características. No estágio de classificação, um algoritmo LVQ (*Learnig Vector Quantization*) é usado como classificador. Este sistema apresenta resultados satisfatórios apenas para os numerais manuscritos bem definidos, uma vez que o contorno extraído não pode apresentar falhas, ou seja, alguns *pixels* em branco.

Lee et al. [5], propuseram um sistema de reconhecimento para numerais manuscritos

que usa as wavelets de Haar no processo de extração de características multiresolucionais, e uma rede neural multicamadas com agrupamento, treinada com o algoritmo de retropropagação do erro estocástico, para classificação das características extraídas. Dois tipos de vetores com as características foram consideradas. O primeiro tipo usa apenas as características extraídas no primeiro nível de resolução, e o segundo tipo todas as características obtidas nos dois primeiros níveis de resolução. Este esquema apresentou bons resultados, e mostrou-se tolerante aos vários estilos de escrita e tamanho apresentados pelos numerais.

O presente trabalho, baseia-se no algoritmo de Lee et. al, e determina um sistema de reconhecimento que utiliza a wavelet biortogonal CDF 3/7 na extração de características e uma rede neural multicamadas com agrupamento, treinada com o algoritmo de retropropagação do erro usando momento, no estágio de classificação [27].

## 2.5 Conclusão

O reconhecimento de caracteres numéricos manuscritos é uma das áreas mais importantes do reconhecimento automático de padrões e nas últimas duas décadas, tem sido bastante estudado pela comunidade científica.

Durante esse período, várias técnicas de extração de características, bem como vários tipos de classificadores foram propostos e investigados. Uma das técnicas mais recentes, é baseada na transformada wavelet, e extrai características globais e direcionais das imagens dos caracteres. Obviamente, a escolha da técnica de extração e do tipo de classificador a ser usado, depende do sistema de reconhecimento de caracteres considerado. Uma vez determinado o sistema, as técnicas para o pré-processamento dos dados são escolhidas de forma a ajustar os caracteres de entrada para a melhor representação possível, capaz de elevar os índices de reconhecimento.

## Capítulo 3

# Transformada Wavelet

A análise wavelet, assim como a análise de Fourier, é uma técnica usada para representar, descrever e processar sinais através da combinação linear de funções elementares. Enquanto a transformada de Fourier utiliza uma série de exponenciais complexas de diferentes frequências, a transformada wavelet utiliza funções que são versões escalonadas e transladadas de uma função modelo conhecida como wavelet mãe. As wavelets, ao contrário das senóides que oscilam infinitamente, são ondas com comprimento e duração limitadas. Por esta razão, são indicadas para análise de sinais não-estacionários, tais como imagens, uma vez que permitem associar características no domínio da frequência com sua localização no tempo [16, 17].

O conceito de wavelets é resultado de uma síntese de idéias originadas independentemente em diversas áreas, como matemática, física quântica, engenharia elétrica e geologia [28]. Embora seus princípios básicos sejam derivados da transformada de Haar de 1909, apenas no início dos anos 80 é que foram matematicamente formuladas. Grossmann e Morlet desenvolveram o estudo da transformada wavelet contínua. Daubechies, Grossmann e Meyer iniciaram o conceito de transformada discreta com a teoria de *frames*. Por fim, Mallat desenvolveu o conceito matemático para a análise multiresolucional, permitindo a construção de bases wavelets e o cálculo da transformada wavelet discreta através de um algoritmo recursivo chamado transformada wavelet rápida [29].

Atualmente, as wavelets podem ser encontradas em uma grande quantidade de apli-



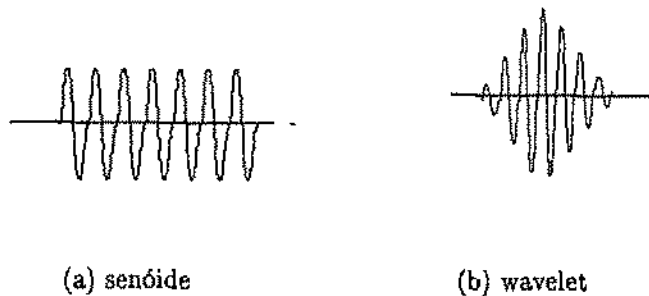


Figura 3.1: Ondas e wavelets.

cações incluindo processamento de sinais e imagens, análise numérica, comunicações digitais, teoria de sistemas, compressão de sinais e computação gráfica, dentre outras [30, 16]. Em particular, no processamento de imagens, wavelets tem sido empregadas para análise de textura, filtragem espacial, detecção de bordas, extração de características e compressão [30, 31].

Neste capítulo, apenas os conceitos básicos necessários para o entendimento do trabalho desenvolvido serão apresentados. Uma abordagem aprofundada sobre as wavelets pode ser encontrada em vários livros [32, 17, 33, 28]. Um maior entendimento sobre o conceito de análise multiresolucional e sua formulação matemática é fornecido nos artigos escritos por Mallat [18, 19].

### 3.1 Propriedades da transformada wavelet

A transformada wavelet tem se mostrado uma ferramenta eficiente na análise de um grande número de sinais, principalmente daqueles não-estacionários. As propriedades que tornam as wavelets úteis e interessantes são [32, 17]:

- *Representação tempo-freqüência.* A transformada wavelet fornece informações importantes sobre o sinal através da análise simultânea de características no domínio do tempo e da freqüência. Em outras palavras, as wavelets são locais em ambos os domínios e podem detectar facilmente as particularidades do sinal.

- *Análise multiresolucional.* A decomposição wavelet de um sinal permite a sua análise em diferentes níveis de resolução ou escalas. Em cada resolução, diferentes características podem ser observadas.
- *Bases wavelets.* A transformada wavelet não é única e existe na literatura um grande número de possíveis funções de base, com diferentes formas e tamanhos, satisfazendo todos os requisitos matemáticos. A seleção de uma delas é dependente da aplicação.
- *Representação esparsa.* Devido à propriedade de localização no plano tempo-frequência, os coeficientes da transformada são esparsamente representados, ou seja, apenas um pequeno número de coeficientes são não nulos. Esta é uma propriedade extremamente importante para aplicações como compressão de sinais e imagens.
- *Cálculo eficiente.* A transformada wavelet discreta pode ser implementada em computadores digitais, através de um algoritmo conhecido como transformada wavelet rápida, com complexidade  $O(N)$ , em que  $N$  representa o comprimento do sinal. Assim, a quantidade de multiplicações de ponto flutuante e adições aumentam linearmente à medida que o comprimento do sinal aumenta.

### 3.2 Conceitos básicos

Uma wavelet mãe, representada por  $\psi(t)$ , é uma função oscilatória de duração finita que deve satisfazer a condição de admissibilidade, isto é,

$$\int_{-\infty}^{\infty} \psi(t) dt = 0, \quad (3.1)$$

para garantir que uma função decomposta em uma combinação linear de wavelets possa retornar a sua forma original.

As wavelets são geradas através de mudanças de escala e translações da wavelet  $\psi(t)$  e podem ser representadas como,

$$\psi_{a,b}(t) = \frac{1}{\sqrt{a}} \psi \left( \frac{t-b}{a} \right) \quad a, b \in \mathbb{R} \quad (a \neq 0) \quad (3.2)$$

em que  $a$  é o fator de escala,  $b$  o fator de translação e  $a^{-1/2}$  o fator de normalização que mantém a mesma energia para todas as wavelets independente da escala utilizada.

O fator de translação é responsável pela localização das wavelets no tempo e o fator de escala por sua localização no domínio da frequência. Altas escalas ( $a > 1$ ) correspondem a versões expandidas da wavelet mãe e proporcionam uma análise em baixas frequências do sinal. De modo contrário, pequenas escalas ( $a < 1$ ) correspondem a versões comprimidas da wavelet mãe e permitem uma análise em altas frequências. Dessa forma, as wavelets são usadas para decompor um sinal em componentes de diferentes faixas de frequências e analisar cada componente com uma resolução ajustada à sua escala. A Figura 3.2 é o exemplo de uma wavelet escalonada por diversos valores. Observa-se que em todos os casos as wavelets possuem o mesmo formato.

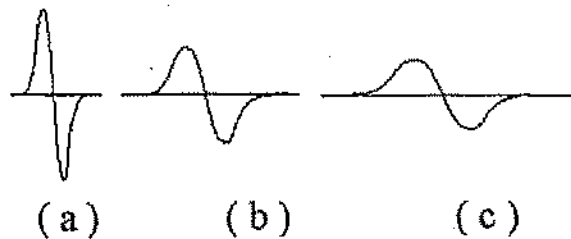


Figura 3.2: A wavelet e suas versões escalonadas: (a)  $a < 1$ , (b)  $a = 1$ , (c)  $a > 1$ .

Quando os fatores  $a$  e  $b$  são restritos a valores discretos, como  $a = a_0^j$  e  $b = ka_0^j b_0$ , em que  $a_0 > 1$ ,  $b_0 > 0$  e  $j, k \in \mathbb{Z}$ , um conjunto discreto de wavelets é obtido, o qual é representado por:

$$\psi_{j,k}(t) = a_0^{-j/2} \psi \left( a_0^{-j} t - kb_0 \right). \quad (3.3)$$

Um caso particular interessante ocorre para os valores  $a_0 = 2$  e  $b_0 = 1$ , que correspondem à amostragem de ordem 2 na frequência. Nesse caso, é possível construir funções  $\psi(\cdot)$  em que o conjunto,

$$\psi_{j,k}(t) = 2^{-j/2} \psi(2^{-j}t - k) \quad (3.4)$$

é ortonormal, ou seja,

$$\langle \psi_{j,k}, \psi_{m,n} \rangle = \begin{cases} 1, & \text{se } j = m \text{ e } k = n \\ 0, & \text{caso contrário.} \end{cases} \quad (3.5)$$

Então, a transformada wavelet discreta usando as bases wavelets ortonormais é representada pela equação,

$$DWT_f(j, k) = \langle f, \psi_{j,k} \rangle = |2|^{-j/2} \int_{-\infty}^{\infty} f(t) \psi(2^{-j}t - k) dt. \quad (3.6)$$

em que o produto interno  $\langle f, \psi_{j,k} \rangle$  é conhecido como coeficiente wavelet discreto. A fórmula de reconstrução de uma função  $f(t) \in L^2(\mathbb{R})$  é fornecida pela equação 3.7:

$$f(t) = \sum_{j,k} \langle f, \psi_{j,k} \rangle \psi_{j,k}(t). \quad (3.7)$$

Nesse caso aplicando-se a transformada direta e depois a transformada inversa, a função  $f$  é totalmente reconstruída, porque as famílias wavelets formam bases ortonormais.

### 3.3 Wavelets e análise multiresolucional

Existem duas maneiras de se calcular a transformada wavelet discreta: a primeira através da formulação matemática descrita na seção anterior e a segunda através da teoria de análise multiresolucional. A seguir serão apresentados os conceitos de análise multiresolucional e os algoritmos de reconstrução e decomposição.

#### 3.3.1 Análise multiresolucional

A análise multiresolucional é uma teoria que provê um modelo matemático para representar sinais em diferentes escalas ou resoluções. Essa análise pode ser definida como a decomposição do espaço das funções de quadrado integrável  $L^2(\mathbb{R})$  em uma seqüência de subespaços de aproximação  $\{V_j\}_{j \in \mathbb{Z}}$  que satisfazem à seguinte relação:

$$V_j \subset V_{j-1} \quad j \in \mathbb{Z}, \quad (3.8)$$

com

$$\bigcap_{j=-\infty}^{\infty} V_j = \{0\}, \quad \bigcup_{j=-\infty}^{\infty} V_j = L^2(\mathbb{R}). \quad (3.9)$$

As bases do subespaço  $V_0$  são geradas através de translações de uma função de norma unitária  $\phi(t)$ , denominada função escalonamento, para diversos valores discretos. Da mesma forma, para um valor fixo  $j$ , as translações  $\{2^{-j/2}\phi(2^{-j}t - k)\}_{k \in \mathbb{Z}}$  formam as bases ortonormais do subespaço  $V_j$ .

Observa-se que à medida que a escala  $j$  decresce, as funções  $2^{-j/2}\phi(2^{-j}t - k)$  se tornam cada vez mais compactas no tempo. Quanto menor a escala maior a resolução, ou seja, maior a representação das características presente no sinal. Logo, a condição de causalidade, definida em 3.8, significa que o subespaço  $V_{j-1}$  possui uma maior resolução que o subespaço  $V_j$ .

Associado a cada subespaço  $V_j$  existe um subespaço  $W_j$  que contém os detalhes necessários para passar do nível de resolução  $j$  para o nível de resolução  $j - 1$ . Dessa forma,  $W_j$  representa o complemento ortogonal de  $V_j$  em relação a  $V_{j-1}$  e pode ser representado pela soma direta,

$$V_{j-1} = V_j \oplus W_j, \quad V_j \perp W_j. \quad (3.10)$$

Dessa forma, toda informação presente em uma determinada resolução, é mantida nos níveis de resolução superiores.

Aplicando-se equação 3.10 recursivamente, obtém-se a seguinte decomposição:

$$V_{j-1} = V_{j+1} \oplus W_{j+1} \oplus W_j, \quad (3.11)$$

$$= V_{j+2} \oplus W_{j+2} \oplus W_{j+1} \oplus W_j, \quad (3.12)$$

em geral,

$$V_j = V_j \oplus W_j \oplus W_{j-1} \oplus W_{j-2} \oplus \dots \oplus W_{J+1} \quad \text{para } j > J. \quad (3.13)$$

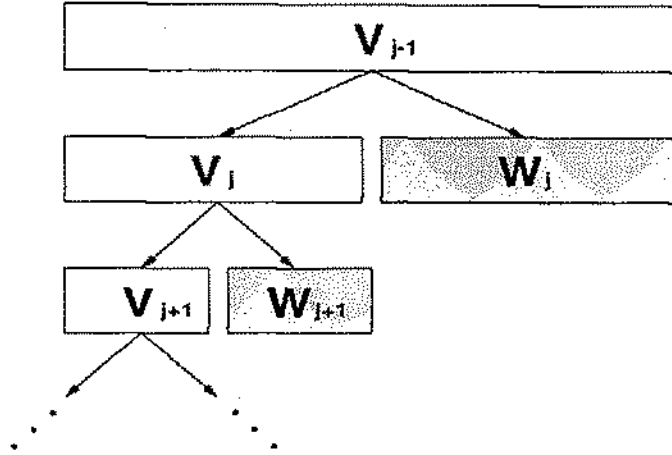


Figura 3.3: Decomposição wavelet representada pelos subespaços de aproximação  $\{V_j\}_{j \in \mathbb{Z}}$  e detalhes  $\{W_j\}_{j \in \mathbb{Z}}$ .

Quando  $j$  tende a infinito, a relação 3.13 pode ser representada por,

$$L^2(\mathbb{R}) = \bigoplus_{j=-\infty}^{\infty} W_j = \dots \oplus W_{-1} \oplus W_0 \oplus W_1 \oplus \dots \quad (3.14)$$

a qual é conhecida como decomposição wavelet, uma vez que cada subespaço  $W_j$  é gerado através das funções wavelets  $\{2^{j/2}\psi(2^j t - k)\}_{k \in \mathbb{Z}}$ .

### 3.3.2 Algoritmos de decomposição e reconstrução

Desde que  $V_0$  e  $W_0$  são subespaços de  $V_{-1}$ :

$$V_0 \subset V_{-1} \text{ e } W_0 \subset V_{-1}, \quad (3.15)$$

pode-se representar as funções  $\phi(t) \in V_0$  e  $\psi(t) \in W_0$  como combinações lineares das funções base de  $V_{-1}$ :

$$\phi(t) = \sum_n h(n)\phi(2t - n), \quad (3.16)$$

$$\psi(t) = \sum_n g(n)\phi(2t - n), \quad (3.17)$$

em que as seqüências finitas  $\{h(n)\}$  e  $\{g(n)\}$  representam coeficientes de filtros discretos passa-baixas e passa-altas, respectivamente. As equações 3.16 e 3.17 são conhecidas como equações de dilatação entre duas escalas. Para que a função escalonamento exista é necessário que os coeficientes do filtro passa-baixas satisfaçam à seguinte condição:

$$\sum_n h(n) = 1. \quad (3.18)$$

Como  $V_{j-1} = V_j \oplus W_j$ , uma função  $f(t) \in L^2(R)$  expressa em termos das funções base de  $V_{j-1}$ , pode ser representada pelas funções base de  $V_j$  e  $W_j$ , ou seja,

$$f(t) = \sum_k c_{j-1}(k) 2^{-(j-1)/2} \phi(2^{-(j-1)}t - k), \quad (3.19)$$

$$= \sum_k c_j(k) 2^{-j/2} \phi(2^{-j}t - k) + d_j(k) 2^{-j/2} \psi(2^{-j}t - k) \quad (3.20)$$

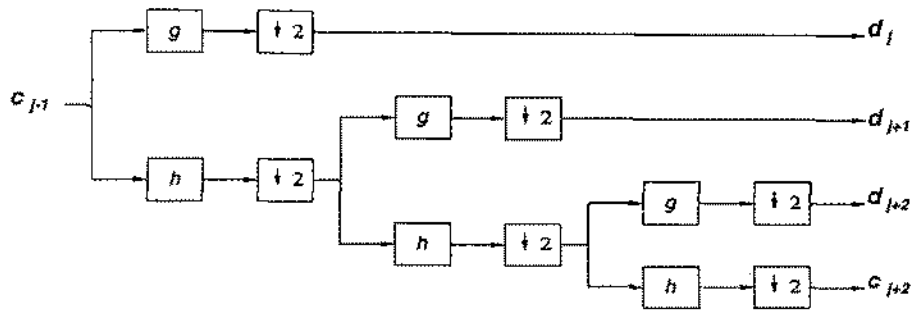
em que os coeficientes  $c_j(k)$  e  $d_j(k)$  são definidos por:

$$c_j(k) = \sum_m h(m - 2k) c_{j-1}(m), \quad (3.21)$$

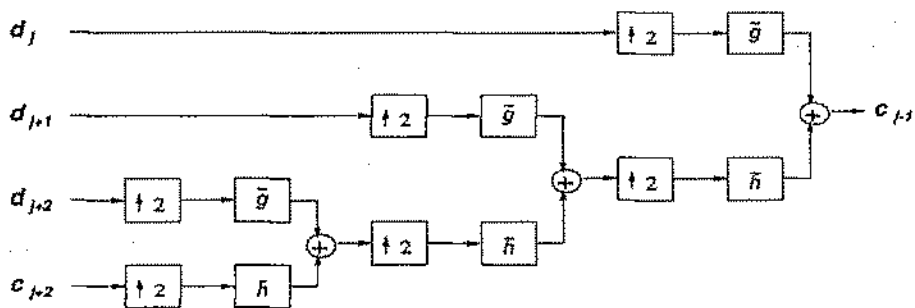
$$d_j(k) = \sum_m g(m - 2k) c_{j-1}(m). \quad (3.22)$$

O algoritmo da transformada wavelet rápida é obtido através das equações 3.21 e 3.22. Estas equações mostram que os coeficientes wavelets  $c_j(k)$  e  $d_j(k)$  são obtidos a partir de convoluções discretas com os filtros de análise passa-baixas  $h(n)$  e passa-altas  $g(n)$ , respectivamente. O fator  $2k$  no índice dos filtros representa a decimação por um fator 2. Na literatura de processamento de sinais, este algoritmo é conhecido como banco de filtros e pode ser visto na Figura 3.4.

Para cada nível de resolução  $j$  é obtida uma aproximação  $c_j(k)$  e um detalhe  $d_j(k)$ . Os termos aproximação e detalhe correspondem, respectivamente, às componentes de baixas e altas frequências. Se o sinal original  $c_{j-1}(k)$  possui  $N$  amostras, então os sinais  $c_j(k)$  e  $d_j(k)$  possuem  $2^{-j}N$  amostras. Dessa forma, a decomposição wavelet é formada por seqüências que possuem o mesmo tamanho do sinal de entrada. Na prática  $N$  deve ser potência de 2.



(a) Decomposição para três níveis de resolução.



(b) Reconstrução para três níveis de resolução

Figura 3.4: Algoritmos de decomposição e reconstrução wavelet uni-dimensionais.

A transformada wavelet inversa pode ser obtida de maneira similar. Nesse caso, o sinal  $c_{j-1}(k)$  é reconstruído a partir dos sinais  $c_j(k)$  e  $d_j(k)$  através da equação:

$$c_{j-1}(k) = \sum_m \tilde{h}(k - 2m)c_j(m) + \tilde{g}(k - 2m)d_j(m), \quad (3.23)$$

como mostra a Figura 3.4. Em cada nível de resolução  $j$ ,  $c_j(k)$  e  $d_j(k)$  são combinados para recuperar o sinal  $c_{j-1}(k)$ . O fator  $2m$  representa uma interpolação por um fator 2. Isto significa que zeros são inseridos entre as amostras.

Os filtros de síntese  $\tilde{h}(n)$  e  $\tilde{g}(n)$  não são necessariamente iguais aos filtros de análise  $h(n)$  e  $g(n)$  usados na decomposição. Os filtros de síntese são filtros duais e correspondem as funções duais escalonamento e wavelet,  $\tilde{\phi}(t)$  e  $\tilde{\psi}(t)$ , que formam os subespaços



duais  $\tilde{V}_0$  e  $\tilde{W}_0$ . As equações relacionando essas quantidades duais são equivalente às descritas anteriormente.

Quando as wavelets são ortonormais os filtros de análise e síntese possuem o mesmo número de coeficientes e são reversos no tempo, ou seja,  $\tilde{h}(n) = h(-n)$  e  $\tilde{g}(n) = g(-n)$ . Entretanto, filtros ortonormais com resposta ao impulso finita FIR simétrica só podem ser obtidos para um comprimento 2, como é o caso das wavelets de Haar. Consequentemente, filtros FIR possuindo resposta ao impulso simétrica com comprimento maior que 2 precisam ser biortogonais e, neste caso  $\tilde{h}(n)$  e  $\tilde{g}(n)$  são diferentes de  $h(n)$  e  $g(n)$ . Além disso, os filtros passa-baixas e passa altas,  $h(n)$  e  $g(n)$ , biortogonais, podem possuir tamanhos diferentes. Por exemplo, a wavelet biortogonal Cohen-Daubechies-Feauveau 3/7 possui  $h(n)$  de comprimento 4 e  $g(n)$  de comprimento 12 [28]. Para o caso biortogonal, os filtros duais são relacionados por:

$$g(n) = (-1)^n \tilde{h}(1-n), \quad (3.24)$$

$$\tilde{g}(n) = (-1)^n h(1-n). \quad (3.25)$$

Os requisitos que os filtros biortogonais passa-baixas e passa-altas devem satisfazer para garantir a reconstrução perfeita do sinal original, são os seguintes:

$$\sum_n h(n) \tilde{h}(n+2k) = \delta(k), \quad (3.26)$$

$$\sum_n h(n) \tilde{g}(n+2k) = 0, \quad (3.27)$$

$$\sum_n \tilde{h}(n) g(n+2k) = 0, \quad (3.28)$$

$$\sum_n g(n) \tilde{g}(n+2k) = \delta(k). \quad (3.29)$$

### 3.3.3 Transformada Wavelet Aplicada a Imagens

Como visto anteriormente, a transformada wavelet multiresolucional é calculada através de projeção recursiva da função a ser representada nas funções wavelet e nas funções escalonamento. Na análise wavelet bi-dimensional, a função escalonamento  $\phi(x, y)$  pode ser definida como:

$$\phi(x, y) = \phi(x)\phi(y), \quad (3.30)$$

em que  $\phi(x)$  é uma função escalonamento uni-dimensional.

Considere uma wavelet unidimensional  $\psi(x)$  associada à função escalonamento  $\phi(x)$ . Então, três wavelets bi-dimensionais podem ser definidas:

$$\psi^1(x, y) = \phi(x)\psi(y), \quad (3.31)$$

$$\psi^2(x, y) = \psi(x)\phi(y), \quad (3.32)$$

$$\psi^3(x, y) = \psi(x)\psi(y). \quad (3.33)$$

O algoritmo da transformada wavelet rápida em duas dimensões, assim como no caso uni-dimensional, pode ser implementado através de filtragens individuais das linhas e colunas de uma imagem com filtros discretos passa-baixas  $h(n)$  e passa-altas  $g(n)$ , como mostra a Figura 3.5. Nesse caso, as linhas da imagem  $A_{j-1}$  são inicialmente convoluídas com os filtros  $h(n)$  e  $g(n)$  e, em seguida as duas imagens resultantes são decimadas por um fator 2, retirando-se as linhas ímpares. Da mesma forma, as colunas das imagens decimadas são convoluídas com os filtros passa-baixas e passa-altas e as linhas ímpares são excluídas no processo de decimação. Como resultado, quatro subimagens são obtidas  $A_j, D_j^1, D_j^2$  e  $D_j^3$ . As equações 3.34 a 3.37 mostram a representação matemática para cada uma delas:

$$A_j(k_1, k_2) = \sum_{m_1, m_2} h(m - 2k_1)h(m - 2k_2)A_{j-1}(m_1, m_2), \quad (3.34)$$

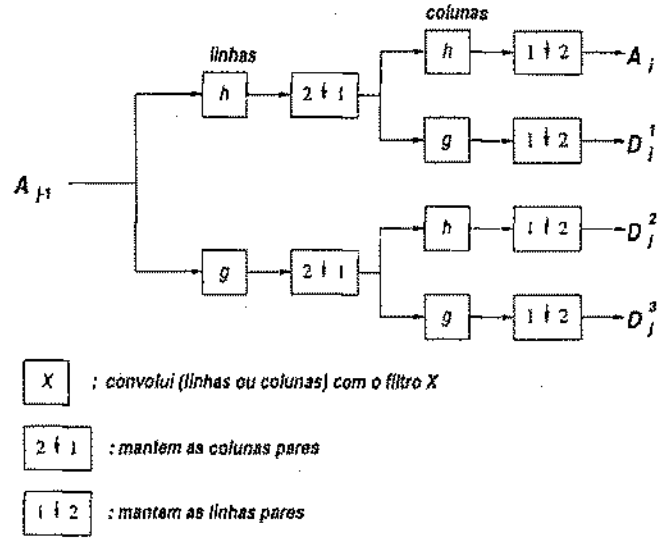


Figura 3.5: Algoritmo de decomposição wavelet bi-dimensional para um nível de resolução. A imagem  $A_{j-1}$  é decoposta nas subimagens  $A_j$  (características gerais),  $D_j^1$  (detalhes horizontais),  $D_j^2$  (detalhes verticais) e  $D_j^3$  (detalhes diagonais).

$$D_j^1(k_1, k_2) = \sum_{m_1, m_2} h(m - 2k_1)g(m - 2k_2)A_{j-1}(m_1, m_2), \quad (3.35)$$

$$D_j^2(k_1, k_2) = \sum_{m_1, m_2} g(m - 2k_1)h(m - 2k_2)A_{j-1}(m_1, m_2), \quad (3.36)$$

$$D_j^3(k_1, k_2) = \sum_{m_1, m_2} g(m - 2k_1)g(m - 2k_2)A_{j-1}(m_1, m_2). \quad (3.37)$$

A subimagem  $A_j$  fornece os componentes de baixa frequência (informação global),  $D_j^1$  os componentes de alta frequência na direção vertical (detalhes horizontais),  $D_j^2$  os componentes de alta frequência na direção horizontal (detalhes verticais) e  $D_j^3$  os componentes de alta frequência em ambas as direções verticais e horizontais (detalhes diagonais). A Figura 3.6 ilustra o exemplo da decomposição de um quadrado em dois níveis de resolução. Como esperado, as subimagens de detalhes do sinal caracterizam as direções horizontal, vertical e diagonais das bordas do quadrado. A disposição de cada subimagem na representação wavelet pode ser observada na Figura 3.7.

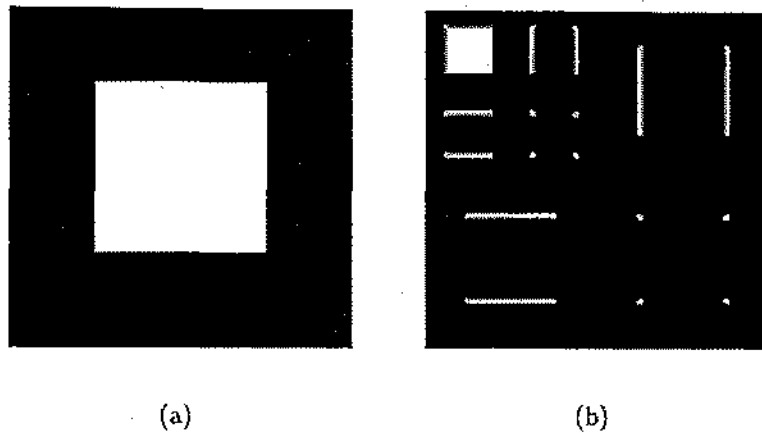


Figura 3.6: Decomposição wavelet da imagem de um quadrado em dois níveis de resolução usando as wavelets de Haar: (a) Imagem original de dimensão  $128 \times 128$  *pixels*, (b) Imagem com os coeficientes wavelets de dimensão  $128 \times 128$  *pixels*. Os *pixels* pretos e brancos correspondem, respectivamente, a valores nulos e não-nulos.

Se a imagem original possui  $N \times N$  *pixels*, cada subimagem possui  $2^{-j}(N \times N)$  *pixels*. O total de elementos na representação wavelet é igual ao número de elementos da imagem original, e o volume de dados não é alterado. Na prática o valor de  $N$  deve ser uma potência de dois.

O algoritmo de reconstrução pode ser facilmente obtido realizando o processo inverso. Seu esquema é mostrado na Figura 3.8.

### 3.3.4 Wavelet versus Fourier

A transformada de Fourier [10, 40] é uma técnica empregada na análise de sinais que utiliza exponenciais complexas de diferentes frequências para decompor funções de uma e duas dimensões. Essa ferramenta pode ser vista como uma transformação de um sinal no domínio do tempo para o domínio da frequência.

Para um grande número de sinais, a análise de Fourier é extremamente útil, uma vez que o conteúdo em frequência é muito importante. Entretanto, no processo de transformação informações sobre o comportamento do sinal no tempo são perdidas.

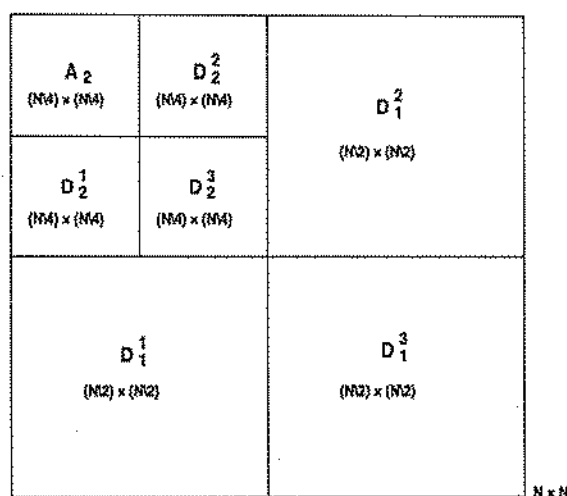
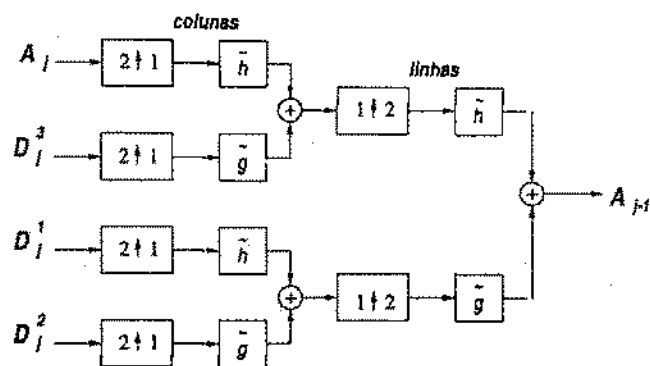


Figura 3.7: Disposição das subimagens de coeficientes wavelets  $A_j$ ,  $D_j^1$ ,  $D_j^2$  e  $D_j^3$  para dois níveis de resolução ( $j = 1, 2$ ).

Dessa forma, a transformada de Fourier de um sinal não fornece a localização no tempo (sinais uni-dimensionais) ou no espaço (sinais bi-dimensionais) de um evento específico. Assim, quando sinais apresentam transitórios ou informações de bordas, o início e o final destes eventos, são informações importantes que não podem ser perdidas, dependendo da aplicação. Logo, a análise de Fourier, nesses casos, não é um método adequado para a representação de sinais [32].

A transformada wavelet, como citado anteriormente, possui propriedades interessantes que permitem a análise simultânea do sinal em ambos os domínios, do tempo e da frequência. Logo, é indicada para a representação de sinais que apresentam transitórios e descontinuidades, já que as características locais podem ser identificadas, bem como a sua posição no tempo ou espaço. Tal fato é evidenciado na Figura 3.6, que ao contrário da decomposição de Fourier para a mesma imagem original (Figura 3.9), fornece a localização dos componentes de baixas frequências e altas frequências nas direções vertical, horizontal e diagonais.



- $X$  : convolui (linhas ou colunas) com o filtro  $X$
- $2 \uparrow 1$  : adiciona zeros entre as colunas
- $1 \uparrow 2$  : adiciona zeros entre as linhas

Figura 3.8: Algoritmo de reconstrução wavelet bi-dimensional para um nível de resolução. A imagem  $A_{j-1}$  é reconstruída a partir das subimagens  $A_j$  (características gerais),  $D_j^1$  (detalhes horizontais),  $D_j^2$  (detalhes verticais) e  $D_j^3$  (detalhes diagonais).

### 3.4 Famílias wavelets

Existe um grande número de famílias wavelets na literatura cujas qualidades variam de acordo com alguns critérios. Dentre eles, os principais são [34]:

- *Suporte compacto.* Se as wavelets e a função escalonamento possuem suporte compacto, os filtros  $h(n)$  e  $g(n)$  possuem resposta ao impulso finita, e os somatórios que definem a transformada wavelet discreta são finitos. Caso contrário, os filtros possuem resposta ao impulso infinita.
- *Simetria.* Se as wavelets e a função escalonamento são simétricas os filtros possuem fase linear. A ausência desta propriedade pode levar à distorção de fase, de grande importância no processamento de imagens, uma vez que, a imagem reconstruída a partir de coeficientes obtidos por wavelets não-simétricas pode apresentar borramentos [33].

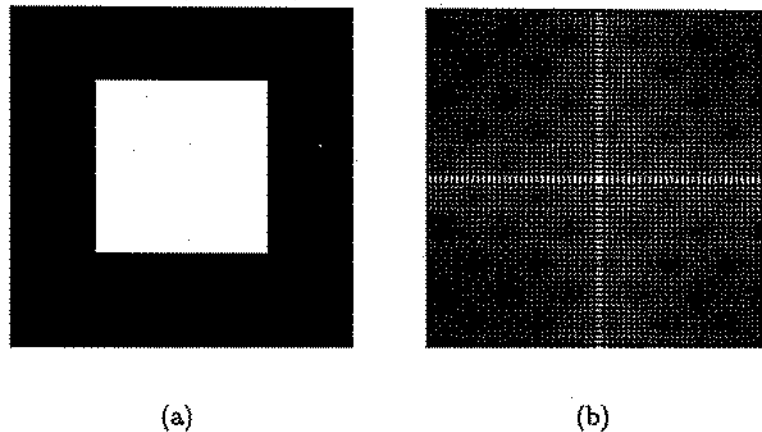


Figura 3.9: Decomposição de Fourier da imagem de um quadrado: (a) Imagem original de dimensão  $128 \times 128$  *pixels*, (b) Imagem com os coeficientes de Fourier de dimensão  $128 \times 128$  *pixels*.

- *Suavidade.* A suavidade é determinada pelo comprimento dos filtros. Quanto maior o comprimento mais suave são as funções escalonamento e wavelet. Esta é uma propriedade interessante, uma vez que filtros suaves são indicados para a representação de sinais suaves, com aplicação na análise de textura de imagens, e filtros não suaves são indicados para problemas de compressão e extração de características.
- *Ortogonal ou biortogonal.* Os filtros podem ser relacionados de maneiras diferentes em cada um desses sistemas como comentado na seção anterior.

A Tabela 3.1 cita as famílias wavelets usadas neste trabalho e as diferencia segundo alguns dos critérios listados anteriormente. Os coeficientes dos filtros para cada uma delas se encontra no livro de Daubechies [28].

### 3.5 Conclusão

A transformada wavelet é uma ferramenta eficiente para analisar características gerais e locais de sinais. Neste capítulo foram descritas as principais propriedades e conceitos

Tabela 3.1: Classificação das famílias wavelets.

Características	Haar	Daubechies	Coiflets	Cohen-Daubechies-Feauveau
Suporte Compacto	X	X	X	X
Simétrica	X			X
Quase Simétrica			X	
Assimétrica		X		
Análise Ortogonal	X	X	X	
Análise Biorrtogonal	X			X

básicos da teoria de wavelets. Em particular foi definido o conceito de análise multiresolucional e apresentado a transformada wavelet rápida para os casos uni-dimensional e bi-dimensional. Os coeficientes obtidos no caso de duas dimensões extraem aspectos importantes de imagens, como uma aproximação global e características direcionais. Estas características foram usadas no sistema de reconhecimento proposto para representar os numerais manuscritos. Por fim, algumas famílias wavelets foram analisadas.



## Capítulo 4

### Sistema Implementado

Os coeficientes da transformada wavelet de imagens de caracteres manuscritos contém uma grande quantidade de informação, podendo ser usados em um sistema de reconhecimento de numerais manuscritos.

Lee et al. [5] propuseram um sistema que consiste de três estágios: normalização em escala, extração de características multiresolucionais usando as wavelets de Haar e uma rede neural multicamadas com agrupamento para a classificação. O sistema implementado nesta dissertação é semelhante ao algoritmo de Lee exceto por três diferenças: a técnica usada para a normalização dos caracteres, a wavelet usada para a extração de características e o algoritmo de treinamento da rede. As seções seguintes descrevem cada um dos métodos utilizados.

#### 4.1 Normalização em escala

O propósito principal da etapa de pré-processamento é melhorar a informação pictorial das imagens dos caracteres, no intuito de torná-las as mais uniformes possíveis, e consequentemente, aumentar o desempenho dos algoritmos de extração de características e classificação.

As variações de tamanho apresentadas pelos numerais manuscritos podem degradar o desempenho de um sistema de reconhecimento. Visando atenuar essas variações, geralmente, utilizam-se algoritmos para normalizar o tamanho da imagem do caracter.

A normalização em escala é uma forma de pré-processamento muito utilizada para uniformizar as imagens dos caracteres em uma mesma dimensão.

O procedimento de normalização em escala é realizado principalmente em sistemas de reconhecimento por redes neurais, em que a entrada da rede é dependente da quantidade de *pixels* existentes na imagem do caracter [35]. Sendo assim, a padronização da imagem é um requisito fundamental para a definição de uma arquitetura fixa para a rede neural.

A escolha de um bom algoritmo de normalização é um fator muito importante. Quando o algoritmo é usado para diminuir as dimensões das imagens, normalmente ocorrem perdas de informação. Caso as informações perdidas sejam relevantes, o desempenho do sistema pode ser comprometido como um todo. Outro problema, ocorre quando o algoritmo realiza uma interpolação dos dados (aumento de dimensões), pois os *pixels* adicionados à imagem original podem deformá-la, prejudicando o desempenho do sistema [8].

O algoritmo selecionado para a normalização em escala dos numerais da base de dados utilizada nesta dissertação foi proposto por Veloso [8], baseado no método desenvolvido por Yuceer [35], fornecendo uma baixa complexidade de implementação e bons resultados. Esse algoritmo utiliza fatores de escala diferentes em cada um dos eixos  $x$  e  $y$  da imagem, para realizar a correção. Esse tipo de correção em escala tem-se mostrado eficiente na normalização de padrões digitalizados com resoluções e tamanhos diferentes [35].

A função de mapeamento para o escalonamento da imagem é dada pela equação:

$$f_s(x_i, y_j) = f(S_x x_i, S_y y_j). \quad (4.1)$$

A função  $f(x, y)$  representa a imagem original e a função  $f_s(x, y)$  representa a imagem normalizada.  $S_x$  e  $S_y$  são os fatores de escala na direção  $x$  e  $y$ , respectivamente, calculados através das seguintes equações:

$$S_x = \frac{X}{R_x}, \quad (4.2)$$

$$S_y = \frac{Y}{R_y}. \quad (4.3)$$

Os parâmetros  $R_x$  e  $R_y$  são as dimensões desejadas e  $X$  e  $Y$  são as dimensões originais da imagem. Na equação 4.1,  $S_x x_i$  e  $S_y y_j$  podem assumir valores fracionários e neste caso, deve-se fazer uma aproximação por limiar para torná-los numerais inteiros. Nas simulações foi adotada a seguinte convenção: se a parte fracionária for  $\leq 0,5$  aproxima-se para menos, caso seja  $> 0,5$  aproxima-se para mais.

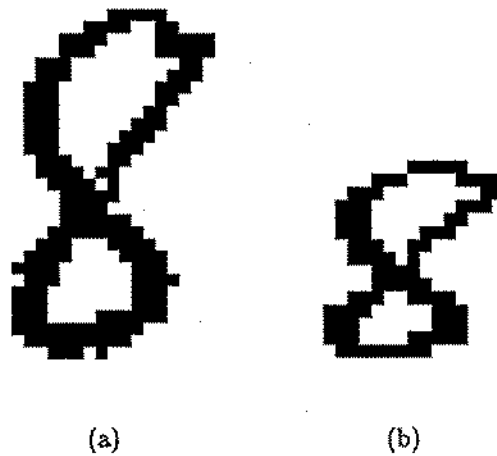


Figura 4.1: Exemplo de uma imagem normalizada do numeral 8: (a) Caracter original com  $29 \times 18$  *pixels*, (b) Caracter normalizado com  $16 \times 16$  *pixels*.

## 4.2 Extração de características

Os numerais manuscritos são essencialmente formados por linhas, ou seja estruturas uni-dimensionais em um espaço bi-dimensional. Dessa forma, a detecção local dos segmentos de linhas e a detecção global das estruturas das linhas aparecem como um método adequado para a extração de características [21]. Isto é possível usando a transformada wavelet.

O procedimento usado para extrair as características corresponde ao algoritmo da transformada wavelet rápida bi-dimensional descrito no capítulo anterior. Este decompõe a imagem do caracter em subimagens contendo características globais e características direcionais.

Considere  $F$  a imagem discreta do caracter arranjada na forma de uma matriz  $N \times N$  como expressa a equação 4.4,

$$F = \begin{bmatrix} a_{0,0}^{j-1} & a_{0,1}^{j-1} & \dots & a_{0,N-1}^{j-1} \\ a_{1,0}^{j-1} & a_{1,1}^{j-1} & \dots & a_{1,N-1}^{j-1} \\ \vdots & \vdots & \ddots & \vdots \\ a_{N-1,0}^{j-1} & a_{N-1,1}^{j-1} & \dots & a_{N-1,N-1}^{j-1} \end{bmatrix}_{N \times N} \quad (4.4)$$

As quatro subimagens obtidas a partir da convolução da imagem  $F$  com filtros discretos passa-baixas e passa-altas podem ser representadas pelas matrizes:

$$A_j = \begin{bmatrix} a_{0,0}^j & a_{0,1}^j & \dots & a_{0,(N/2)-1}^j \\ a_{1,0}^j & a_{1,1}^j & \dots & a_{1,(N/2)-1}^j \\ \vdots & \vdots & \ddots & \vdots \\ a_{(N/2)-1,0}^j & a_{(N/2)-1,1}^j & \dots & a_{(N/2)-1,(N/2)-1}^j \end{bmatrix}_{(N/2) \times (N/2)} \quad (4.5)$$

$$D_j^l = \begin{bmatrix} d_{l,0,0}^j & d_{l,0,1}^j & \dots & d_{l,0,(N/2)-1}^j \\ d_{l,1,0}^j & d_{l,1,1}^j & \dots & d_{l,1,(N/2)-1}^j \\ \vdots & \vdots & \ddots & \vdots \\ d_{l,(N/2)-1,0}^j & d_{l,(N/2)-1,1}^j & \dots & d_{l,(N/2)-1,(N/2)-1}^j \end{bmatrix}_{(N/2) \times (N/2)} \quad (4.6)$$

$l = 1, 2, 3.$

A subimagem  $A_j$  representa as informações globais,  $D_j^1$  as características horizontais,  $D_j^2$  fornece as características verticais e  $D_j^3$  as características diagonais.

Os índices  $(j-1)$  e  $j$  nas matrizes da imagem original  $F$  e nas subimagens  $A_j$  e  $D_j^l$  ( $l = 1, 2, 3$ ), respectivamente, representam o nível de resolução. De acordo com a teoria de análise multiresolucional [18, 19] o nível  $(j-1)$  possui uma maior resolução que o nível  $j$ .

1	1
1	1

(a)

1	-1
1	-1

(b)

1	1
-1	-1

(c)

1	-1
-1	1

(d)

Figura 4.2: Máscaras de Haar: (a) características gerais, (b) características verticais, (c) características horizontais, (d) características diagonais.

Para o caso das wavelets de Haar, a decomposição pode, também, ser realizada através da convolução da imagem do caracter com máscaras de Haar da seguinte forma:

$$D_j^k = \frac{1}{4} I_{j-1} \otimes H_k, \quad k = 0, 1, 2, 3 \quad (4.7)$$

em que  $\otimes$  é o operador de convolução,  $I_{j-1}$  é a imagem na resolução  $j - 1$ ,  $D_k^j$  são as subimagens na resolução  $j$ , e  $H_k$  são as máscaras de Haar definidas na Figura 4.2. Neste caso  $D_k^0$  fornece as características globais.

A Figura 4.3 ilustra o processo de extração de características multiresolucionais. O vetor de características é construído a partir das subimagens escolhidas para representar o numeral. Lee et al [5] utilizaram dois vetores, o primeiro continha todas as subimagens do primeiro nível de resolução e o segundo todas as subimagens do primeiro e segundo níveis de resolução obtidas através da convolução com as wavelets de Haar. No intuito de determinar o melhor conjunto de características outros vetores foram testados neste trabalho, e os resultados são descritos no próximo capítulo.

A escolha da wavelet é dependente da aplicação. Dessa forma, no intuito de aumentar o desempenho do sistema proposto por Lee et al. [5], outras famílias wavelets foram testadas no problema de reconhecimento de caracteres. A princípio, para determinar o efeito do comprimento do filtro, comparou-se o desempenho obtido por wavelets de várias ordens e de uma mesma família. Em seguida, o desempenho obtido com famílias diferentes também foi testado. O resultado destas comparações é descrito no capítulo seguinte. As wavelets selecionadas foram escolhidas de acordo com a facilidade de acesso aos coeficientes dos filtros e são as seguintes:

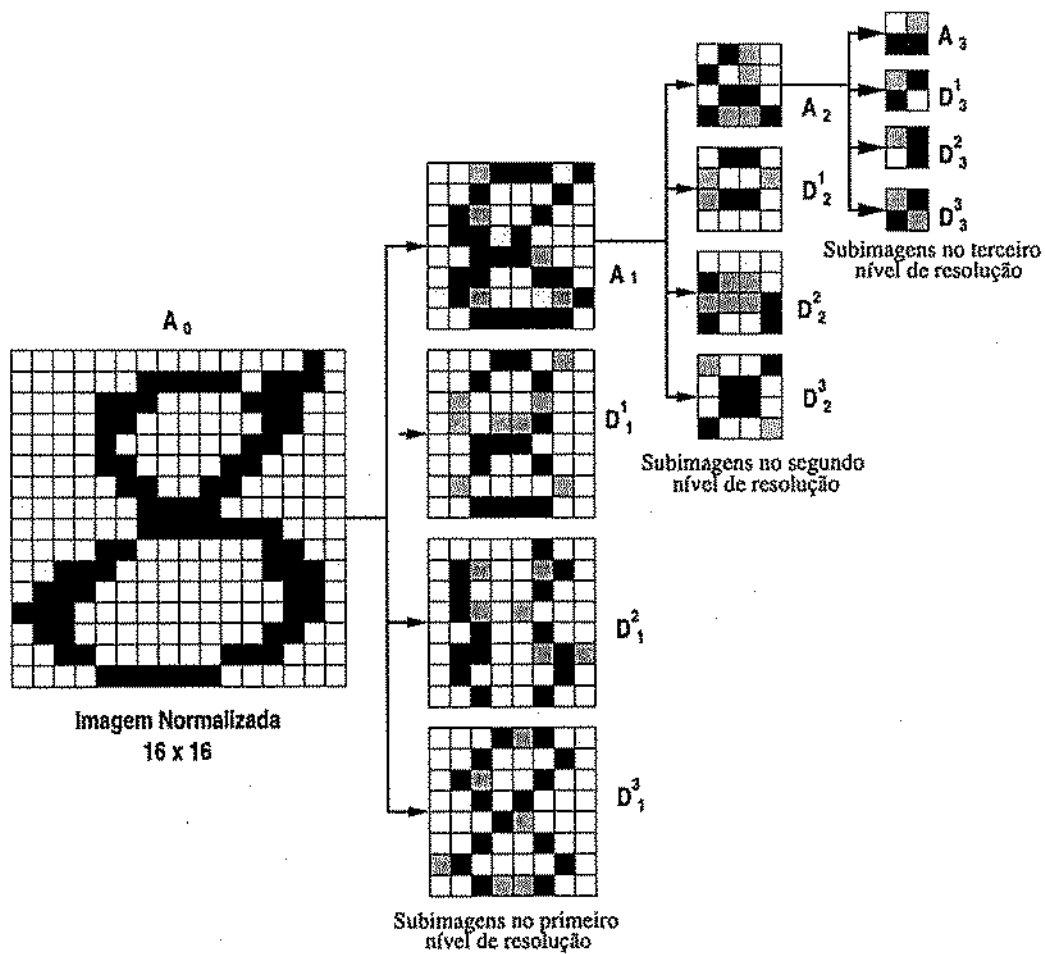


Figura 4.3: Extração de características usando as wavelets biortogonais Cohen-Daubechie-Feauveau 3/7 em três níveis de resolução.

1. Daubechies: 4,6,8,10,12,16 e 20;
2. Coiflets: 4, 6 e 8;
3. Cohen-Daubechies-Feauveau (CDF): 2/2, 2/4, 3/3 e 3/7.

As famílias wavelets Daubechies e Coiflets são ortogonais e as famílias CDF são biortogonais. Os coeficientes dos filtros para cada uma delas encontram-se disponíveis no livro de Daubechies [28].

Os coeficientes da transformada wavelet apresentam uma faixa muito grande de valores. Objetivando aumentar o desempenho do classificador, os coeficientes de cada subimagem foram normalizados para uma faixa de valores entre 0 e 1.

### 4.3 Redes neurais multicamadas com retropropagação do erro

As redes neurais multicamadas treinadas com o algoritmo de retropropagação do erro estão entre os mais difundidos e versáteis modelos de classificadores. Foi mostrado, em trabalhos anteriores [21, 36], que este tipo de rede neural contendo uma camada intermediária e uma função de ativação não-linear é um classificador universal. Isto é, tais redes podem determinar limiares de decisão de complexidade arbitrária. Em particular, têm-se usado as redes neurais com retropropagação extensivamente para o problema de reconhecimento de caracteres [7].

Tipicamente, a arquitetura de uma rede neural multicamadas consiste de um conjunto de neurônios que constituem a camada de entrada, uma ou mais camadas escondidas e a camada de saída. O número de neurônios na camada de entrada é determinada pela dimensão do vetor com as características, a quantidade de neurônios na saída é determinado pela resposta desejada, e, a quantidade de camadas escondidas e o número de neurônios em cada uma destas camadas são determinados pelo projetista da rede [37], em geral empiricamente. A Figura 4.4 apresenta um exemplo da rede neural contendo três camadas.

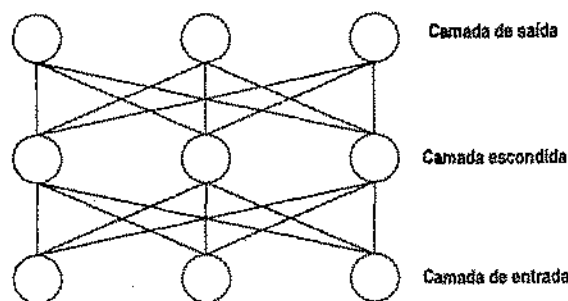


Figura 4.4: Arquitetura de uma rede neural com três camadas.

Os neurônios na camada de entrada servem para distribuir as informações contidas em um vetor de características para todos os neurônios da camada escondida. A ativação de saída associada à camada escondida  $i$  é dada pela soma ponderada de todas as entradas, ou seja:

$$net_{pi} = \sum_k w_{ki} \cdot i_{pk} \quad (4.8)$$

em que  $net_{pi}$  é o valor de ativação do padrão  $p$  no  $i$ -ésimo neurônio da camada escondida,  $i_{pk}$  é o  $k$ -ésimo neurônio do padrão  $p$  na camada de entrada e  $w_{ki}$  é o peso sináptico associado à conexão entre o  $k$ -ésimo neurônio de entrada e o  $i$ -ésimo neurônio escondido. O resultado na saída do neurônio é dado por uma função de ativação não-linear. Existem vários tipos de funções de ativação, dentre elas a mais empregada é a sigmóide:

$$\varphi(net_{pi}) = \frac{1}{1 + e^{-K net_{pi}}} = O_{pi} \quad (4.9)$$

em que  $K$  é uma constante positiva definida como o parâmetro de inclinação da função. Variando essa constante é possível obter sigmóides com diferentes inclinações. Usualmente, é adotado o valor  $K = 1$ , o qual fornece um valor de saída entre 0 e 1 [38].

Uma transferência de informação similar ocorre entre a camada escondida e a camada de saída. Sendo assim, cada neurônio da camada de saída tem seu valor associado à soma ponderada de todas as saídas dos neurônios da camada escondida:

$$net_{pj} = \sum_i w_{ij} \cdot O_{pi} \quad (4.10)$$

em que  $net_{pj}$  é o valor de ativação do padrão  $p$  no  $j$ -ésimo neurônio da camada de saída,  $O_{pi}$  é o  $i$ -ésimo neurônio do padrão  $p$  na camada escondida e  $w_{ij}$  é o peso sináptico associado à conexão entre o  $i$ -ésimo neurônio escondido e o  $j$ -ésimo neurônio de saída. O resultado na saída dos neurônios na última camada é dado por  $O_{pj} = \varphi(net_{pj})$ , e para todos os neurônios da camada de saída deve existir apenas um com saída não nula.

Todo o conhecimento necessário para classificar os padrões de entrada se concentra nos pesos das conexões entre os neurônios da rede. Inicialmente, os pesos mais adequa-



dos para um determinado problema são desconhecidos e precisam ser determinados. Caso contrário, a rede não desenvolve a habilidade para realizar o reconhecimento. O processo de encontrar os pesos adequados dado um problema, é chamado de treinamento. Treinar a rede consiste no ajuste dos pesos sinápticos de forma que todas as saídas produzidas pela mesma representem um erro inferior a um determinado fator de vigilância. No algoritmo de treinamento com retropropagação do erro, o erro médio quadrático entre a saída atual e a saída desejada para todos os padrões a serem treinados, é adotado como a medida de erro a ser calculada. Assim, o erro é dado por:

$$E_p = \frac{1}{2} \sum_j (t_{pj} - O_{pj})^2 \quad (4.11)$$

em que  $E_p$  é a função erro médio quadrático do padrão  $p$ ,  $t_{pj}$  representa a saída desejada do padrão  $p$  no neurônio  $j$  e  $O_{pj}$  representa a saída atual no neurônio  $j$ . A capacidade de uma rede neural aprender uma dada função está diretamente relacionada com a quantidade de neurônios nas camadas escondidas da rede.

O treinamento da rede é do tipo supervisionado e, portanto, requer que cada vetor de entrada venha junto com um vetor de saídas desejadas. A rede neural é treinada com um certo número de pares de treinamento (entrada, saída desejada). Ao conjunto de pares de treinamento dá-se o nome conjunto de treinamento.

Uma vez que a rede foi treinada, os pesos sinápticos são mantidos, e um conjunto de teste é aplicado na entrada da rede, para medir a eficiência da classificação. Os dados presentes no conjunto de teste para cada classe, contém apenas os valores de entrada e não fazem parte do conjunto de treinamento.

Os passos do algoritmo de retropropagação do erro para o treinamento da rede neural multicamadas são os seguintes:

1. Inicializa os pesos sinápticos e limiares. Estes devem ser inicializados com pequenos valores determinados aleatoriamente. A razão da estreita faixa é reduzir a probabilidade dos neurônios saturarem, ou seja, encontrarem uma solução estável que não provê uma saída correta.
2. Seleciona o par do conjunto de treinamento, aplicando os valores de entrada e de saída desejada.

3. Calcula a saída atual
4. Calcula o erro entre o valor de saída obtido e o valor desejado
5. Ajusta os pesos da rede visando minimizar o erro
6. Repete os passos de 2 a 5 para outros pares do conjunto de treinamento até que o erro global (obtido para todo o conjunto) seja suficientemente baixo.

Os passos 2 e 3 são conhecidos como passos diretos. Um vetor de entrada é aplicado e a saída resultante é computada. As computações são realizadas camada por camada.

Os passos 4 e 5 são conhecidos como passos para trás. No passo 4, o vetor de saída calculado é subtraído do vetor com as saídas desejadas do conjunto de treinamento, produzindo um erro. Esse erro é então utilizado pelo passo 5 para ajustar os pesos da rede. Esse ajuste é feito calculando-se os gradientes locais  $\delta$  da rede para a camada de saída,

$$\delta_{pj} = KO_{pj} (1 - O_{pj}) (t_{pj} - O_{pj}), \quad (4.12)$$

e para camada escondida,

$$\delta_{pj} = KO_{pj} (1 - O_{pj}) \sum_k \delta_{pk} w_{jk}. \quad (4.13)$$

Então, o ajuste dos pesos sinápticos da rede é feito de acordo com a equação

$$w_{ij} (n + 1) = w_{ij} (n) + \eta \delta_{pj} O_{pj} + \alpha [w_{ij} (n) - w_{ij} (n - 1)] \quad (4.14)$$

em que  $\eta$  é o parâmetro de aprendizagem e  $\alpha$  o momento.

O parâmetro de aprendizagem determina a variação do ajuste dos pesos sinápticos da rede. Para um pequeno valor de  $\eta$  o ajuste é lento e a rede demora para convergir. Por outro lado, aumentando o valor de  $\eta$  demasiadamente, pode provocar instabilidade na rede, uma vez que o ajuste é feito bruscamente. Dessa forma, o momento  $\alpha$  foi adicionado ao algoritmo de retropropagação do erro padrão, visando aumentar a convergência, sem no entanto, tornar a saída da rede instável. Usualmente seu valor é positivo [37].

### 4.3.1 Rede neural multicamadas com agrupamento (RNMCA)

O modelo da rede neural usada nesta dissertação, chamada de rede neural multicamadas com agrupamento (RNMCA), é um caso particular da rede descrita anteriormente.

Normalmente, uma rede neural multicamadas apresenta-se totalmente conectada, ou seja, todo neurônio é conectado a todos os neurônios das camadas antecedentes e subsequentes. Em algumas situações, este modelo de rede apresenta certas deficiências. Se o número de pesos for mantido pequeno, corre-se o risco da rede não aprender o conjunto de treinamento. Por outro lado, se o número de pesos na rede for grande, o excessivo número de parâmetros causa uma sobrecarga, degradando o seu desempenho [26, 5].

Uma alternativa para a redução das conexões entre os neurônios é usar uma rede conectada localmente. Lee et al. [5] propôs um modelo para esta rede baseada no tipo de informações extraídas com a transformada wavelet de duas dimensões. Cada uma das subimagens obtidas no processo de decomposição, forma um grupo, que é apresentado a rede neural independentemente um do outro. No intuito de tornar esta definição clara, a Figura 4.5 apresenta a arquitetura desta rede para um nível de resolução. Como a imagem do caracter encontra-se normalizada na dimensão  $16 \times 16$ , cada subimagem  $\{A_1, D_1^1, D_1^2, D_1^3\}$  contém  $8 \times 8$  pixels. Dessa forma, a camada de entrada consiste de quatro grupos, cada um deles com 64 neurônios. Observa-se que todos os neurônios de um grupo são totalmente conectados ao grupo correspondente na camada escondida, e não há conexões entre grupos distintos. Cada grupo da camada escondida, por sua vez, são totalmente conectados a todos os neurônios da camada de saída. A camada de saída é composta por dez unidades, uma por classe de numeral. Quando o numeral pertence a classe  $i$ , a saída desejada na unidade  $i$  é 1, e 0 para todas as outras unidades.

Todos os pesos da rede são iniciados aleatoriamente com pequenos valores, sendo ajustados através do algoritmo de retropropagação do erro descrito anteriormente.

As conexões entre um grupo na camada de entrada e o grupo correspondente na camada escondida formam um subrede. No exemplo da Figura 4.5, existem quatro

subredes. Uma vantagem apresentada pela RNMCA é a independência entre as subredes, pois se uma delas estiver confusa com os dados de entrada, as outras podem não estar, uma vez que estas iniciaram com valores diferentes e aprendem com diferentes características [26, 5].

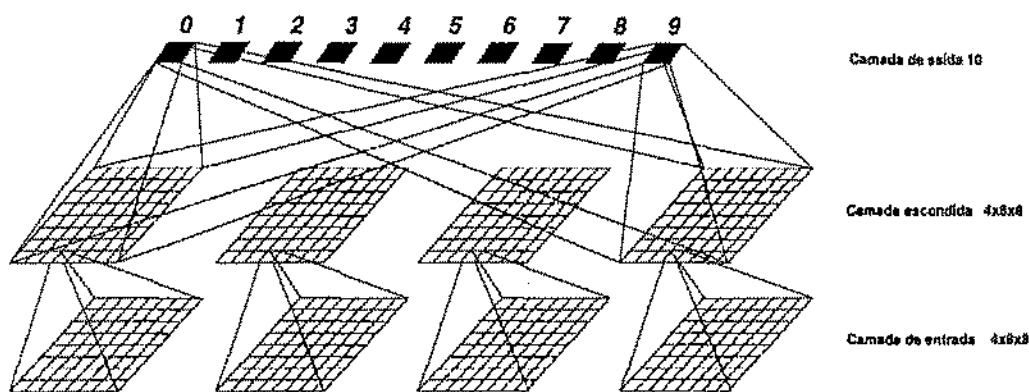


Figura 4.5: Arquitetura da rede neural multicamadas com agrupamento (RNMCA) para um nível de resolução.

A arquitetura desta rede pode sofrer modificações na camada de entrada e na camada escondida, pois estas são dependentes da quantidade de subimagens fornecidas após a extração de características. Porém, o princípio de funcionamento é o mesmo, ou seja, para cada subimagem é atribuído um grupo, e as conexões da rede são mantidas apenas entre os grupos equivalentes entre a camada de entrada e a camada escondida.

## 4.4 Conclusão

Neste capítulo, foi descrito o sistema de reconhecimento de numerais manuscritos implementado nesta dissertação. Este é composto por três estágios: normalização em escala, extração de características e classificação via redes neurais multicamadas com agrupamento treinadas com o algoritmo de retropropagação do erro. No capítulo seguinte, os testes efetuados e os resultados obtidos serão apresentados.

## Capítulo 5

# Resultados Experimentais

Neste capítulo, os sistemas de reconhecimento de numerais manuscritos implementados são avaliados através dos resultados obtidos experimentalmente. Esta avaliação é baseada na classificação de caracteres pertencentes entre as dez possíveis classes de numerais, usando parâmetros que medem a eficiência e precisão do sistema de reconhecimento.

As técnicas de pré-processamento e o algoritmo para o cálculo dos coeficientes wavelets foram implementados em linguagem C ANSI e a rede neural foi desenvolvida no ambiente de simulação desenvolvido pela Universidade de Stuttgart [39].

A seguir são apresentados a base de dados, os parâmetros de avaliação de desempenho e os testes realizados.

### 5.1 Base de dados

Os caracteres manuscritos usados para a validação dos algoritmos implementados fazem parte da base de dados numérica desenvolvida no CENPARMI (*Center for Pattern Recognition and Machine Intelligence, Concordia University, Canada*). Esta contém 17.771 numerais isolados, originalmente coletados dos envelopes de cartas “mortas” pelo *U.S. Postal Service* em diferentes localizações dos Estados Unidos e digitalizados em imagens binárias com uma resolução de aproximadamente 166 PPI (*pixels per inch*) [20]. Assumindo que o código postal norte-americano seja composto por cinco dígitos,

estima-se que os numerais foram escritos por no mínimo 3400 indivíduos [2].

As amostras que compõem a base de dados encontram-se irregularmente distribuídas entre as dez classes de numerais, como pode ser visto na Tabela 5.1. Para evitar que os algoritmos fossem treinados ou testados com uma quantidade distinta de amostras, foram selecionados os primeiros 600 numerais de cada classe, totalizando 6000 numerais, dos quais 4000 foram usados para formar o conjunto de treinamento e 2000 o conjunto de teste.

Tabela 5.1: Distribuição dos numerais por classe na base de dados do CENPARMI.

Classe	Amostras
0	2792
1	3697
2	1701
3	2213
4	1509
5	817
6	1372
7	1741
8	1273
9	716

As Figuras 5.1 e 5.2 mostram alguns exemplos de numerais representativos e confusos da base de dados que foram utilizados neste trabalho, respectivamente. Observa-se que diferentes tipos de estilos, tamanhos e largura de traçado podem ser identificados nos caracteres apresentados.

## 5.2 Parâmetros para avaliação de desempenho

O comportamento de um sistema de reconhecimento de caracteres pode ser avaliado através de uma matriz de classificação, denominada *matriz de confusão*. As colunas

0	1	2	3	4	5	6	7	8	9
0	1	2	3	4	5	6	7	8	9
0	1	2	3	4	5	6	7	8	9
0	1	2	3	4	5	6	7	8	9
0	1	2	3	4	5	6	7	8	9

Figura 5.1: Amostras de numerais representativos da base de dados do CENPARMI.

0	-	2	3	4	5	0	7	8	9
0	-	2	3	4	5	6	7	8	9
0	-	2	3	4	5	6	7	8	9

Figura 5.2: Amostras de numerais confusos da base de dados do CENPARMI.

dessa matriz representam os dados a serem classificados, e as linhas representam o número de classificações para o conjunto de dados analisado [40].

A distribuição dos dados na matriz de confusão fornece informações valiosas que podem ser usadas para medir o desempenho dos sistemas de reconhecimento. Para tanto, quatro parâmetros são comumente utilizados [2, 8]:

- Reconhecimento: a taxa de reconhecimento ou a taxa de sucesso é a percentagem de caracteres classificados corretamente;
- Erro: a taxa de erro ou taxa de substituição é a percentagem de caracteres classificados erroneamente pelo sistema;
- Rejeição: a taxa de rejeição é a percentagem de caracteres que não são identificados pelo sistema;
- Confiabilidade: a taxa de confiabilidade mede a percentagem de caracteres classificados corretamente dentre todos os caracteres classificados pelo sistema, sejam

eles corretos ou não. Esta pode ser obtida através da equação:

$$\text{Confiabilidade} = \frac{\text{Reconhecimento}}{\text{Reconhecimento} + \text{Erro}} \times 100\%. \quad (5.1)$$

A confiabilidade do sistema, também pode ser definida como sendo a taxa de caracteres reconhecidos corretamente após a exclusão dos caracteres rejeitados, ou seja,

$$\text{Confiabilidade} = \frac{\text{Reconhecimento}}{100 - \text{Rejeição}} \times 100\%. \quad (5.2)$$

As equações 5.1 e 5.2 fornecem os mesmos resultados, uma vez que a soma das taxas de reconhecimento, erro e rejeição devem ser iguais a 100%.

Um sistema de reconhecimento eficiente deve fornecer taxas de reconhecimento e confiabilidade elevadas. Uma alta taxa de confiabilidade pode ser obtida através da redução da taxa de erro. Por esta razão, muitos sistemas adotam a rejeição, pois é preferível descartar um caracter para o processamento manual, do que classificá-lo erroneamente.

### 5.2.1 Regra de Rejeição

Na maioria dos sistemas de reconhecimento, a unidade de saída  $i$  da rede neural que apresenta o maior valor numérico, identifica o caracter como pertencente a  $i$ -ésima classe. Neste trabalho, a regra de decisão utilizada é baseada no seguinte procedimento:

1. Identifique as duas unidades de saída da rede neural que apresentam os maiores valores numéricos;
2. Calcule a diferença entre estes valores;
3. Caso o resultado da diferença seja  $\geq 0,2$ , classifique o caracter como pertencente à classe indicada pela unidade de saída da rede que apresenta o maior valor numérico. Caso contrário, o caracter é rejeitado. O valor limiar 0,2 foi determinado empiricamente através de uma grande quantidade de testes, e pode ser alterado para outra base de dados ou método de reconhecimento.



## 5.3 Testes Efetuados e Análise de Desempenho

No intuito de melhorar o desempenho do sistema de reconhecimento proposto por Lee et al [5] uma grande quantidade de experimentos foram realizados no decorrer da pesquisa. O objetivo desta seção é apresentar os resultados de cada experimento e analisar os resultados obtidos. Em todos os sistemas implementados os numerais da bases de dados selecionados para formar o conjunto de treinamento e testes foram normalizados para a dimensão  $16 \times 16$  *pixels*.

### 5.3.1 Influência do número de níveis de resolução

A decomposição wavelet bi-dimensional, descrita no capítulo 3, fornece subimagens contendo características globais e direcionais do caracter, as quais são usadas para compor o vetor de características no sistema de reconhecimento apresentado neste trabalho. Uma vez que as imagens normalizadas dos numerais possuem  $16 \times 16$  *pixels*, estas podem ser decompostas, no máximo, em três níveis de resolução.

A arquitetura da rede neural multicamadas com agrupamento (RNMCA) utilizada durante o estágio de classificação é dependente da quantidade e tamanho das subimagens escolhidas para a representação do numeral. No intuito de determinar o melhor conjunto de características e definir a arquitetura da rede, foram testadas quatro configurações, de acordo com as subimagens utilizadas:

1. todas as subimagens do primeiro nível de resolução;
2. todas as subimagens do segundo nível de resolução;
3. todas as subimagens do terceiro nível de resolução;
4. todas as subimagens do primeiro e segundo nível de resolução.

A Tabela 5.2 apresenta os vetores de características. Para as três primeiras configurações, a RNMCA possui 256 neurônios na camada de entrada, 256 neurônios na camada escondida e 10 neurônios na camada de saída. Para a última configuração, a

Tabela 5.2: Configurações testadas.

Configuração	Subimagens	Dimensão do vetor de características
1	$\{A_1, D_1^1, D_1^2, D_1^3\}$	$4 \times 8 \times 8$ pixels
2	$\{A_2, D_2^1, D_2^2, D_2^3, D_1^1, D_1^2, D_1^3\}$	$4 \times 4 \times 4 + 3 \times 8 \times 8$ pixels
3	$\{A_3, D_3^1, D_3^2, D_3^3, D_2^1, D_2^2, D_2^3, D_1^1, D_1^2, D_1^3\}$	$4 \times 2 \times 2 + 3 \times 4 \times 4 + 3 \times 8 \times 8$ pixels
4	$\{A_2, D_2^1, D_2^2, D_2^3, A_1, D_1^1, D_1^2, D_1^3\}$	$4 \times 4 \times 4 + 4 \times 8 \times 8$ pixels

RNMCA possui 320 neurônios na camada de entrada, 320 neurônios na camada escondida e 10 neurônios na camada de saída. As conexões entre os neurônios da camada de entrada e da camada escondida, nas redes neurais implementadas, são modificadas dependendo das subimagens utilizadas.

Os parâmetros usados durante o treinamento das redes, pelo algoritmo de retropropagação do erro, foram os seguintes:

- taxa de aprendizagem: 0,01;
- erro usado no teste de convergência: 0,01;
- momento: 0,9.

As Tabelas 5.3 a 5.6 contém os dados de reconhecimento para cada uma das configurações. Nesse caso, foram utilizadas as wavelets de Haar durante o processo de extração de características.

Comparando os resultados das três primeiras configurações observa-se que à medida que o nível de resolução aumenta, as taxas de reconhecimento e confiabilidade diminuem e, as taxas de erro e rejeição aumentam. A teoria de wavelets afirma que o aumento de níveis de resolução fornece uma quantidade maior de detalhes sobre o sinal analisado. Dessa forma, seria esperado que o sistema tivesse melhores resultados para um nível maior de resolução, entretanto, neste caso isto não se verifica. As subimagens do segundo e principalmente do terceiro nível de resolução contém mais informações, porém, estas se encontram distribuídas em matrizes muito pequenas, como mostra a

Tabela 5.3: Resultados obtidos para a configuração 1.

Numeral	Reconhecimento (%)	Erro (%)	Rejeição (%)	Confiabilidade (%)
0	98,0	0,5	1,5	99,49
1	97,0	0,0	3,0	100,0
2	95,0	1,0	4,0	98,95
3	91,0	4,5	4,5	95,28
4	94,5	2,0	3,5	97,92
5	86,5	5,0	8,5	94,53
6	96,5	1,0	2,5	98,97
7	91,0	5,0	4,0	94,79
8	86,0	3,0	11,0	96,62
9	91,5	2,0	6,5	97,86
Média	92,7	2,4	4,9	97,44

Tabela 5.2 e a Figura 4.3, uma vez que a imagem normalizada do caracter possui uma dimensão reduzida de  $16 \times 16$  *pixels*. Sendo assim, a classificação via redes neurais a partir de vetores características com as subimagens das configurações 2 e 3, torna-se muito confusa, o que acaba prejudicando o desempenho do sistema de reconhecimento. A quarta configuração obteve o melhor desempenho mas apresenta desvantagens, como a grande quantidade de elementos usados para representar o caracter. Esse fato aumenta o tempo necessário para treinar a rede neural. Dessa forma, dentre as quatro configurações decidiu-se manter a primeira para os testes relatados na próxima seção.

As configurações 1 e 4 testadas com as wavelets de Haar, são resultados do trabalho original de Lee et al. [5]. A Tabela 5.7 mostra as taxas médias obtidas no trabalho original e nos testes realizados nesta dissertação. Observa-se que a adição do critério de rejeição no sistema implementado permitiu um aumento do desempenho, mesmo na configuração 4, pois embora o sistema original tenha reconhecido mais caracteres, o sistema implementado errou menos e apresenta uma melhor confiabilidade.

Tabela 5.4: Resultados obtidos para a configuração 2.

Numeral	Reconhecimento (%)	Erro (%)	Rejeição (%)	Confiabilidade (%)
0	98,0	0,5	1,5	99,49
1	97,0	0,0	3,0	100,0
2	92,5	2,5	5,0	97,36
3	87,5	4,0	8,5	95,62
4	94,5	2,0	3,5	97,92
5	84,0	4,5	11,5	94,91
6	94,0	1,5	4,5	98,42
7	89,5	4,0	6,5	95,72
8	85,5	4,0	10,5	95,53
9	89,5	3,0	7,5	96,75
Média	91,2	2,6	6,2	97,17

### 5.3.2 Escolha da melhor wavelet

A escolha da família wavelet é dependente da aplicação. Objetivando determinar a melhor wavelet capaz de aumentar o desempenho da configuração 1, foram selecionadas três famílias wavelets, duas ortogonais e uma biortogonal, com bases possuindo filtros de diversos comprimentos:

- Daubechies (Daub): 2,4,6,8,10,12 e 20;
- Coiflets (Coif): 2,4 e 6;
- Conhen-Daubechies-Feauveau (CDF): 2/2, 2/4,3/3,3/7.

Como descrito no capítulo 3, os filtros ortogonais passa-baixas  $h(n)$  e passa-altas  $g(n)$  possuem o mesmo comprimento e os filtros biortogonais podem possuir comprimentos diferentes. A seleção dessas bases foi determinada pela facilidade de acesso aos coeficientes dos filtros. Estes podem ser encontrados no livro de Daubechies [28].

Tabela 5.5: Resultados obtidos para a configuração 3.

Numeral	Reconhecimento (%)	Erro (%)	Rejeição (%)	Confiabilidade (%)
0	93,5	2,5	4,0	97,39
1	99,0	0,0	1,0	100,0
2	92,5	2,0	5,5	97,88
3	85,5	3,5	11,0	96,06
4	90,5	3,0	6,5	96,79
5	78,5	6,5	15,0	92,35
6	91,5	2,0	6,5	97,86
7	91,5	3,0	5,5	96,82
8	74,5	8,5	17,0	89,75
9	87,0	4,0	9,0	95,60
Média	88,4	3,5	8,1	96,0

O comprimento do filtro tem uma influência importante na determinação do melhor desempenho. Filtros de pequeno comprimento são indicados para extração de detalhes e informações de borda e filtros com comprimento grande para análise de textura. Entretanto, nos gráficos da Figura 5.3 observa-se que quando o comprimento do filtro diminui o desempenho do sistema também diminui. Isto ocorre porque nesse caso os coeficientes dos filtros são representados por matrizes esparsas. Segundo Yuceer [35] quanto mais zeros são apresentados a rede pior o seu desempenho. Dessa forma, conclui-se que existe um compromisso entre a extração de características e o classificador: a wavelet escolhida não pode possuir filtros de comprimento pequeno para não prejudicar a rede, nem tampouco pode possuir um comprimento demasiadamente grande, senão nenhuma característica é extraída, ou seja, há pouca informação sobre os componentes de altas frequências nas direções vertical, horizontal e diagonal das subimagens de detalhes.

Nos resultados obtidos com a família de Daubechies observa-se claramente esta relação, e deduz-se que o comprimento de filtro ideal está entre 6 e 10 coeficientes.

Tabela 5.6: Resultados obtidos para a configuração 4.

Numeral	Reconhecimento (%)	Erro (%)	Rejeição (%)	Confiabilidade (%)
0	99,0	0,0	1,0	100,0
1	97,0	0,0	3,0	100,0
2	97,0	0,5	2,5	99,48
3	94,5	4,5	1,0	95,45
4	94,0	2,0	4,0	97,91
5	89,0	4,0	7,0	95,69
6	95,5	2,0	2,5	98,93
7	91,0	4,0	5,0	95,78
8	90,0	3,5	6,5	96,25
9	93,0	2,5	4,5	97,38
Média	94,0	2,3	3,65	97,68

A Tabela 5.8 apresenta os melhores resultados alcançados com as wavelets testadas. Dentre as famílias as CDF se destacaram. Estas apresentam uma série de vantagens, como o fato de serem extremamente locais e simétricas [41]. Por seus desempenhos, a CDF 3/7 foi selecionada para ser utilizada no sistema proposto. As Tabelas 5.9 a 5.12 mostram as matrizes de confusão e as tabelas de reconhecimento para cada uma das bases CDF. Analisando cada uma delas nota-se que o desempenho da base CDF 3/7 é melhor no geral, mas não para numerais isolados. Por exemplo, o numeral 5 teve um decréscimo na taxa de reconhecimento com relação a outras bases CDFs analisadas. Provavelmente, este fato ocorre devido as amostras deste numeral na base de dados utilizada serem muito confusas e, a medida que os coeficientes do filtro passa-altas aumentam, as informações extraídas se parecem com a de outros numerais semelhantes.

A Tabela 5.13, mostra o tempo de processamento em cada etapa do sistema final obtido com a base CDF 3/7.

### 5.3.3 Análise Comparativa

Os desempenhos de alguns sistemas de reconhecimento de numerais manuscritos descritos na literatura, que usaram a base de dados do CENPARMI para a validação experimental, são mostrados na Tabela 5.14. É importante observar que esses sistemas não podem ser comparados simplesmente pelo resultados obtidos, desde que os numerais selecionados para formar o conjunto de treinamento e teste, em cada caso, podem ter sido bastante diferentes.

Normalmente, os extratores de características além de representar os dados buscam diminuir o conjunto do vetor de características. No sistema proposto, o vetor possui a mesma quantidade de elementos da imagem normalizada. Assim, testou-se um sistema em que as imagens normalizadas entravam diretamente na rede neural multicamadas. Os resultados encontram-se na Tabela 5.15 e são bem inferiores aos obtidos com os sistemas que utilizaram wavelets.

## 5.4 Conclusão

Neste capítulo foram apresentados os resultados dos testes realizados e a comparação entre o sistema proposto e outros sistemas de reconhecimento. Conclui-se que os desempenhos obtidos foram bastantes satisfatórios, uma vez que o sistema usando a base CDF 3/7 apresentou uma boa confiabilidade e errou menos que o sistema original proposto por Lee et al. [5].

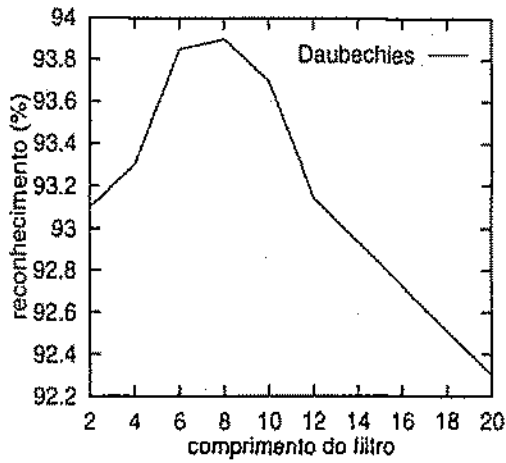
Tabela 5.7: Comparação entre os resultados obtidos por Lee et al. e o sistema implementado.

Taxas Médias	Configuração 1		Configuração 4	
	Lee et al.	Implementado	Lee et al.	Implementado
Reconhecimento (%)	92,3	92,7	96,8	94,0
Erro (%)	7,7	2,4	3,2	2,3
Rejeição (%)	0,0	4,9	0,0	3,65
Confiabilidade (%)	92,3	97,44	96,8	97,68

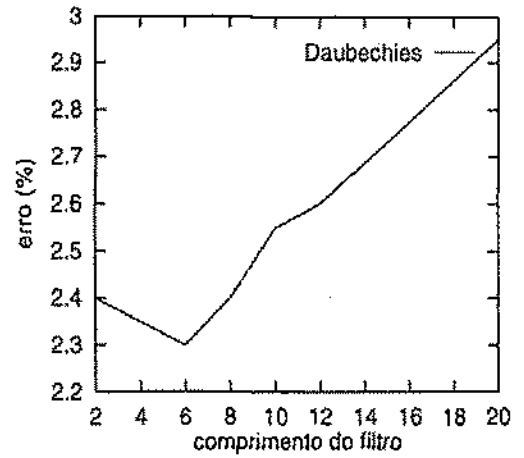
Tabela 5.8: Taxas médias obtidas com as famílias wavelets Daubechies (Daub), Coiflets (Coif) e Cohen-Daubechies-Feauveau (CDF) testadas.

Família	Reconhecimento (%)	Erro (%)	Rejeição (%)	Confiabilidade (%)
Daub6	93,85	2,3	3,85	97,6
Daub8	93,9	2,4	3,7	97,5
Daub10	93,7	2,55	3,75	97,35
Coif2	93,25	2,65	4,1	97,2
Coif4	93,45	2,85	3,7	97
Coif6	93,65	2,3	4,05	97,57
CDF 2/2	93,3	2,6	4,4	97,25
CDF 2/4	93,4	2,5	4,1	97,35
CDF 3/3	94,45	2,3	3,25	97,59
CDF 3/7	94,7	1,8	3,5	98,11

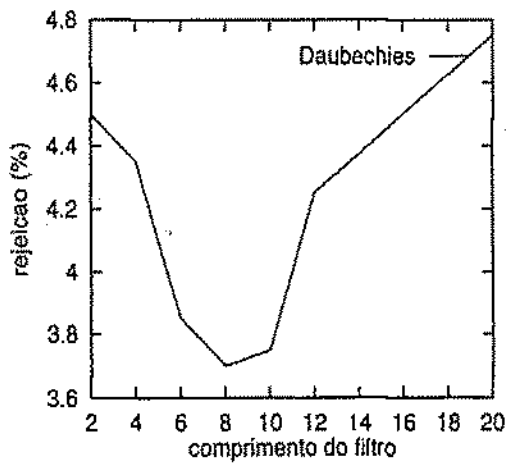




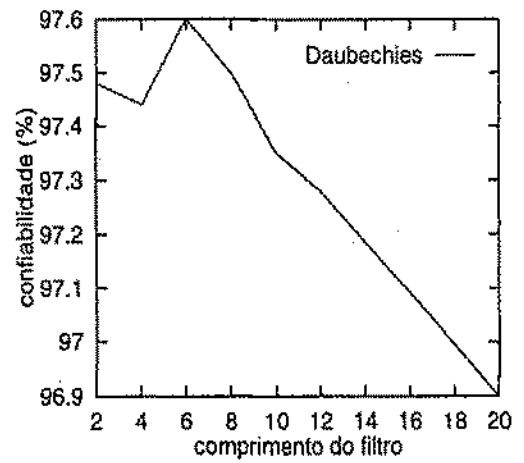
(a) Taxas de reconhecimento



(b) Taxas de erro



(c) Taxas de rejeição



(d) Taxas de confiabilidade

Figura 5.3: Influência do comprimento do filtro no sistema de reconhecimento de numerais manuscritos usando as wavelets de Daubechies.

Tabela 5.9: Matriz de confusão da família CDF 2/2.

Classe	0	1	2	3	4	5	6	7	8	9	Rejeição
0	197					1					2
1		194		1							5
2			190			2					8
3			4	186		1			2		7
4		1			187				1	5	6
5			4	2	1	177			1		15
6			2			1	194				3
7			1		2			184		5	8
8		1	1	1		3			177	1	16
9				2	1	1		2	2	180	12

Tabela 5.10: Matriz de confusão da família CDF 2/4.

Classe	0	1	2	3	4	5	6	7	8	9	Rejeição
0	198										2
1		194									6
2			192	1		1					6
3			6	187		1			3		3
4		1			186					4	9
5			3	2	1	171	1	1	1	1	19
6			1			1	193				5
7			1		2			187		4	6
8		1	1	1		1			177	3	16
9		1		1	1	1		1	2	183	10

Tabela 5.11: Matriz de confusão da família CDF 3/3.

Classe	0	1	2	3	4	5	6	7	8	9	Rejeição
0	198										2
1		194									6
2			195			2					3
3			3	185					5		7
4		1			186					4	9
5			3	3		179	4				11
6			1				196				3
7					2			183		8	7
8	1		1			1			183	2	12
9					2	1		2		190	5

Tabela 5.12: Matriz de confusão da família CDF 3/7.

Classe	0	1	2	3	4	5	6	7	8	9	Rejeição
0	200										0
1		194									6
2			195								5
3			4	190					2		4
4			1		186					3	10
5			2	2		176	2			1	17
6			2			1	190				7
7					2			189		5	4
8		1				1			184	2	12
9					2	1		2		190	5

Tabela 5.13: Tempo de processamento para cada etapa do sistema usando as wavelets CDF 3/7.

Processo	Tempo(ms/caracter)
Normalização	42,7
Extração de Características	70,82
Classificação	3,59

Tabela 5.14: Resultados obtidos por diversos sistemas que utilizaram a base de dados do CENPARMI.

Método	Reconhecimento (%)	Erro (%)	Rejeição (%)	Confiabilidade (%)
E1 - Suen et al. [2]	93,05	0,0	6,95	100,0
E2 - Suen et al. [2]	86,05	2,25	11,7	97,45
E3 - Suen et al. [2]	93,1	2,95	3,95	96,98
E4 - Suen et al. [2]	93,9	1,6	4,5	98,32
Veloso [8]	84,72	4,53	10,75	95,47
Cho [20]	96,05	3,95	0,0	96,05
M1- Lee et al. [5]	92,3	7,7	0,0	92,3
M2- Lee et al. [5]	96,8	3,2	0,0	96,8
Método Proposto	94,7	1,8	3,5	98,13

Tabela 5.15: Resultados obtidos sem a etapa de extração de características.

Numeral	Reconhecimento (%)	Erro (%)	Rejeição (%)	Confiabilidade (%)
0	85,0	1,5	13,0	98,26
1	89,5	2,5	8,0	97,28
2	78,0	2,0	20,0	97,5
3	61,5	9,5	29,0	86,61
4	78,0	2,0	20,0	97,5
5	58,5	7,5	34,0	88,63
6	71,5	4,0	24,5	94,7
7	71,0	2,5	26,5	96,59
8	58,0	6,0	36,0	90,62
9	62,0	4,0	34,0	93,93
Média	71,3	4,16	24,5	94,16

## Capítulo 6

### Conclusão

O reconhecimento automático de caracteres é um problema de difícil solução devido à quantidade de variações existentes na escrita manual. No intuito de desenvolver sistemas capazes de interpretar o significado de letras e numerais com um nível de desempenho semelhante ao dos humanos, muitos métodos têm sido apresentados na literatura. Nesta dissertação, é proposto um sistema de reconhecimento de caracteres numéricos manuscritos, baseado no trabalho de Lee et al. [5] que utiliza a transformada wavelet bi-dimensional para extrair características globais e direcionais dos numerais manuscritos. O sistema é composto por três estágios: pré-processamento, extração de características e classificação. O pré-processamento padroniza as imagens dos caracteres em uma mesma dimensão através da normalização em escala. A extração de características usa os coeficientes da transformada wavelet obtidos pela família biortogonal Cohen-Daubechies-Feauveau (CDF) 3/7 para formar o vetor de características. Por fim, o classificador toma a decisão final de a qual classe pertence o caracter, através de uma rede neural multicamadas com agrupamento treinada com o algoritmo de retropropagação do erro.

A transformada wavelet provê uma representação dos numerais em diferentes bandas de frequência, representando o formato geral do caracter, bem como seus detalhes ao longo das direções horizontal, vertical e diagonal. Por esta razão, pode ser usada no reconhecimento de numerais manuscritos. Sua eficiência é comprovada através dos resultados obtidos, tendo demonstrado robustez às variações de forma apresentados

pelos caracteres da base de dados do CENPARMI [2], fornecendo uma taxa média de reconhecimento de 94.7%.

## 6.1 Contribuições

Podem ser citadas como contribuições originais deste trabalho:

- Aumento das taxas de reconhecimento e confiabilidade do sistema original proposto por Lee et al., através da substituição da wavelet de Haar pelas wavelets biortogonais CDF 3/7 e inserção de uma regra de rejeição;
- Avaliação comparativa entre bases wavelets para determinar o efeito do comprimento do filtro e a indentificação da melhor família wavelet para o problema de reconhecimento de caracteres manuscritos;
- Iniciar estudos sobre o uso da transformada wavelet para o reconhecimento de numerais manuscritos. Estes estudos são pioneiros no âmbito do grupo de Análise e Processamento de Sinais e Imagens do DEE/UFPb e, possivelmente, no país.

## 6.2 Perspectivas de Trabalhos Futuros

- Investigação de técnicas, como a transformada de Karhunen-Loeve e quantização vetorial, para redução da quantidade de elementos do vetor de características sem afetar o desempenho já alcançado pelo sistema;
- Estudo aprofundado da topologia de redes neurais, de maneira a melhorar ou desenvolver uma nova arquitetura de rede para o sistema proposto, capaz de aumentar o seu desempenho;
- Investigação do uso das wavelets *packets*, no sentido de determinar as melhores bases para o problema do reconhecimento automático de numerais manuscritos;
- Análise aprofundada dos caracteres que compõem a base de dados do CENPARMI, no intuito de determinar o melhor conjunto de treinamento, capaz de fornecer

o maior número possível de numerais representativos e conseqüentemente melhorar a eficiência do sistema;

- Avaliação de desempenho do sistema proposto usando caracteres numéricos pertencentes a outras base de dados, que contenham dígitos coletados em outras regiões (de preferência no Brasil), uma vez que os numerais podem apresentar diferenças na forma da escrita de uma região para outra. Como por exemplo, o numeral 7 nos Estados Unidos não costuma possuir o traço horizontal dividindo-o, como no Brasil.



# Bibliografia

- [1] S. N. Srihari. High-performance reading machines. *Proceedings of IEEE*, 7(7):1120–1132, July 1992.
- [2] C. Y. Suen, C. Nadal, R. Legault, T. A. Mai and L. Lam. Computer recognition of unconstrained handwritten numerals. *Proceedings of the IEEE*, 80(7):1162–1180, 1992.
- [3] C. Y. Suen, M. Berthod and S. Mori. Automatic recognition of handprinted characters - the state of the art. *Proceedings of the IEEE*, 68(4):469–487, 1980.
- [4] G. S. Kapogiannopoulos and M. Papadakis. Character recognition using a biorthogonal discrete wavelet transform. *Wavelet Applications in Signal and Image Processing IV, SPIE*, 2825:384–393, 1996.
- [5] S.-W. Lee, C.-H. Kim and Y. Y. Tang. Multiresolution recognition of unconstrained handwritten numerals with wavelet transform and multilayer cluster neural network. *Pattern Recognition*, 29(12):1953–1961, 1996.
- [6] N. R. Gomes and L. L. Ling. Um Algoritmo Sequencial Utilizando Redes Neurais Para o Reconhecimento de numerais manuscritos. In *XIV- Simpósio brasileiro de telecomunicações*, 1996.
- [7] O. D. Trier, A. K. Jain and T. Taxt. Feature extraction methods for character recognition - a survey. *Pattern Recognition*, 29(4):641–662, 1996.

- [8] L. R. Veloso. Reconhecimento de caracteres numéricos manuscritos. *Dissertação de Mestrado*, Universidade Federal da Paraíba - Centro de Ciências e Tecnologia - Departamento de Engenharia Elétrica, 1998.
- [9] G.-E. Wang and J.-F. Wang. A new hierarchical approach for recognition of unconstrained handwritten numerals. *IEEE- Transactions on Consumer Electronics*, 40(3):428-436, 1994.
- [10] R. C. Gonzalez and R. E. Woods. *Digital Image Processing*. Addison-Wesley, 1992.
- [11] L. O'Gorman and R. Kasturi. *Document Image Analysis*. IEEE Computer Society, 1997.
- [12] H. M. Gomes. Investigação de técnicas automáticas para o reconhecimento off-line de assinaturas. *Dissertação de Mestrado*, Universidade Federal de Pernambuco - Recife, 1994.
- [13] J. Cai and Z.-Q. Liu. Integration of structural and statistical information of unconstrained handwritten numeral recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 21(3):263-270, March 1999.
- [14] A. Sinha. An improved recognition module for the identification of handwritten digits. *Dissertação de Mestrado*, Department of Electrical and Computer Science, MIT, 1999.
- [15] M. Shridhar and A. Badreldin. High accuracy character recognition algorithm using fourier and topological descriptors. *Pattern Recognition*, 17(5):515-524, 1984.
- [16] A. Graps. *An Introduction to Wavelets*. Disponível na URL <http://www.amara.com/IEEEwave/Iw-fbi.html>.
- [17] A. D. Poularikas. *The Transforms and Applications Handbook*. CRC Press and IEEE Press.