

Universidade Estadual de Campinas
Faculdade de Engenharia Elétrica
Departamento de Comunicações

Algoritmo Seqüencial para Reconhecimento de Numerais
Manuscritos Desconectados Utilizando Redes Neurais

Autor

Natanael Rodrigues Gomes

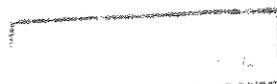
Tese apresentada à Faculdade de Engenharia
Elétrica, Departamento de Comunicações,
como parte dos requisitos para a obtenção do
título de Mestre em Engenharia Elétrica.
Orientador: Prof. Lee Luan Ling

Este exemplar corresponde à versão final da tese
defendida por Natanael Rodrigues Gomes
em defesa pública em 22 / 02 / 1996.
Orientador Lee Luan Ling

CAMPINAS

1996

7604/96



UNIDADE	80
N.º CHAMADA:	T/UNICAMP
	G585a
V. Ex.	
IMP. Nº	2/27602
PRG.	667/96
C	<input type="checkbox"/>
D	<input checked="" type="checkbox"/>
PREÇO	R\$ 11,00
DATA	03/05/96
N.º CPD	011.00087591-9

FICHA CATALOGRÁFICA ELABORADA PELA
BIBLIOTECA DA ÁREA DE ENGENHARIA - BAE - UNICAMP

G585a Gomes, Natanael Rodrigues.
 Algoritmo seqüencial para reconhecimento de numerais
 manuscritos desconectados utilizando redes neurais /
 Natanael Rodrigues Gomes.--Campinas, SP: [s.n.], 1996.

 Orientador: Lee Luan Ling.
 Dissertação (mestrado) - Universidade Estadual de
 Campinas, Faculdade de Engenharia Elétrica.

 1. Reconhecimento de padrões. 2. Redes neurais
 (Computação). 3. * Numerais manuscritos. I. Lee, Luan
 Ling. II. Universidade Estadual de Campinas. Faculdade
 de Engenharia Elétrica. III. Título.

*Dedico este trabalho aos meus pais,
Luiz e Palmira, pelo inestimável apoio
e compreensão em todos os momentos.*

AGRADECIMENTOS

Um especial agradecimento à minha família, pelo carinho e incentivo que sempre me foi dedicado.

À minha esposa Ana Paula, pela valiosa ajuda e apoio em todo o desenvolvimento deste trabalho.

Aos colegas do Núcleo de Estudos e Pesquisas Aeroespaciais, na Universidade Federal de Santa Maria, e ao Professor Nelson Schuch, pela importante ajuda em meu desenvolvimento pessoal e profissional.

Ao meu orientador, Professor Lee, pelas indispensáveis observações e explicações no decorrer deste trabalho.

A todas as pessoas que, de uma maneira ou de outra, me ajudaram a desenvolver este trabalho.

SUMÁRIO

LISTA DE FIGURAS	VIII
LISTA DE TABELAS.....	XI
RESUMO	XII
ABSTRACT	XIII
CAPÍTULO 1 - INTRODUÇÃO.....	1
1.1 - Reconhecimento de Caracteres Manuscritos.....	1
1.2 - Histórico do Reconhecimento de Caracteres.....	1
1.3 - Aplicações dos Sistemas de Reconhecimento de Caracteres.....	2
1.4 - Objetivos da Dissertação.....	3
1.5 - Organização da Dissertação.....	4
CAPÍTULO 2 - ALGUNS ASPECTOS DO RECONHECIMENTO DE CARACTERES	7
2.1 - Introdução	7
2.2 - Estrutura dos Sistemas de Reconhecimento de Caracteres	7
2.3 - Principais Problemas do Reconhecimento de Caracteres Manuscritos.....	9
2.4 - Métodos mais Empregados no Reconhecimento de Caracteres	10
2.4.1 - Transformações Globais.....	11
2.4.2 - Distribuição de Pontos	12
2.4.3 - Características Topológicas e Geométricas.....	13
2.5 - Algumas Considerações Finais.....	15
CAPÍTULO 3 - REDES NEURAIIS	16
3.1 - Introdução	16
3.2 - Conceitos Básicos Sobre as Redes Neurais.....	17

3.2.1 - Neurônio Artificial.....	18
3.2.2 - Funções de Ativação	19
3.2.3 - Aprendizagem.....	19
3.2.3.1 - Treinamento Supervisionado	20
3.2.3.2 - Treinamento não Supervisionado.....	20
3.2.3.3 - Pesos Fixados.....	20
3.3 - Rede de Hopfield Discreta.....	21
3.3.1 - Arquitetura da Rede de Hopfield Discreta	21
3.3.2 - Operação da Rede de Hopfield Discreta	22
3.3.3 - Treinamento da Rede de Hopfield Discreta.....	24
3.3.4 - Capacidade de Armazenamento da Rede de Hopfield Discreta.....	24
3.3.5 - Padrões Correlatos - Método da Projeção	26
3.3.6 - Função de Energia da Rede de Hopfield Discreta.....	27
3.4 - Sistemas LSSM Aplicados ao Modelo da Rede de Hopfield Discreta.....	28
3.4.1 - Sistemas LSSM.....	29
3.4.2 - Síntese dos Sistemas LSSM.....	31
3.5 - Algumas Considerações Finais.....	33
CAPÍTULO 4 - MORFOLOGIA MATEMÁTICA E APRESENTAÇÃO DO SISTEMA.....	35
4.1 - Introdução	35
4.2 - Morfologia Matemática.....	35
4.2.1 - Algumas Definições Básicas da Morfologia Matemática.....	36
4.2.2 - Transformações "Tudo ou Nada" por um Elemento Estruturante	39
4.3 - Descrição Geral do Sistema de Reconhecimento de Numerais Manuscritos.....	40
4.4 - Algumas Considerações Finais.....	43
CAPÍTULO 5 - AQUISIÇÃO E PRÉ-PROCESSAMENTO DA IMAGEM	44
5.1 - Introdução	44
5.2 - Aquisição da Imagem.....	46
5.3 - Pré - Processamento da Imagem.....	48
5.3.1 - Normalização em Escala.....	48

5.3.2 - Afinamento e Dilatação de Caracteres.....	49
5.3.2.1 - Afinamento.....	50
5.3.2.2 - Eliminação dos Ramos Parasitas do Esqueleto.....	52
5.3.2.3 - Dilatação.....	54
CAPÍTULO 6 - EXTRAÇÃO DE CARACTERÍSTICAS	57
6.1 - Introdução	57
6.2 - Cavidades da Imagem dos Numerais.....	58
6.3 - Seqüência de Cruzamento de um Numeral.....	61
6.4 - Interseções com o Eixo Principal e Eixo Secundário de um Numeral.....	62
6.5 - Distribuição Pictorial.....	63
CAPÍTULO 7 - CLASSIFICAÇÃO DE NUMERAIS MANUSCRITOS	65
7.1 - Introdução	65
7.2 - Classificação via Características Extraídas.....	65
7.2.1 - Classificação dos Numerais Observando a Existência de Cavidades Centrais.....	66
7.2.2 - Classificação dos Numerais que não Apresentam Cavidades Centrais.....	72
7.2.2.1 - Grupo de Numerais que Apresentam 3 Interseções com o Eixo Principal.....	72
7.2.2.2 - Grupo de Numerais que Apresentam Outro Número de Interseções com o Eixo Principal.....	75
7.3 - Classificação Via Rede Neural.....	76
7.3.1 - Rede de Hopfield Empregando o Método da Projeção.....	77
7.3.2 - Rede de Hopfield Empregando o Procedimento de Síntese para Sistemas LSSM	81
CAPÍTULO 8 - RESULTADOS EXPERIMENTAIS E DISCUSSÃO FINAL	83
8.1 - Introdução	83
8.2 - Resultados Obtidos	83

8.2.1 - Teste do Sistema Implementado com a Rede de Hopfield com Pesos Calculados pelo Método da Projeção.....	84
8.2.2 - Teste do Sistema Implementado com a Rede de Hopfield com Pesos Calculados pelo Processo de Síntese para Sistemas LSSM.....	86
8.2.3 - Análise dos Resultados Obtidos.....	87
8.3 - Conclusões Finais sobre o Método de Reconhecimento Empregado.....	88
8.4 - Propostas para Aperfeiçoamento do Sistema	89
REFERÊNCIAS BIBLIOGRÁFICAS	91

LISTA DE FIGURAS

Fig. 2.1. Diagrama de blocos de um típico sistema de reconhecimento padrões.....	9
Fig. 2.2. Exemplos de segmentos de reta e curvas a serem extraídos da forma do caracter.....	14
Fig. 2.3. Exemplo de características extraídas da topologia do caracter.....	14
Fig. 3.1. Rede neural com duas camadas de neurônios.....	17
Fig. 3.2. Neurônio artificial.....	18
Fig. 3.3. Esquema da rede de Hopfield discreta.....	21
Fig. 4.1. Translação do conjunto A	36
Fig. 4.2. Reflexão do conjunto C	37
Fig. 4.3. Complemento do conjunto A	37
Fig. 4.4. União dos conjuntos A e B	38
Fig. 4.5. Interseção dos conjuntos A e B	38
Fig. 4.6. Diferença entre os conjuntos A e B	39
Fig. 4.7. Exemplo de uma transformação “tudo ou nada”	40
Fig. 4.8. Diagrama funcional do sistema de reconhecimento de numerais manuscritos.....	41
Fig. 5.1. Amostra dos numerais a serem digitalizados.....	46
Fig. 5.2. Processo de enquadramento de um caracter.....	48
Fig. 5.3. Esqueleto da imagem X	50
Fig. 5.4. Máscara de afinamento.....	52

Fig. 5.5.	Família de elementos estruturantes usados no processo de pruning.....	53
Fig. 5.6.	Elementos estruturantes com centro em um, usados no processo de dilatação dos numerais.....	55
Fig. 5.7.	Fase de pré-processamento do numeral 4.....	55
Fig. 6.1.	Cavidades que constituem o fundo da imagem do numeral 6.....	58
Fig. 6.2.	Exemplo de aplicação do algoritmo para contagem de cavidades.....	60
Fig. 6.3.	Seqüências de cruzamento do numeral 6.....	62
Fig. 6.4.	Eixo principal e eixo secundário do numeral 6.....	63
Fig. 7.1.	Diagrama de classificação dos numerais desconhecidos de acordo com suas cavidades centrais.....	67
Fig. 7.2.	Ausência de pixels pretos à esquerda do eixo principal entre a 1ª e 2ª interseções no numeral 6.....	68
Fig. 7.3.	Numerais 2 e 8 com pixels pretos à esquerda do eixo principal entre a 1ª e 2ª interseções.....	68
Fig. 7.4.	Número de pixels pretos apresentados na última linha das matrizes representativas dos numerais 4 e	69
Fig. 7.5.	Segmento de reta apresentado pelo numeral 4.....	70
Fig. 7.6.	Numerais 2 e 8 com cavidades centrais superiores e pixels pretos à esquerda do eixo principal entre a 2ª e a última linha das matrizes representativas destes numerais.....	71
Fig. 7.7.	Pixels pretos à esquerda dos numerais 4 e 9.....	71
Fig. 7.8.	Diagrama de classificação de numerais que não apresentam cavidades centrais.....	73

Fig. 7.9. Número de pixels pretos apresentados na última linha das matrizes representativas dos numerais 7 e 3.....	74
Fig. 7.10. Diagrama da classificação dos caracteres manuscritos via rede neural.....	78

LISTA DE TABELAS

Tabela 8.1 - Total de Padrões Analisados e Classificados pelo Sistema.....	85
Tabela 8.2 - Total de Padrões Analisados e Classificados pelo 1º Estágio de Classificação.....	85
Tabela 8.3 - Padrões Analisados e Classificados pelo 2º Estágio de Classificação - Método da Projeção.....	86
Tabela 8.4 - Padrões Analisados e Classificados pelo 2º Estágio de Classificação - Método de Síntese para Sistemas - LSSM.....	87

RESUMO

A principal dificuldade em reconhecimento de caracteres manuscritos está em desenvolver métodos que possibilitem um ótimo índice de reconhecimento, apesar da grande variabilidade dos caracteres. Este trabalho apresenta um sistema para reconhecimento de numerais manuscritos desconectados, baseado na análise da topologia e distribuição pictorial de numerais e no emprego da rede de Hopfield discreta como memória associativa. O processo de classificação de numerais é dividido em dois estágios. No primeiro estágio, o numeral desconhecido é classificado de acordo com características extraídas de sua topologia e distribuição pictorial. Caso isto não seja possível, devido a distorções e ruídos na imagem do numeral desconhecido, a classificação é efetuada no segundo estágio de classificação do sistema, via rede de Hopfield. A rede de Hopfield é implementada de dois modos. No primeiro modo, a rede tem seus pesos calculados pelo método da projeção, no segundo modo, a rede tem os pesos calculados pelo método de síntese para sistemas lineares operando em modo saturado (LSSM). O sistema é testado com 1500 numerais manuscritos, sendo obtido uma taxa de 85% de reconhecimento com o sistema utilizando, no segundo estágio, a rede de Hopfield implementada conforme o primeiro modo. Uma taxa de reconhecimento de 84,4% é obtida com o sistema utilizando, no segundo estágio, a rede de Hopfield implementada de acordo com o segundo modo.

ABSTRACT

The main difficulty in handwritten character recognition consists in developing methods that provide a high recognition rate, although of the large degree of variability of the characteres. This work presents a system for recognition of disconnected handwritten numerals, based in analysis of the topology and distribution of pixels from the numerals, and application of a discrete Hopfield neural net used as associative memory. In the system, the process of classification is divided in two stages. In the first stage, the unknown numeral is classified considering features extracted from its topology and distribution of pixels. If it is not possible, due to distortions and noise in numeral image, the classification is effectuated in the second stage, via Hopfield net. The Hopfield net is implemented of two manners. In the first manner, the net has weights calculated by projection method and, in the second manner, the net has weights calculated by sinthesis procedures to linear systems operating in the saturated mode (LSSM systems). The system is tested with 1500 handwritten numerals. A recognition rate over 85% is obtained with the system making use in the second stage of the Hopfield net implemented by the first manner. A recognition rate over 84,4% is obtained with the system making use in the second stage of the Hopfield net implemented by the second manner.

CAPÍTULO 1 - INTRODUÇÃO

1.1 - RECONHECIMENTO DE CARACTERES MANUSCRITOS

O reconhecimento de padrões é parte essencial em um sistema de análise de imagens, e tem por objetivo analisar determinado evento ou objeto, e classificá-lo dentre as várias categorias preestabelecidas. O reconhecimento de caracteres é um exemplo típico de reconhecimento de padrões e, há algum tempo, vem sendo objeto de pesquisas para muitos cientistas. Têm-se procurado desenvolver algoritmos ou sistemas capazes de realizar o reconhecimento de caracteres manuscritos com uma performance semelhante a do ser humano, o que se justifica devido a grande importância científica e econômica das aplicações do reconhecimento de caracteres. Apesar de ser um assunto extensivamente estudado, o reconhecimento de caracteres manuscritos ainda permanece um difícil problema para os pesquisadores, devido a grande dificuldade em se desenvolver programas para computadores que possibilitem o reconhecimento das inúmeras variações de formas dos caracteres existentes.

1.2 - HISTÓRICO DO RECONHECIMENTO DE CARACTERES

A história do reconhecimento de caracteres teve início em 1870, quando Carey inventou um sistema de transmissão de imagens denominado *the retina scanner*, e em 1890 quando NipKow inventou o *scanner seqüencial* [1]. De qualquer modo, o reconhecimento de caracteres primeiramente apareceu como uma ajuda para deficiências visuais, e a primeira aplicação com sucesso foi declarada pelo cientista russo Tyurin em 1900 [1]. A versão moderna para os sistemas de reconhecimento de caracteres, também chamados *OCR* (Optical Character Recognition), surgiu em meados de 1940 com o desenvolvimento do computador

digital, quando pela primeira vez passou-se a utilizar estes sistemas no processamento aproximado de dados, visando aplicações comerciais.

1.3 - APLICAÇÕES DOS SISTEMAS DE RECONHECIMENTO DE CARACTERES

Atualmente uma das aplicações mais importantes dos sistemas de reconhecimento de caracteres está no processamento de informações em escritórios e repartições públicas, onde a maior parte do processo de armazenamento de informações vem sendo feito em papel. Os sistemas de reconhecimento de caracteres são capazes de fazer a conversão automática das informações contidas no papel para a forma digital, permitindo que estas informações sejam armazenadas em disquetes ou fitas magnéticas. Pode-se dividir as aplicações do reconhecimento de caracteres em 4 principais ramos [1-2]:

1- Processamento em sistemas bancários

Em atividades bancárias o reconhecimento de caracteres pode ser utilizado na verificação de cheques, os quais contêm um conjunto limitado de caracteres formado por números e símbolos especiais. O sistema deve reconhecer os números e símbolos que especificam o valor de um cheque com um alto grau de confiabilidade, tendo em vista a natureza das operações bancárias.

2- Processamento de textos

Este tipo de OCR analisa documentos com um formato de página normal em ambientes onde ocorre a edição de textos, como jornais e empresas de publicidade. O sistema de reconhecimento de caracteres é capaz de reconhecer estilos de caracteres alfanuméricos impressos, tais como, *courier*, *times new roman* e outros tipos de fontes de caracteres usados na edição de textos.

3- Processos de automação

Uma das aplicações principais deste ramo de reconhecimento de caracteres pode ser encontrada nas agências de correio, onde os OCR são usados para controlar algum tipo de processo responsável pela distribuição automática de correspondências para seu apropriado destino. O sistema de reconhecimento de caracteres é responsável pela leitura e interpretação das informações de endereçamento das correspondências tais como, cidade, país e código postal.

4- Aparelho de leitura para deficientes visuais

Uma aplicação de OCR muito importante para o homem é o desenvolvimento de aparelhos de leitura para deficientes visuais. Estes aparelhos em geral seriam compostos de um sistema de reconhecimento de caracteres combinado a um sistema de síntese de voz, facilitando entendimento de documentos escritos e livros por deficientes visuais.

1.4 - OBJETIVOS DA DISSERTAÇÃO

Nesta dissertação é tratado o reconhecimento de numerais manuscritos, o qual também é denominado a partir de agora, como reconhecimento de caracteres. O principal objetivo desta dissertação é a apresentação do sistema de reconhecimento de caracteres desenvolvido no decorrer de nossas pesquisas, bem como a descrição das técnicas empregadas na elaboração das funções que constituem o sistema. Conforme foi relatado, muitas técnicas estão sendo pesquisadas e empregadas na área de reconhecimento de caracteres e algumas são bastantes eficientes [3-7], permitindo o desenvolvimento de sistemas com altos índices de reconhecimento. Durante nosso trabalho procuramos pesquisar, dominar e implementar muitas

destas técnicas, para que fosse possível desenvolver nosso próprio sistema de reconhecimento de caracteres, o qual acreditamos ter alcançado um bom desempenho. Ainda, como parte dos objetivos desta dissertação, está a apresentação de forma sucinta de alguns dos métodos mais empregados no reconhecimento de caracteres, com a finalidade de mostrar diferentes maneiras de tratar o problema do reconhecimento de caracteres manuscritos.

1.5 - ORGANIZAÇÃO DA DISSERTAÇÃO

Após terem sido estabelecidos na seção anterior os principais objetivos, é definida nesta seção a estrutura desta dissertação. Começando a partir do capítulo 2, são descritas as fases em que se dividem os sistemas de reconhecimento de caracteres e apresentados os termos utilizados neste tipo de reconhecimento. Logo a seguir são descritos os problemas existentes no reconhecimento de caracteres manuscritos e alguns dos métodos mais empregados para solucionar estes problemas.

O capítulo 3 é dedicado à apresentação da rede neural de Hopfield discreta empregada no processo de classificação de nosso trabalho. Inicialmente, são apresentados conceitos básicos sobre as redes neurais em geral, os quais envolvem: a) Neurônio artificial; b) Funções de ativação; c) Treinamento ou aprendizagem da rede. Após, é descrita a filosofia da rede de Hopfield discreta, seu funcionamento, capacidade de armazenamento e sua função de energia. Finaliza-se o capítulo com a apresentação de sistemas lineares operando em modo saturado (Sistemas LSSM), suas equações e o modo como o processo de síntese, utilizado para estes sistemas, é aplicado a rede de Hopfield visando melhorar a performance da rede.

O capítulo 4 refere-se a conceitos básicos da morfologia matemática, área que envolve os processos de dilatação, afinamento e pruning usados no pré-processamento das imagens dos

numerais manuscritos. São apresentadas transformações básicas da morfologia matemática, um conceito para elemento estruturante e operações “tudo ou nada”. A parte final do capítulo é dedicada a uma apresentação geral de nosso sistema de reconhecimento de numerais manuscritos, para que se possa ter uma visão global de seu funcionamento antes de serem apresentadas cada uma de suas partes.

No capítulo 5 passa-se a descrição das primeiras fases que compõem nosso sistema de reconhecimento de numerais manuscritos. Inicialmente, o capítulo apresenta o processo de aquisição dos numerais manuscritos, digitalização e armazenamento da imagem. O restante do capítulo é dedicado a descrição das técnicas utilizadas na fase de pré-processamento do sistema: a) Afinamento; b) Pruning; c) Dilatação; d) Normalização em escala. Estas técnicas têm o objetivo de melhorar a imagem dos numerais manuscritos .

O Capítulo 6 descreve a parte do sistema de reconhecimento de numerais que trata da extração de características das imagens, fase responsável pela obtenção dos atributos usados para classificar os caracteres desconhecidos. Primeiramente, é descrita a técnica empregada para contagem dos elementos conexos da imagem de um numeral. Usando esta técnica, extrai-se os primeiros atributos do numeral desconhecido, sendo estes o número e a posição das cavidades do numeral. Após, são explicados os atributos obtidos através das seqüências de cruzamento do numeral, que são a quantidade de blocos de pixels pretos existentes em cada linha e coluna da matriz representativa da imagem do numeral. O capítulo é concluído com a descrição dos atributos extraídos das interseções da imagem com seu eixo principal e eixo secundário, e os atributos obtidos da distribuição pictorial da imagem.

O capítulo 7 é dedicado à fase de classificação dos caracteres, que é dividida em dois estágios: 1) Utilização dos atributos obtidos na fase de extração de características para

classificar aproximadamente 80% dos caracteres analisados pelo sistema de reconhecimento de caracteres; 2) Utilização da rede de Hopfield discreta para classificar o restante dos caracteres (aproximadamente 20%), que não são classificados no primeiro estágio. Neste capítulo são testados dois métodos para implementação da rede de Hopfield usada no segundo estágio de classificação, são eles: método da projeção e processo de síntese para os sistemas LSSM, ambos apresentados no capítulo 3.

Finalmente no capítulo 8, são apresentados os resultados obtidos nos testes de nosso sistema de reconhecimento de numerais manuscritos, conclusões finais e sugestões para um trabalho futuro, com o objetivo de aperfeiçoar o sistema.

CAPÍTULO 2 - ALGUNS ASPECTOS DO RECONHECIMENTO DE CARACTERES

2.1 - INTRODUÇÃO

Os sistemas de reconhecimento de caracteres, em geral, utilizam técnicas que procuram descrever um caracter da forma mais completa possível. Estas técnicas são colocadas na forma de um programa de computador, tornando possível a extração de características e classificação de um caracter desconhecido dentre vários outros caracteres.

Este capítulo apresenta uma visão geral sobre os sistemas de reconhecimento de caracteres e sobre as técnicas mais empregadas para realizar o reconhecimento de caracteres.

2.2 - ESTRUTURA DOS SISTEMAS DE RECONHECIMENTO DE CARACTERES

Esta seção, inicialmente, apresenta as definições para processamento de imagens, análise de imagens e reconhecimento de padrões da forma como elas foram introduzidas pelos pioneiros destas áreas, K. S. Fu e A. Rosenfeld [8]:

- **Processamento de Imagens:** Operação que transforma imagens em outras, produzindo uma versão melhorada das imagens originais, com condições necessárias para serem usadas em uma determinada aplicação.
- **Análise de Imagens:** Estudo das características que descrevem uma imagem. Estas características são expressas na forma de estruturas, reproduzindo as relações e as propriedades entre as partes formadoras da imagem.
- **Reconhecimento de Padrões:** Primeiramente diz respeito a descrição e classificação de medidas obtidas de um processo físico ou mental. Mas também

denomina-se reconhecimento de padrões de imagens, quando no contexto da análise e classificação de uma determinada imagem.

Do ponto de vista de desenvolvimento de um sistema de reconhecimento de padrões, o processamento da imagem de entrada é denominado pré-processamento e a análise da imagem é chamada de extração de características ou descrição do padrão [8].

Os sistemas de reconhecimento de padrões geralmente possuem a estrutura mostrada pela figura 2.1, podendo ser separada nas seguintes fases [9]:

- **Aquisição da imagem:** É o processo em que a imagem é convertida em um conjunto de números que podem ser manipulados via computador.
- **Pré-processamento:** O pré-processamento visa remover ruídos, melhorar a imagem e se necessário segmentá-la em regiões, para que cada uma destas regiões possa ser analisada separadamente.
- **Extração de características:** Na fase de extração de características tenta-se representar a imagem por um conjunto de características próprias, visando a remoção de redundâncias e redução da dimensão do conjunto de dados.
- **Classificação:** Na classificação é atribuído um rótulo para a imagem desconhecida através da análise das características extraídas e comparação destas com uma classe de padrões aprendida pelo classificador durante um estágio de treinamento. Idealmente um bom classificador deve ser capaz de reconhecer um objeto independente de variações ocorridas na sua orientação, tamanho e localização na cena.

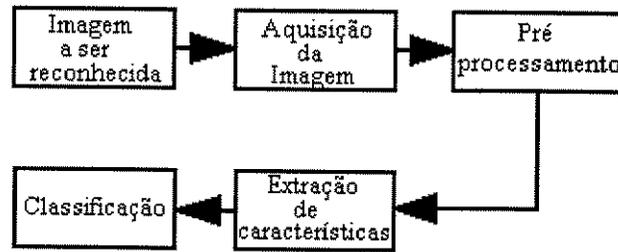


Fig. 2.1. Diagrama de blocos de um sistema de reconhecimento de padrões.

2.3 - PRINCIPAIS PROBLEMAS DO RECONHECIMENTO DE CARACTERES MANUSCRITOS

O avanço da tecnologia computacional tem permitido a elaboração de métodos cada vez mais eficientes para tratar o problema do reconhecimento de caracteres, no entanto, o reconhecimento de caracteres manuscritos ainda é um problema complexo e não totalmente resolvido. São descritos abaixo os principais fatores de erros de classificação em sistemas de reconhecimento de caracteres manuscritos, esses fatores tornam o reconhecimento de caracteres uma tarefa complexa. Note que estes fatores também determinam a sensibilidade de um sistema à deformação da imagem do caracter. São eles:

- a) **Ruídos** - Elementos não pertencentes a imagem original, podendo ser caracterizados por segmentos de linhas desconectados, lacunas, pontos isolados, segmentos de linhas espúrios, pontos que não deveriam estar conectados, etc.;
- b) **Distorções** - As distorções incluem variações locais do caracter, cantos arredondados, saliências impróprias, variações de tamanho, etc.;
- c) **Variações de estilo** - Consiste no uso de diferentes formas para representar um mesmo caracter, bem como traços e inclinações usados para personalizar o caracter;
- d) **Translação** - É definida como o movimento do caracter inteiro ou de seus componentes em relação a uma determinada posição de referência;

e) **Rotação** - É a mudança na orientação do caracter.

O ruído é, geralmente, introduzido devido ao processo de digitalização da imagem, superfície de escrita ou instrumento de escrita. As distorções e variações de estilo são principalmente produzidas pelo escritor. Translações e rotações podem resultar de ambos, escritor e parte mecânica do sistema (digitalizador, superfície de escrita ou instrumento de escrita). Estes fatores podem ser atenuados ou eliminados na fase de pré-processamento da imagem e a análise deles juntamente com os outros fatores envolvidos na implementação prática do sistema, que serão descritos na próxima seção, determinam a eficiência do método de reconhecimento.

2.4 - MÉTODOS MAIS EMPREGADOS NO RECONHECIMENTO DE CARACTERES

Numerosos métodos têm sido desenvolvidos para realizar o reconhecimento automático de caracteres manuscritos. Estes métodos variam de acordo com as características escolhidas, o modo de extração destas características e o esquema de classificação usado. Podem ser divididos em três categorias: *transformações globais*, *distribuição de pontos* e *características topológicas e geométricas* [10]. Estas categorias serão descritas a partir da subseção 2.4.1. Os méritos de cada um destes métodos podem ser analisados de acordo com o critério da sensibilidade à deformação da imagem do caracter, causada pelos fatores descritos na seção anterior e o critério da implementação prática do método, conforme destacado a seguir:

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:** Quão rápido 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 determinada performance sem ser necessário o uso de técnicas suplementares.

2.4.1 - TRANSFORMAÇÕES GLOBAIS

As características extraídas pelas transformações globais representam uma imagem como um todo, não refletem nenhuma propriedade local, geométrica ou topológica de uma região da imagem. Algumas destas técnicas foram desenvolvidas para reconhecer apenas caracteres impressos. Abaixo, são apresentadas as técnicas que fazem parte das transformações globais:

- **Similaridade e correlação (Template matching):** Esta técnica mede a similaridade entre uma 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 [11-15], [24].
- **Transformadas e expansões em séries:** Estes métodos examinam a imagem com relação ao modo como ela é composta por séries de funções periódicas. Foram criados com a finalidade de extrair características que fossem invariantes a alguma deformação global da imagem, tais como, rotação e translação. As mais exploradas têm sido as séries de Fourier, a transformada de Fourier [16-18], a transformada de Walsh [19], a transformada de Haar e Hadamard [20-21], bem como a transformada de Karhunen-Loeve [22-23], [40].

O principal problema apresentado pelas transformações globais é sua alta sensibilidade às distorções e variações de estilos de escrita. Por exemplo, manchas ou ruídos podem deslocar os pontos da imagem de sua posição de alinhamento, afetando seriamente as características obtidas por estas técnicas.

2.4.2 - DISTRIBUIÇÃO DE PONTOS

As técnicas envolvidas neste grupo extraem atributos derivados da distribuição estatística dos pontos que formam a imagem do caracter, produzindo um conjunto de características de dimensão reduzida. Abaixo, são citadas técnicas mais empregadas dentro deste grupo:

- Divisão em zonas (zoning):** 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 nestas regiões [4].
- **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 [9].
 - **Características Loci:** Para cada pixel branco do fundo da imagem, um conjunto de vetores vertical e horizontal são gerados. O conjunto de características usado neste caso, é dado pelo número de interseções que estes vetores fazem com os segmentos de linha que formam o caracter [25].
 - **Distâncias e cruzamentos:** A característica de cruzamento é obtida do número de vezes com que o caracter é cortado por segmentos de linha especificados em

determinadas direções [4], [26]. As distâncias medidas dos pontos que formam o caracter a um ponto especificado na imagem, por exemplo, os pontos que determinam o limite do retângulo que contém o caracter, formam um outro conjunto de características.

A considerável tolerância às distorções e às pequenas variações de estilos são observadas nas técnicas descritas acima, devido a algumas características levarem em conta algum tipo de informação topológica do caracter. Este fato é evidenciado principalmente pelas técnicas de características Loci, distâncias e cruzamentos.

Note que a implementação de máscaras para estas técnicas não é trivial, devido a grande variação de forma dos caracteres manuscritos. Por outro lado, estas técnicas permitem uma alta velocidade de reconhecimento e uma baixa complexidade de implementação.

2.4.3 - CARACTERÍSTICAS TOPOLÓGICAS E GEOMÉTRICAS

Esta categoria baseia-se na extração de características que descrevem aspectos importantes da geometria e da topologia do desenho do caracter. As técnicas incluídas nesta categoria podem representar propriedades globais e locais do caracter. Alguns exemplos são apresentados a seguir:

- **Segmentos de reta e linhas curvas em várias direções:** A figura 2.2 exemplifica segmentos de reta e curvas, comumente extraídos do caracter e utilizados para representar sua geometria [4], [17], [27].

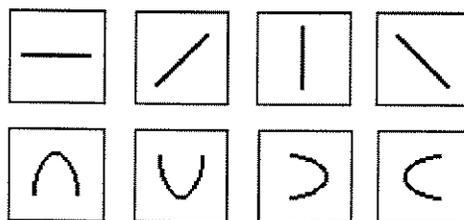


Fig. 2.2. Exemplos de segmentos de reta e curvas a serem extraídos da forma do caracter.

- **Pontos finais, interseções de linhas e loops:** Estes tipos de características são representativos da topologia do caracter, a figura 2.3 mostra alguns exemplos destas características [27].

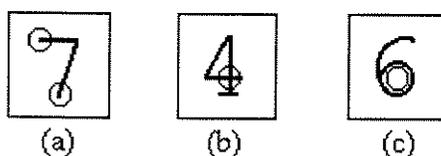


Fig. 2.3. Exemplos de características extraídas da topologia dos caracteres. (a) Os círculos nas extremidades do número 7 marcam os pontos finais do numeral. (b) O círculo sobre o número 4 mostra um ponto de interseção existente na topologia do caracter. (c) O círculo nesta figura mostra a posição do loop encontrado na parte inferior do numeral 6.

Devido a 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 estas características. Entretanto, uma vez implementada estas técnicas, elas permitem ao sistema uma grande velocidade de processamento e um alto índice de classificações corretas dos caracteres.

2.5 ALGUMAS CONSIDERAÇÕES FINAIS

Os itens 2.4.1, 2.4.2 e 2.4.3 resumem as técnicas mais empregadas no reconhecimento de caracteres, observado o critério de sensibilidade à deformação da imagem, cujas causas

foram descritas na seção 2.3, e o critério da implementação prática do sistema, descrito no início desta seção. Estas técnicas abordam as mais importantes e comuns propriedades dos caracteres manuscritos, bem como, os principais problemas do reconhecimento de caracteres estudados pelos pesquisadores. Muitos sistemas combinam duas ou mais destas técnicas, com o objetivo de obter um maior número de características e eliminar aquelas que provaram ser defeituosas ou indesejáveis. Desta maneira, uma técnica complementa a outra, tendo em vista a sensibilidade às deformações dos caracteres, resultando em um sistema de reconhecimento de caracteres mais robusto.

Atualmente o reconhecimento de caracteres via redes neurais vem se tornando uma técnica bastante usada pelos pesquisadores. A capacidade das redes neurais em aprender determinado conhecimento, armazenar este conhecimento e usá-lo mais tarde para executar uma tarefa, tem proporcionado bons resultados na área de reconhecimento de caracteres [4], [6], [28]. O próximo capítulo trata, mais especificamente, o tema das redes neurais, descrevendo a rede de Hopfield discreta utilizada neste trabalho.

CAPÍTULO 3 - REDES NEURAI

3.1 - INTRODUÇÃO

A possibilidade de dar às máquinas a habilidade que o ser humano possui em aprender, reter informações, recordá-las e aplicá-las na solução de diversos tipos de problemas, tem levado muitos pesquisadores a procurar desenvolver modelos computacionais para os mesmos fins. Tais modelos são chamados sistemas neurais artificiais ou simplesmente redes neurais.

Uma rede neural tem a capacidade de aprender um dado tipo de conhecimento, armazenar este conhecimento e usá-lo para executar certa tarefa. Podemos citar, entre outras, as seguintes aplicações das redes neurais:

- Conversão de textos em representações fonéticas;
- Reconhecimento de voz;
- Reconhecimento de caracteres;
- Reconstrução de imagens;
- Compressão de imagens.

Este capítulo faz uma breve descrição dos conceitos fundamentais da teoria de redes neurais necessários a compreensão deste trabalho. Especificamente, apresenta o modelo da rede de Hopfield discreta empregado no desenvolvimento de nosso sistema de reconhecimento de numerais manuscritos.

3.2 - CONCEITOS BÁSICOS SOBRE AS REDES NEURAIS

Em geral, uma rede neural consiste de um grande número de unidades básicas de processamento denominadas neurônios, células ou nós, como mostra a figura 3.1. Cada neurônio é diretamente conectado a outro neurônio, e a cada conexão é associado um peso, representando informações que a rede dispõe para resolver um problema.

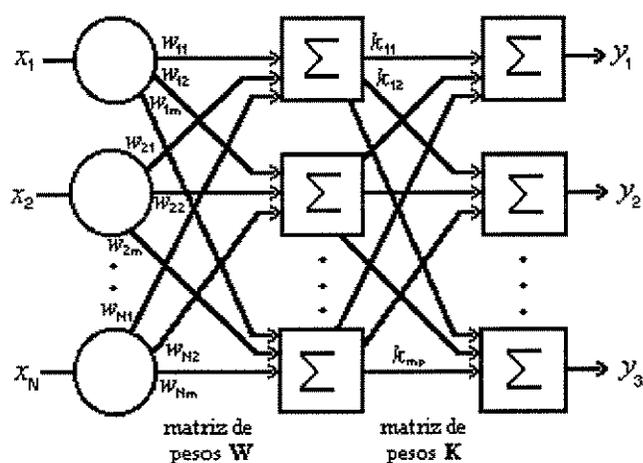


Fig. 3.1. Rede neural com duas camadas de neurônios [29].

Uma rede neural é caracterizada por:

- 1- Tipo de conexões entre neurônios (arquitetura da rede);
- 2- Modo como os pesos entre conexões são determinados (aprendizagem ou treinamento);
- 3- Sua função de ativação.

A seguir, é apresentada uma definição de neurônio artificial, são descritos seus principais componentes e princípio de operação.

3.2.1 - NEURÔNIO ARTIFICIAL

A figura 3.2 mostra esquematicamente os elementos que compõem um neurônio artificial.

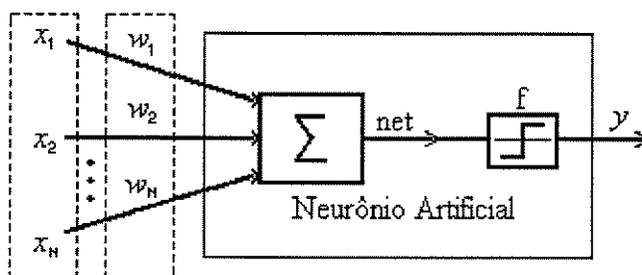


Fig. 3.2. Neurônio artificial.

Um vetor $\mathbf{x} = (x_1, \dots, x_n)$ é o conjunto de entradas aplicado ao neurônio artificial e representa os sinais nas sinapses de um neurônio biológico. O conjunto de entradas é multiplicado por um vetor $\mathbf{W} = (w_1, \dots, w_n)$ de pesos, representando um processo químico nas sinapses de um neurônio biológico, que modifica o sinal de entrada ao transmiti-lo. O corpo do neurônio, representado por Σ , soma algebricamente todos os sinais x_j multiplicados por seus correspondentes pesos. O resultado desta operação, denotado net , pode ser expresso por:

$$net = \sum_{j=1}^N x_j w_j. \quad (3.1)$$

Em seguida o sinal net é processado por uma função de ativação f , cuja finalidade é ativar o neurônio somente quando o valor de net exceder a um determinado valor de limiar. Quando isto ocorre o neurônio produz um sinal de saída y , que é enviado a outros neurônios.

$$y = f(net). \quad (3.2)$$

3.2.2 - FUNÇÕES DE ATIVAÇÃO

Vários tipos de funções de ativação podem ser usadas para ativar um neurônio artificial, porém o uso de uma particular função depende do tipo de dados de saída (contínuo ou discreto) e da faixa em que se encontram estes dados (por exemplo de -1 a 1). As funções de ativação, mais comuns, encontradas na literatura são:

a) Função linear:

$$f(x) = x;$$

b) Função degrau unitário:

$$f(x) = \begin{cases} 1, & \text{se } x \geq 0 \\ 0, & \text{se } x < 0 \end{cases};$$

c) Função bipolar:

$$f(x) = \begin{cases} 1, & \text{se } x \geq 0 \\ -1, & \text{se } x < 0 \end{cases};$$

d) Função tangente hiperbólica:

$$f(x) = \text{tgh}(x);$$

e) Função Sigmoidal:

$$f(x) = \frac{1}{(1 + e^{-x})}.$$

3.2.3 - APRENDIZAGEM

Antes de uma rede neural executar certa tarefa é necessário que ela seja treinada, isto é, os valores de pesos sejam ajustados gradualmente de acordo com seu algoritmo de

treinamento, convergindo para valores, tais que, cada vetor de entrada produza o vetor de saída desejado. Abaixo, são citados tipos de treinamento encontrados na literatura [28].

3.2.3.1 - Treinamento Supervisionado

Neste tipo de treinamento é apresentado à rede uma seqüência de vetores, cada um associado a um vetor de saída. Os vetores de treinamento são aplicados seqüencialmente a rede, os erros são calculados e os pesos são ajustados para cada vetor, até que o erro para o conjunto inteiro atinja um valor mínimo aceitável.

3.2.3.2 - Treinamento não Supervisionado

Neste treinamento são aplicados a rede apenas os vetores de entrada, sem seus correspondentes vetores de saída. O algoritmo de treinamento modifica os valores de pesos para produzir uma saída consistente, isto é, se aplicado um dos vetores de treinamento ou um vetor suficientemente similar a ele, ambos deverão produzir uma mesma saída.

3.2.3.3 - Pesos Fixados

Neste caso, não há nenhum processo de ajuste iterativo de peso, os valores de pesos são fixados de maneira a dar informações sobre a saída desejada de acordo com as restrições e as quantidades a serem minimizadas ou maximizadas.

A próxima seção apresenta o modelo da rede de Hopfield discreta usada no processo de classificação de numerais manuscritos, descrito no capítulo 7 desta dissertação. A seção descreve o funcionamento básico da rede de Hopfield, seu algoritmo de treinamento, sua capacidade de armazenamento e sua função de energia.

3.3 - REDE DE HOPFIELD DISCRETA

A rede de Hopfield como rede neural aplicada ao reconhecimento de padrões, visa reconhecer um padrão, parcialmente ou totalmente distorcido, como sendo um de seus padrões previamente armazenados, e produzir como saída a versão completa e perfeita do padrão memorizado [30], [41]. Devido a esta propriedade, nosso sistema de reconhecimento de caracteres manuscritos é implementado empregando a rede de Hopfield no reconhecimento de numerais manuscritos. É considerado que as diversas variações de formas de um numeral sejam versões distorcidas de um padrão previamente armazenado.

3.3.1 - ARQUITETURA DA REDE DE HOPFIELD DISCRETA

A arquitetura da rede de Hopfield é ilustrada através de um esquema pela figura 3.3.

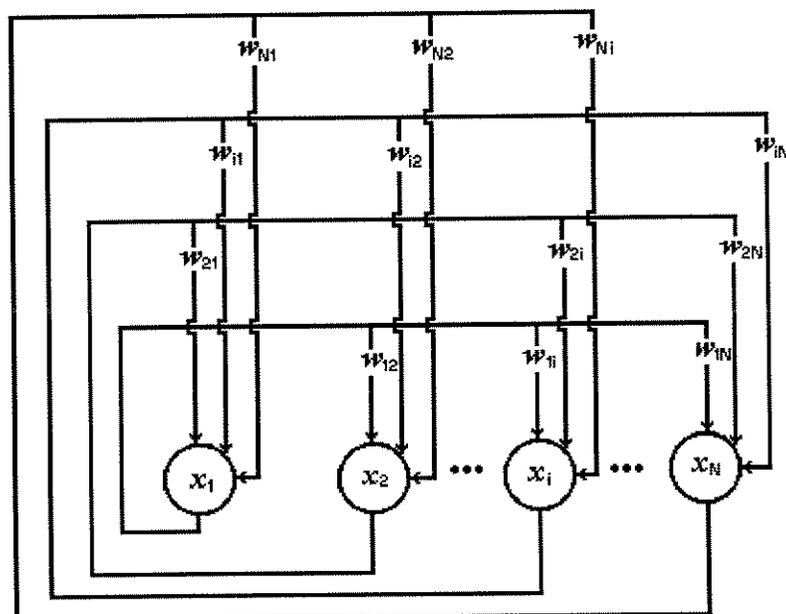


Fig. 3.3. Esquema da rede de Hopfield discreta [28].

$\mathbf{x} = (x_1, \dots, x_n)$ representa um vetor de entrada aplicado a rede neural, uma matriz de pesos $\mathbf{W} = [\mathbf{W}_{ij}]$ é calculada de forma a permitir que a rede armazene diferentes padrões de treinamento (estados da rede). Se qualquer um dos padrões armazenados é apresentado à entrada da rede, a rede permanece naquele estado e produz como saída um vetor $\mathbf{Y} = (y_1, \dots, y_n)$, que é o próprio vetor de entrada. Se uma versão distorcida de um padrão armazenado é apresentada à rede, esta evoluirá para o estado memorizado que mais se aproxima do padrão apresentado e produzirá, como saída \mathbf{Y} , o estado memorizado [31].

3.3.2 - OPERAÇÃO DA REDE DE HOPFIELD DISCRETA

O processamento feito por cada unidade da rede de Hopfield é similar ao discutido na subseção 3.2.1 deste capítulo. Cada unidade calcula a soma do produto de suas entradas e seus pesos para determinar o valor total de entrada no neurônio, como mostra a equação abaixo:

$$net_i = \sum_{j=1}^N x_j w_{ji}, \quad i = 1, \dots, N \text{ e } j = 1, \dots, N \quad (3.3)$$

onde:

net_i : valor total de entrada no neurônio i ;

$\mathbf{x} = (x_1, \dots, x_n)$: vetor de entrada que representa o padrão a ser analisado pela rede;

$\mathbf{W} = [\mathbf{W}_{ij}]$: matriz de pesos da rede;

N : dimensão do vetor de entrada.

A função de ativação usada para determinar a saída y de um neurônio qualquer da rede, é uma função bipolar do tipo:

$$y_i = f(\text{net}_i) = \begin{cases} 1, & \text{se } \text{net}_i \geq 0 \\ -1, & \text{se } \text{net}_i < 0 \end{cases} \quad (3.4)$$

Cada neurônio tem sua saída atualizada de maneira assíncrona, isto é, a cada interação um neurônio é escolhido aleatoriamente para ter sua saída modificada de acordo com a equação (3.4). Os seguintes passos são seguidos no emprego da rede neural de Hopfield neste trabalho:

Passo 0 - Escolhe-se aleatoriamente um neurônio i para ter sua saída atualizada;

Passo 1 - Calcula-se o valor da entrada net_i , do neurônio i , através da equação (3.3);

Passo 2 - Determina-se o valor da saída y , do neurônio escolhido no passo 0, através da equação (3.4);

Passo 3 - Verifica-se se a condição de convergência:

$$y_i = x_i, \quad (3.5)$$

é verdadeira. Se for verdadeira o passo 5 é executado, caso contrário o passo 4 é efetuado ;

Passo 4 - Altera-se o valor de x_i de maneira a se ter:

$$x_i = y_i. \quad (3.6)$$

Os passos de 1 a 3 são repetidos até que a condição do passo 3 seja verdadeira;

Passo 5 - Escolhe-se aleatoriamente outro neurônio para ter sua saída atualizada;

Passo 6 - Os passos de 1 a 5 são repetidos até que todas as unidades tenham sido atualizadas.

3.3.3 - TREINAMENTO DA REDE DE HOPFIELD DISCRETA

O processo de treinamento para a rede de Hopfield discreta é também chamado processo de armazenamento de vetores. Para armazenar um conjunto com P vetores de treinamento $\mathbf{X}^p = (x_1^p, \dots, x_n^p)$ $p = 1, \dots, P$, na rede neural, calcula-se a matriz de pesos $\mathbf{W} = [\mathbf{W}_{ij}]$ da seguinte forma:

$$w_{ij} = \sum_{p=1}^P x_i^p x_j^p \quad i = 1, \dots, N \text{ e } j = 1, \dots, N. \quad (3.7)$$

Esta regra de aprendizagem é denominada *regra de Hebb* ou *regra generalizada de Hebb* e a obtenção de valores corretos de pesos depende da correlação entre os vetores de treinamento [28]. Se os vetores de treinamento forem incorrelatos (ortogonais) a regra de Hebb produzirá pesos corretos, e a rede quando testada com um dos vetores de treinamento recuperará o próprio vetor. Se os vetores de entrada não forem ortogonais aos vetores de treinamento, a rede produzirá uma resposta com erro, que tem a contribuição de cada vetor de treinamento. Este erro é chamado de *cross talk* [28], [41], o qual será tratado no tópico a seguir.

3.3.4 - CAPACIDADE DE ARMAZENAMENTO DA REDE DE HOPFIELD DISCRETA

A rede de Hopfield é considerada estável quando produz na sua saída um dos padrões de treinamento armazenados, ou seja, quando a seguinte condição ocorre:

$$f(\text{net}_i^v) = x_i^v, \quad \text{para todo } i. \quad (3.8)$$

Onde, net_i^v dado por (3.9), é o valor total de entrada calculado para o neurônio i quando apresentado a rede um padrão v .

$$\text{net}_i^v = \sum_{j=1}^N x_j^v w_{ji}. \quad (3.9)$$

Substituindo a equação (3.7) em (3.9), tem-se:

$$net_i^v = \sum_{j=1}^N \sum_{p=1}^P x_j^p x_i^p x_j^v. \quad (3.10)$$

Separando do somatório em p o termo com $p = v$ da equação (3.10), obtém-se:

$$net_i^v = x_i^v + \underbrace{\sum_{j=1}^N \sum_{p \neq v}^P x_j^p x_i^p x_j^v}_{\text{Cross talk}}. \quad (3.11)$$

O segundo termo do lado direito da equação (3.11), denominado *cross talk*, deve ser nulo para ser obtido o resultado apresentado pela equação (3.8). Isto somente acontece se o vetor de entrada, \mathbf{X}^v , é ortogonal. Se \mathbf{X}^v não é ortogonal e o termo de *cross talk* for menor do que 1, o sinal de x_i^v não é alterado, e a função de ativação produz o valor esperado, a rede então recupera o correto padrão armazenado. Se o termo de *cross talk* é maior do que 1, o sinal de x_i^v é alterado, e a rede não recupera o valor desejado, atingindo um estado espúrio de estabilidade.

Em muitos casos de interesse se o número total P , de padrões armazenados, for pequeno o suficiente para que o termo de *cross talk* não mude o sinal de x_i^v , a rede será estável. Hopfield encontrou experimentalmente o número P de padrões que podem ser armazenados e recuperados pela rede com razoável precisão [28], [36]:

$$P = 0.15 \times N, \quad (3.12)$$

onde N é a dimensão do vetor de entrada \mathbf{X} .

Abu-Mostafa e St. Jacques [32] realizaram uma detalhada análise teórica sobre a capacidade de armazenamento da rede de Hopfield, e mostraram que o número de estados da rede não pode exceder a N .

3.3.5 - PADRÕES CORRELATOS - MÉTODO DA PROJEÇÃO

Como discutido anteriormente, o termo de *cross talk* é um limite para a capacidade de armazenamento da rede de Hopfield. No entanto, quando os padrões de entrada são correlatos, e na maioria das vezes eles o são, o problema torna-se mais grave, pois a rede não consegue recuperar padrões confiáveis, estabilizando-se em estados espúrios, onde os padrões recuperados não fazem parte do conjunto de referência armazenado. Uma solução para este problema pode ser obtida via método de aproximação pelo cálculo da pseudo-inversa. O método trabalha com P padrões ($P < N$) linearmente independentes e uma matriz $\mathbf{Q} = [\mathbf{Q}_{pv}]$ de dimensão $P \times P$ é definida por:

$$q_{pv} = \sum_{i=1}^N x_i^p x_i^v, \quad p = 1, \dots, P \text{ e } v = 1, \dots, P. \quad (3.13)$$

Se a matriz \mathbf{Q} é não singular, os elementos da matriz de pesos \mathbf{W} podem ser obtidos por:

$$w_{ij} = \sum_p \sum_v x_i^p (q^{-1})_{pv} x_j^v. \quad (3.14)$$

Se \mathbf{Q} não pode ser invertida, há alguma dependência linear entre os padrões de treinamento. Neste caso, pode-se restringir o conjunto de padrões de treinamento a um subconjunto de padrões linearmente independentes.

Note que a matriz \mathbf{W} , calculada através da equação (3.14), projeta os vetores de entrada sobre um subespaço de dimensão P gerado pelos vetores de treinamento da rede [41]. Quando os vetores de entrada não são correlatos, a rede produz o correto padrão armazenado. Se os vetores de entrada são correlatos, a rede recupera o padrão armazenado mais próximo do vetor de entrada, minimizando o erro quadrático médio [41], [44].

O método da projeção é uma das regras de aprendizagem usadas por nós para implementação da rede de Hopfield neste trabalho. Na seção 3.4 é descrita uma outra regra de aprendizagem também empregada à rede de Hopfield. Os resultados da aplicação destas 2 regras de aprendizagem são apresentadas no capítulo 8.

3.3.6 - FUNÇÃO DE ENERGIA DA REDE DE HOPFIELD DISCRETA

Uma função de energia tem como principal característica seu decréscimo à medida que a rede evolui, de acordo com o algoritmo de aprendizagem adotado, alcançando seu ponto de mínimo e assegurando a estabilidade da rede. A função apresentada a seguir, que concorda com a característica relatada acima, é denominada função de Lyapunov [28-29].

$$E = -1/2 \sum_i \sum_j y_i y_j w_{ij} - \sum_i I_j y_j + f_j y_j, \quad (3.15)$$

onde:

E = função de energia da rede neural;

w_{ij} = peso da saída do neurônio i para a entrada do neurônio j ;

y_j = saída do neurônio j ;

I_j = entrada externa ao neurônio j ;

f_j = função de ativação do neurônio j .

Note que quando ocorrer uma mudança de Δy_j na saída y_j , do neurônio j , a energia do sistema muda por um valor:

$$\Delta E = - \left[\sum_{i \neq j} y_j w_{ij} + I_j - f_j \right] \Delta y_j. \quad (3.16)$$

Pode-se verificar pela equação (3.16) que Δy_j é positiva (saída y_j passa de -1 para +1) somente se $\left[\sum_{i \neq j} y_i w_{ij} + I_j - f_j \right]$ positivo, e Δy_j é negativa (y_j passa de +1 para -1) somente se $\left[\sum_{i \neq j} y_i w_{ij} + I_j - f_j \right]$ negativo. Isso demonstra que a energia somente decresce ou permanece em um mesmo estado durante o processo de interação do sistema. Como a energia é limitada, a rede deve alcançar um estado de equilíbrio no qual a energia não mudará, mesmo que continue as interações do sistema. Quando isto acontece a rede é dita estável e não ocorrem mais mudanças nos valores de sua saída.

Na próxima seção são introduzidos os sistemas LSSM, os quais serão utilizados para calcular outra matriz de pesos para a rede de Hopfield discreta, apresentando assim, uma outra solução para o problema de padrões correlatos discutidos nos itens 3.3.4 e 3.3.5.

3.4 - SISTEMAS LSSM APLICADOS AO MODELO DA REDE DE HOPFIELD DISCRETA

Esta seção apresenta uma definição de sistemas LSSM e a maneira com que eles são aplicados ao modelo da rede de Hopfield utilizado em nosso trabalho. Nosso objetivo é reduzir as limitações da rede de Hopfield causadas pelo termo de *cross talk* (3.11) e, por conseguinte, reduzir o número de estados espúrios da rede. Para isto, é empregado o processo de síntese para sistemas LSSM, conforme descritos nos parágrafos a seguir. É importante salientar que apenas os resultados necessários para aplicação dos sistemas LSSM à rede de Hopfield são apresentados aqui. Maiores detalhes sobre a formulação de problemas, desenvolvimento teórico e análise e prova dos resultados podem ser encontrados em [33], [42-43].

3.4.1 - SISTEMAS LSSM

A classe de redes neurais descrita por um sistema de equações lineares ordinárias de primeira ordem, definidas em um subespaço fechado \mathbf{D}^n de \mathbf{R}^n , é referida como *sistemas lineares operando em modo saturado (LSSM)* [33]. São consideradas especificamente as redes descritas por equações da forma:

$$\frac{dx}{dt} = \mathbf{T}\mathbf{x} + \mathbf{I}, \quad (3.17)$$

com a seguinte restrição:

$$-1 \leq x_i \leq 1, \quad i = 1, \dots, n.$$

Onde:

$$\mathbf{x} = (x_1, \dots, x_n)^T \in \mathbf{D}^n;$$

$$\mathbf{D}^n = \{\mathbf{x} \in \mathbf{R}^n : -1 \leq x_i \leq 1, \quad i = 1, \dots, n\};$$

$\mathbf{T} = [\mathbf{T}_{ij}]$ é uma matriz $n \times n$ de constantes;

$\mathbf{I} = (I_1, \dots, I_n)^T$ é um vetor de constantes.

A principal diferença entre os sistemas LSSM dados pela equação (3.17) e os demais sistemas lineares, é que os últimos são definidos em subespaços abertos de \mathbf{R}^n , enquanto que os sistemas LSSM são definidos em um subespaço fechado \mathbf{D}^n de \mathbf{R}^n . Para os sistemas dados por (3.17) pode-se introduzir um novo tipo de soluções, denominadas soluções no modo *saturado*, as quais são definidas sobre a fronteira de \mathbf{D}^n , denotada $\partial\mathbf{D}^n$. Os resultados estabelecidos a partir destas definições [33] podem ser divididos em dois tipos de processos, processo de *análise* e processo de *síntese* para sistemas LSSM.

O processo de análise estuda o comportamento de sistemas LSSM a partir do desenvolvimento de uma definição de soluções para a equação (3.17). Na fronteira ∂D^n , soluções em modo *saturado* são consideradas. Fazendo uso desta definição de soluções, o processo de análise mostra que, para qualquer condição inicial, existe uma única solução (3.17) podendo ser estendida para um intervalo infinito de tempo. O processo de análise, então, desenvolve um eficiente algoritmo de determinação de estados de equilíbrio estáveis e instáveis do sistema (3.17) e estabelece uma função de energia para o sistema (3.17).

O processo de síntese baseia-se em resultados obtidos no processo de análise para sistemas LSSM para desenvolver um sistema representado pela equação (3.17). Isto é feito através do armazenamento de um adequado conjunto de vetores, permitindo calcular T e I para (3.17) de modo que:

- 1) Os vetores armazenados sejam estados de equilíbrio assintoticamente estáveis do sistema;
- 2) O sistema não tenha soluções periódicas;
- 3) O número de estados de equilíbrio espúrios do sistema seja o menor possível;
- 4) O domínio de atração de cada vetor [41] seja tão amplo quanto possível.

A seguir, é feita uma breve descrição do processo de síntese, para demonstrar o modo como são calculados novos pesos aplicados à rede de Hopfield.

3.4.2 - SÍNTESE DOS SISTEMAS LSSM

O processo de síntese descrito nesta subseção é baseado em teoremas elaborados no processo de análise descrito em [33], e visa a partir de um adequado conjunto de vetores, encontrar uma matriz $\mathbf{T} = [\mathbf{T}_{ij}]$ $n \times n$ e um vetor $\mathbf{I} = (I_1, \dots, I_n)^T$, conforme relatado em 3.4.1.

Supondo que temos m vetores $\mathbf{a}_1, \dots, \mathbf{a}_m$ definidos em $\mathbf{B}^n = \{\mathbf{x} = (x_1, \dots, x_n) \in \mathbf{R}^n : x_i = 1 \text{ ou } -1, i = 1, \dots, n\}$, representando m imagens dos numerais usados para treinamento da rede de Hopfield, o seguinte algoritmo deve ser usado para garantir que esses vetores sejam estados de equilíbrio estáveis de um sistema n dimensional dado por (3.17):

- 1) Construir uma matriz \mathbf{Y} de dimensão $n \times (m-1)$ da seguinte maneira:

$$\mathbf{Y} = [\mathbf{a}_1 - \mathbf{a}_m, \dots, \mathbf{a}_{m-1} - \mathbf{a}_m]. \quad (3.18)$$

- 2) Proceder a decomposição singular de \mathbf{Y} obtendo as matrizes \mathbf{U}, \mathbf{V} e Σ tal que

$\mathbf{Y} = \mathbf{U}\Sigma\mathbf{V}^T$, onde \mathbf{U} e \mathbf{V} são matrizes unitárias e Σ é uma matriz diagonal com os autovalores de \mathbf{Y} em sua diagonal. Denotamos:

$$\mathbf{Y} = [y_1, \dots, y_{m-1}];$$

$$\mathbf{U} = [u_1, \dots, u_n].$$

Das propriedades da decomposição singular sabemos que:

$k = \text{rank de } \Sigma$;

$(u_1, \dots, u_k) = \text{uma base ortonormal que gera } (y_1, \dots, y_{m-1})$;

$(u_1, \dots, u_n) = \text{uma base ortonormal de } \mathbf{R}^n$.

Calcular:

$$\mathbf{T}^+ = [\mathbf{T}^+_{ij}] = \sum_{i=1}^k u_i u_i^T ; \quad (3.19)$$

$$\mathbf{T}^- = [\mathbf{T}^-_{ij}] = \sum_{i=k+1}^n u_i u_i^T . \quad (3.20)$$

4) Obter \mathbf{T} e \mathbf{I} do seguinte modo:

$$\mathbf{T} = \mathbf{T}^+ - \tau \mathbf{T}^- \quad \text{e} \quad \mathbf{I} = \alpha_m - \mathbf{T} \alpha_m . \quad (3.21)$$

Onde, $\tau \in \mathbf{R}^n$, é um parâmetro escolhido, que se suficientemente grande, o conjunto de estados de equilíbrio estáveis contido em \mathbf{B}^n é aproximadamente $\mathbf{L}_a \cap \mathbf{B}^n$ [33]. \mathbf{L}_a é o subespaço linear de \mathbf{R}^n gerado por $\mathbf{a}_1, \dots, \mathbf{a}_m$.

O sistema então, pode ser simulado pela equação de diferenças dada por:

$$x((k+1)h) = F(\Phi x(kh) + \Gamma), \quad (3.22)$$

com restrição:

$$-1 \leq x_i \leq 1, \quad 1 \leq i \leq n,$$

onde:

$$\mathbf{x}(kh) \in \mathbf{D}^n;$$

h é o período de amostragem de x ;

F é uma função tal que:

$$F: \mathbf{R}^n \rightarrow \mathbf{D}^n, \quad F(\mathbf{x}) = (F_1(x_1), \dots, F_n(x_n))^T \text{ com:}$$

$$F(\rho) = \begin{cases} 1, & \text{se } \rho > 1 \\ 0, & \text{se } -1 \leq \rho \leq 1 \\ -1, & \text{se } \rho < -1 \end{cases} \quad (3.23)$$

e

$$\Phi = \exp(h\mathbf{T}), \quad (3.24)$$

$$\Gamma = \int_0^h \exp(\rho \mathbf{T}) d\rho \cdot \mathbf{I}. \quad (3.25)$$

Φ e Γ são aplicados diretamente ao modelo da rede de Hopfield, representando sua nova matriz de pesos e vetor de entradas externas a rede, respectivamente. Considerando os resultados demonstrados no processo de análise dos sistemas LSSM [33] e no procedimento de síntese em [33], um número maior de vetores podem ser empregados para o cálculo da matriz Φ do que o permitido para diminuir o efeito do termo de *cross talk*. Em outras palavras, pode-se treinar a rede de Hopfield com um maior número de padrões e garantir um menor número de estados espúrios, melhorando a performance de reconhecimento da rede. Os resultados experimentais obtidos usando Φ e Γ , na rede de Hopfield, são apresentados no capítulo 8 desta dissertação, juntamente com os resultados do método da projeção discutido na subseção 3.3.5.

3.5 - ALGUMAS CONSIDERAÇÕES FINAIS

Nesta seção foi descrito o procedimento de síntese dos sistemas LSSM e as razões de ser aplicado este procedimento diretamente a rede de Hopfield no cálculo de uma nova matriz de pesos. Concluí-se, assim, a apresentação do modelo de rede neural aplicado ao sistema de reconhecimento de numerais manuscritos tratado por esta dissertação. O próximo capítulo

trata os conceitos de morfologia matemática necessários ao entendimento das técnicas usadas para pré-processamento das imagens dos numerais, a última parte do capítulo é dedicada a uma descrição geral de nosso sistema de reconhecimento, antes de serem apresentadas cada uma de suas partes.

CAPÍTULO 4 - MORFOLOGIA MATEMÁTICA E APRESENTAÇÃO DO SISTEMA

4.1 - INTRODUÇÃO

Neste capítulo são descritos alguns conceitos básicos da morfologia matemática e o sistema de reconhecimento de numerais implementado neste trabalho. A apresentação de conceitos básicos da morfologia matemática visa facilitar a compreensão das técnicas morfológicas de afinamento, pruning e dilatação usadas no pré-processamento dos numerais a serem reconhecidos, descrito no capítulo 5. A segunda parte deste capítulo é dedicada a apresentação de nosso sistema de reconhecimento de numerais manuscritos, sob um aspecto geral, nosso objetivo é mostrar as fases que compõem o sistema.

4.2. MORFOLOGIA MATEMÁTICA

A morfologia matemática, elaborada inicialmente por Georges Matheron e Jean Serra, concentra seus esforços no estudo da estrutura geométrica das formas presentes em uma imagem. A morfologia matemática é considerada como parte do processamento e análise não linear de imagens [34], e tem seu formalismo baseado na teoria dos conjuntos, geometria integral e topologia, observando o fato de que uma imagem pode ser representada por um conjunto de pontos de dimensão qualquer. A morfologia matemática pode ser aplicada em várias áreas como visão robótica, controle de qualidade, metalografia, biomédica e reconhecimento de padrões, permitindo entre outros objetivos, processar imagens para realce, segmentação, detecção de bordas, esqueletonização, afinamentos, análise de formas e compressão de imagens.

O princípio básico da morfologia matemática é a comparação de objetos a serem analisados com um outro objeto de forma conhecida, denominado elemento estruturante. Devido as imagens utilizadas neste estudo serem binárias, são consideradas a seguir algumas definições básicas em morfologia matemática aplicadas ao processamento de imagens binárias, e explicado o conceito de transformações em “tudo ou nada” (hit or miss).

4.2.1 - ALGUMAS DEFINIÇÕES BÁSICAS DA MORFOLOGIA MATEMÁTICA

Sejam A, B e C conjuntos em \mathbf{R}^2 , com componentes $a = (a_1, a_2)$, $b = (b_1, b_2)$ e $c = (c_1, c_2)$, respectivamente. As seguintes operações básicas [35] são definidas para formular algumas operações morfológicas binárias, tais como, a operação de dilatação descrita no capítulo 5:

a) Translação

A translação do conjunto A por $x = (x_1, x_2)$, denotado $(A)_x$, é definido como:

$$(A)_x = \{y | y = a + x, \text{ para } a \in A\}.$$

A figura 4.1(b) mostra a translação do conjunto A apresentado na figura 4.1(a).

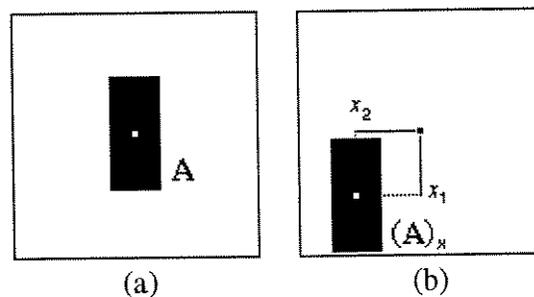


Fig. 4.1. Translação do conjunto A . (a) Conjunto A . (b) Conjunto A transladado por x . O ponto branco em cada uma das figuras indica a origem do conjunto.

b) Reflexão

A reflexão do conjunto C , denotado \hat{C} , é definido como:

$$\hat{C} = \{x \mid x = -c, \text{ para } c \in C\}.$$

A figura 4.2(b) mostra graficamente a reflexão do conjunto C mostrado na figura 4.2(a).

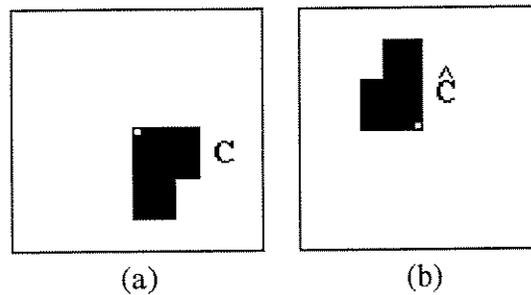


Fig. 4.2. Reflexão do conjunto C . (a) Conjunto C . (b) Conjunto C refletido. O ponto branco em cada uma das figuras indica a origem do conjunto.

c) Complemento

O conjunto que representa o complemento do conjunto A , apresentado pela figura 4.3(a), é mostrado graficamente pela figura 4.3(b). Matematicamente o complemento de A , denotado A^c , é definido como:

$$A^c = \{x \mid x \notin A\}.$$

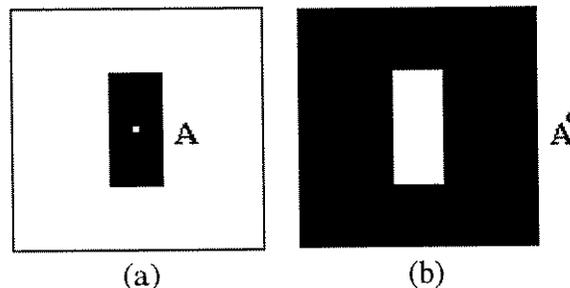


Fig. 4.3. Complemento do conjunto A . (a) Conjunto A . (b) Representação do complemento de A . O ponto branco na primeira figura indica a origem do conjunto.

d) União

A união de dois conjuntos **A** e **B**, denotada $A \cup B$, é definida como:

$$A \cup B = \{x | x \in A \text{ ou } x \in B\}.$$

A figura 4.4(c) apresenta a união dos conjuntos **A** e **B** representados pelas figuras 4.4(a) e 4.4(b), respectivamente.

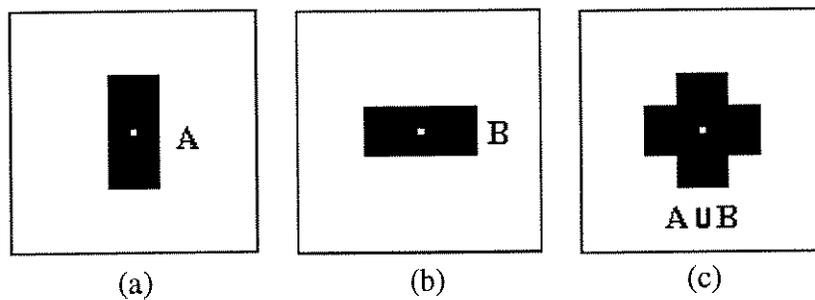


Fig. 4.4. União dos conjuntos **A** e **B**. (a) Conjunto **A**. (b) Conjunto **B**. (c) União de **A** e **B**. O ponto branco em cada uma das figuras indica a origem dos conjuntos.

e) Interseção

A interseção de dois conjuntos **A** e **B**, denotada $A \cap B$, é definida como:

$$A \cap B = \{x | x \in A \text{ e } x \in B\}.$$

A figura 4.5 apresenta a interseção de **A** e **B**.

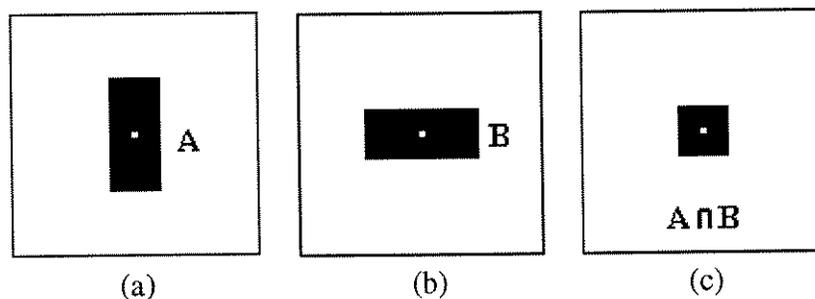


Fig. 4.5. Interseção dos conjuntos **A** e **B**. (a) Conjunto **A**. (b) Conjunto **B**. (c) Interseção de **A** e **B**. O ponto branco em cada uma das figuras indica a origem dos conjuntos

f) Diferença

A diferença entre os conjuntos **A** e **B** é mostrada pela figura 4.6, matematicamente esta diferença é definida como:

$$\mathbf{A - B} = \{x | x \in \mathbf{A}, x \notin \mathbf{B}\} = \mathbf{A} \cap \mathbf{B}^c.$$

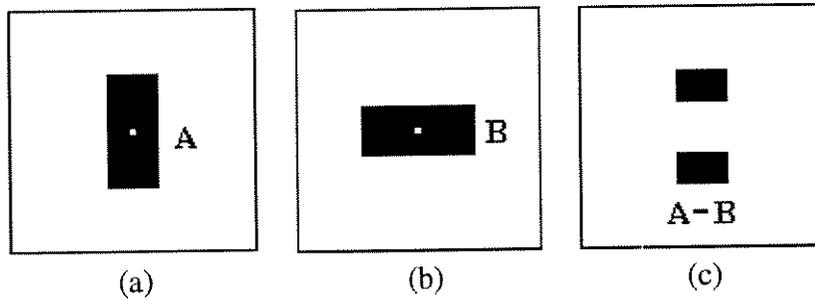


Fig. 4.6. Diferença entre os conjuntos **A** e **B**. (a) Conjunto **A**. (b) Conjunto **B**. (c) diferença entre **A** e **B**. O ponto branco em cada uma das duas primeiras figuras indica a origem dos conjuntos.

4.2.2 - TRANSFORMAÇÕES TUDO OU NADA POR UM ELEMENTO ESTRUTURANTE

Uma transformação “tudo ou nada” quando aplicada sobre um conjunto **K** de pontos a ser analisado, gera um novo conjunto de pontos que é capaz de caracterizar a estrutura do conjunto original. Note que esta informação estrutural não é representada por uma medida, mas sim, por transformações da própria imagem. A medida é definida como sendo o resultado da quantização da transformação resultante.

A transformação “tudo ou nada” sobre uma imagem **K**, consiste em deslocar um elemento estruturante **B**, de geometria conhecida, sobre **K** de tal forma que sua origem passe por todas as posições do espaço que constituem a imagem [34]. E em cada posição do espaço é feita uma pergunta em relação a união ou interseção de **B** com **K**, ou se **B** está contido em **K**. O conjunto dos pontos correspondentes a todas as respostas positivas forma um novo conjunto que constitui a imagem transformada. Abaixo, é mostrado um exemplo de

transformação “tudo ou nada”, onde é verificado se o elemento estruturante B da figura 4.7(a) está contido ou não na imagem K , representada pela figura 4.7(b), o resultado desta operação é mostrado na figura 4.7(c), que representa o conjunto de pontos que apresentam resposta positiva à questão do elemento estruturante B estar contido na imagem K .

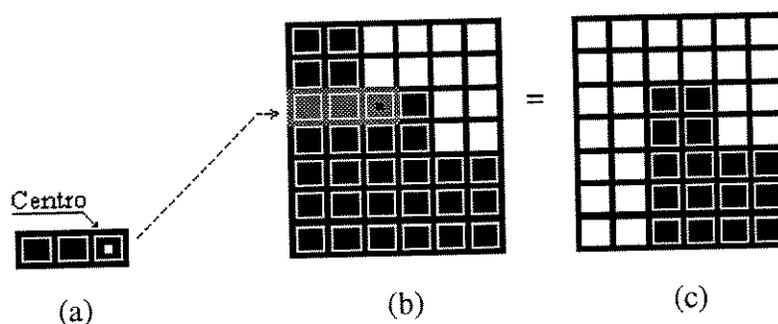


Fig. 4.7. Exemplo de uma transformação “tudo ou nada”. (a) Elemento Estruturante B . (b) Imagem K , sobre a qual é verificado se o elemento estruturante B está contido ou não. (c) Resultado da transformação tudo ou nada sobre a imagem K .

As operações de morfologia matemática aplicadas a este trabalho são afinamento e dilatação, descritos no capítulo 5 desta dissertação. Ainda, no capítulo 5, é descrita uma outra operação da morfologia matemática, denominada pruning, usada para eliminar pequenos segmentos de linha espúrios que surgem na imagem após seu afinamento.

4.3 DESCRIÇÃO GERAL DO SISTEMA DE RECONHECIMENTO DE NUMERAIS MANUSCRITOS

Nesta seção é feita uma descrição do funcionamento global do sistema de reconhecimento de numerais manuscritos. O sistema de reconhecimento de numerais manuscritos é formado pelas seguintes fases:

- Aquisição da imagem;
- Pré-processamento;

- Extração de características;
- Classificação via características extraídas (1º estágio de classificação);
- Classificação via rede neural (2º estágio de classificação).

A figura 4.8 mostra o diagrama de blocos do sistema de reconhecimento de numerais manuscritos.

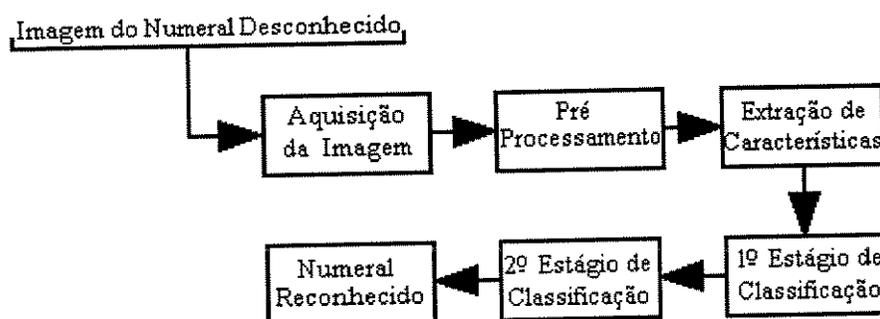


Fig. 4.8. Diagrama funcional do sistema de reconhecimento de numerais manuscritos.

A fase de aquisição da imagem começa com a digitalização via scanner dos numerais manuscritos. A imagem digitalizada é armazenada em arquivo no formato bitmap, para que o sistema localize as regiões nas quais os numerais foram escritos. Uma vez delimitada estas regiões, os numerais são separados e normalizados em escala. Cada numeral é armazenado em uma matriz binária de dimensão 20x19. O sistema passa, a partir deste ponto, para a fase de pré-processamento.

Na fase de pré-processamento da imagem as técnicas de afinamento, dilatação, e normalização de escala são aplicadas sobre a matriz binária que representa o numeral. O objetivo principal do pré-processamento é melhorar a imagem facilitando a extração de características.

Na fase de extração de características os seguintes atributos são extraídos de cada matriz de numeral:

1. Número de cavidades que formam o numeral e suas posições;
2. Número de interseções do numeral com um segmento de linha vertical que passa pelo seu centro de gravidade;
3. Número de interseções com um segmento de linha vertical que passa no ponto médio entre o centro de gravidade do numeral e a primeira coluna da matriz do numeral;
4. Sequências de cruzamento horizontal e vertical do caracter, que são definidas como sendo a quantidade de blocos de pixels pretos existentes em cada linha e em cada coluna da matriz que representa o caracter, respectivamente;
5. Atributos obtidos da distribuição pictorial da imagem dos numerais manuscritos, por exemplo, quantidade de pixels pretos existentes na primeira ou última linha da imagem do numeral, existência ou não de pixels pretos à direita ou à esquerda do segmento de reta que passa pelo centro de gravidade da imagem do numeral.

Após extraídos os atributos citados acima, o sistema inicia a fase de classificação, dividida em dois estágios: classificação via características extraídas e classificação via rede neural. No estágio de classificação via características extraídas são utilizados os atributos obtidos na fase de extração de características para reconhecer os numerais desconhecidos. Os numerais não reconhecidos neste estágio de classificação, passam a ser examinados no estágio de classificação via rede neural. Os numerais são então processados por 4 redes de Hopfield discretas utilizando o método da projeção (ver 3.3.5). Estas 4 redes são arranjadas em 2 blocos, sendo o primeiro bloco formado por apenas uma rede de Hopfield que tem a função de separação dos caracteres em 3 grupos segundo suas características. O segundo bloco é formado por 3 redes de Hopfield, cada uma processa quando solicitada, um dos três grupos

formados no primeiro bloco, respectivamente. A classificação via rede neural também foi implementada com a rede de Hopfield empregando o processo de síntese para sistemas LSSM (ver 3.4), o objetivo é melhorar o desempenho do sistema. Neste esquema de classificação via rede neural apenas uma rede faz a classificação dos numerais.

4.4 - ALGUMAS CONSIDERAÇÕES FINAIS

Todas as descrições feitas acima procuraram estabelecer uma idéia geral do funcionamento de nosso sistema de reconhecimento de numerais manuscritos. É importante ressaltar que até este momento foram apresentados todos os conceitos necessários a compreensão de nosso trabalho. Os capítulos que seguem, são dedicados a apresentação detalhada do sistema, apresentação dos resultados obtidos e conclusões.

CAPÍTULO 5 - AQUISIÇÃO E PRÉ-PROCESSAMENTO DE IMAGEM

5.1 - INTRODUÇÃO

Neste capítulo são apresentadas as fases de aquisição e pré-processamento de imagem, implementadas em nosso sistema de reconhecimento de numerais manuscritos. Tanto a aquisição como o pré-processamento de imagem representam etapas essenciais no reconhecimento de manuscritos. A aquisição é fase inicial em qualquer processo de reconhecimento de caracteres, envolvendo a digitalização da imagem do caracter para processamento computacional.

Uma imagem pode ser denotada por uma função contínua $f(x, y)$, cujo valor em qualquer ponto de coordenadas espaciais (x, y) é proporcional ao brilho da imagem naquele ponto [34-35]. As imagens observadas pelas pessoas em suas atividades visuais, normalmente consistem de luz refletida dos objetos. Basicamente $f(x, y)$ pode ser caracterizada por duas componentes: a quantidade de luz incidente na cena observada e a quantidade de luz refletida pelos objetos da cena observada. Apropriadamente, estas componentes são denominadas iluminância e reflectância, e são denotadas por $i(x, y)$ e $r(x, y)$, respectivamente. O produto da interação entre $i(x, y)$ e $r(x, y)$ forma a função $f(x, y)$.

$$f(x, y) = i(x, y) \cdot r(x, y). \quad (5.1)$$

A digitalização é a discretização de uma imagem contínua $f(x, y)$ em coordenadas espaciais (amostragem da imagem) e em brilho (quantização em níveis de cinza). Uma imagem

$f(x, y)$ pode ser, então, aproximada por amostras igualmente espaçadas, arranjadas na forma de uma matriz $N \times M$ como mostra a expressão (5.2).

$$f(x, y) \equiv \begin{bmatrix} f(0, 0) & f(0, 1) & \dots & f(0, M-1) \\ f(1, 0) & f(1, 1) & \dots & f(1, M-1) \\ \vdots & \vdots & \dots & \vdots \\ f(N-1, 0) & f(N-1, 1) & \dots & f(N-1, M-1) \end{bmatrix}. \quad (5.2)$$

O lado direito da expressão (5.2) representa uma imagem digitalizada. A posição de um elemento na matriz $N \times M$ identifica um ponto em $f(x, y)$, e o correspondente valor deste elemento especifica o nível de cinza naquele ponto em $f(x, y)$. Cada elemento da matriz $N \times M$ é denominado elemento da imagem ou pixel.

A fase de pré-processamento tem o objetivo de melhorar a qualidade de uma imagem digitalizada, corrigindo defeitos e imperfeições, que surgiram durante a fase de aquisição da imagem, causados pelo próprio processo de digitalização da imagem, objeto de escrita ou superfície de escrita. A imagem pré-processada deve possibilitar a escolha de características adequadas ao reconhecimento do carácter.

As seções 5.2 e 5.3 apresentam, respectivamente, o processo de aquisição e as técnicas de pré-processamento utilizadas em nosso sistema de reconhecimento de numerais.

5.2 - AQUISIÇÃO DE IMAGEM

Neste trabalho são considerados apenas numerais desconectados, escritos dentro de retângulos como mostrados na figura 5.1. Os numerais são digitalizados por um scanner óptico, que capta imagens em preto e branco com resolução de 180 pontos por polegada para facilitar processamentos a posteriori. A imagem digitalizada é armazenada em arquivo no formato de imagem bitmap. De fato, o formato bitmap permite fácil processamento e transferência de informações dentro do ambiente windows de programação, devido a disponibilidade de diversas funções necessárias ao carregamento de uma imagem em memória, leitura e obtenção de parâmetros, tais como, altura e largura em pixels de uma imagem e nível de cinza de um pixel em uma posição especificada.

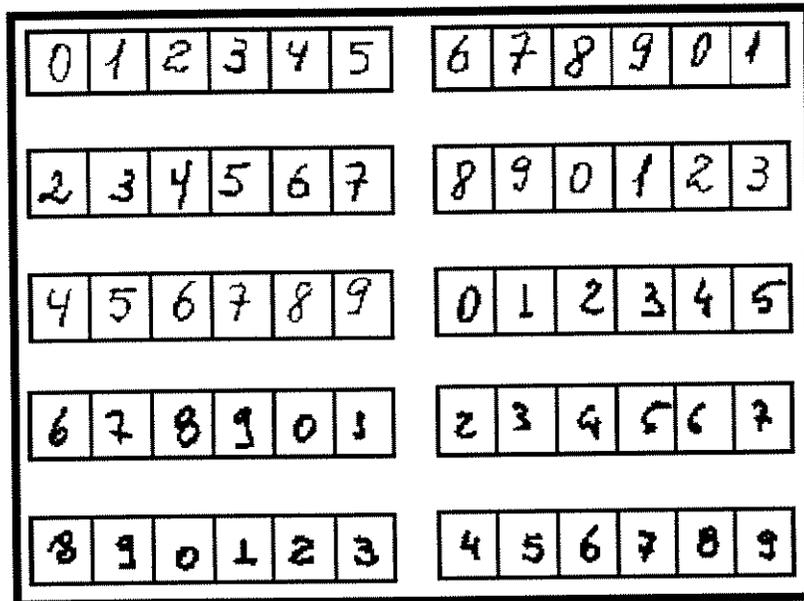


Fig. 5.1. Amostra dos numerais a serem digitalizados.

Durante a fase de aquisição de imagem, o sistema localiza as regiões nas quais os numerais foram escritos. Uma vez encontradas estas regiões, é determinado o menor retângulo

capaz de encobrir o caracter a ser reconhecido. As dimensões deste retângulo são calculadas pelas seguintes equações:

$$w_r = x_{\max} - x_{\min} ; \quad (5.3)$$

$$h_r = y_{\max} - y_{\min} . \quad (5.4)$$

Onde:

w_r = largura do retângulo;

x_{\max} = maior valor de coordenada x dos pixels de um caracter;

x_{\min} = menor valor de coordenada x dos pixels de um caracter;

h_r = altura do retângulo;

y_{\max} = maior valor de coordenada y dos pixels de um caracter;

y_{\min} = menor valor de coordenada y dos pixels de um caracter.

O procedimento de encontrar o menor retângulo capaz de encobrir o caracter é denominado enquadramento do caracter. A figura 5.2(b) mostra o enquadramento do numeral 8, originalmente apresentado pela figura 5.2(a).

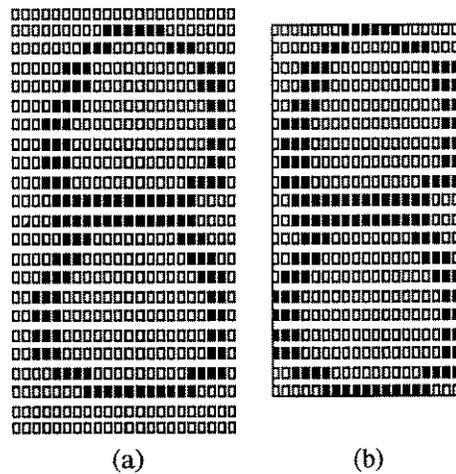


Fig. 5.2. Processo de enquadramento de um caracter. (a) Caracter a ser enquadrado. (b) Menor retângulo capaz de encobrir o caracter.

5.3 - PRÉ- PROCESSAMENTO DA IMAGEM

Esta fase, do sistema de reconhecimento, visa eliminar ruídos da imagem e melhorar sua informação pictorial. O sucesso em reconhecimento de caracteres manuscritos é intimamente ligado a qualidade da imagem usada. Em outras palavras, uma imagem de boa qualidade possibilita maior sucesso no reconhecimento. Nesta seção são descritas a técnica de normalização em escala e as técnicas da morfologia matemática de afinamento, pruning e dilatação, todas usadas no desenvolvimento desta fase do sistema.

5.3.1. NORMALIZAÇÃO EM ESCALA

Como relatado na seção 4.3, a fase de classificação dos numerais em seu segundo estágio, denominado classificação via rede neural, utiliza a rede de Hopfield discreta para reconhecimento dos numerais manuscritos. Estudos experimentais mostraram que este procedimento de reconhecimento, não apresenta bons resultados quando os numerais armazenados como padrões de treinamento, pela rede, apresentarem tamanhos diferentes em

relação aos numerais a serem reconhecidos. O mesmo acontece no primeiro estágio da fase de classificação, onde são comparadas as seqüências de cruzamento dos numerais a serem reconhecidos, com a seqüência de cruzamento de um padrão usado como referência. Para evitar que os numerais tenham tamanhos diferentes e obter melhores resultados, é necessário normalizar os numerais com relação a escala. Basicamente, isto é feito efetuando-se uma operação lógica “E” entre duas linhas ou colunas consecutivas do caracter. Em outras palavras, a ampliação ou redução da imagem baseia-se no seguinte princípio: o operador lógico “E” produz um pixel branco (nível de cinza 1), se todos os pixels envolvidos pelo operador forem brancos, caso contrário um pixel preto (nível de cinza igual a 0) é produzido. O efeito de ampliação ou redução da imagem do caracter é realizado diretamente por funções específicas do ambiente windows de programação. O procedimento de normalização em escala é concluído com a representação da imagem de cada caracter por uma matriz binária de dimensão 20x19.

5.3.2 AFINAMENTO E DILATAÇÃO DE CARACTERES

A técnica de afinamento seguida por uma dilatação, utilizada neste sistema de reconhecimento, visa obter uma mesma espessura de traço para caracteres, tornando-os independentes às variações de traços ocasionadas pelo tipo de caneta usada para escrever ou superfície de escrita. Caracteres com uma mesma espessura de traço são importantes na classificação via redes neurais, pois é feita a comparação entre os caracteres a serem reconhecidos e os padrões de treinamento armazenados pela rede. Testes mostraram que a classificação via redes neurais não apresenta bons resultados quando os padrões de



treinamento possuem espessuras de traço diferentes em relação aos numerais a serem reconhecidos

5.3.2.1 Afinamento

Deseja-se, com o processamento de afinamento, fazer com que todos os numerais tenham aproximadamente a espessura de traço de um pixel. Este procedimento é um modo de obter caracteres com uma mesma espessura de traço como discutido no item 5.3.2.

O afinamento de uma região, produz um esqueleto da região. O esqueleto de uma região pode ser definido através de uma transformação de eixo médio [35], da seguinte forma:

“Seja X uma região com bordas δX . Denomina-se x , um ponto pertencente a X . Para cada ponto x encontra-se seus vizinhos mais próximos em δX . Se x tem mais do que um vizinho, ele faz parte do esqueleto de X (eixo médio).”

A figura 5.3 mostra um exemplo do esqueleto (linhas tracejadas) de uma região X com bordas δX (linhas contínuas) quando a distância Euclidiana é usada.

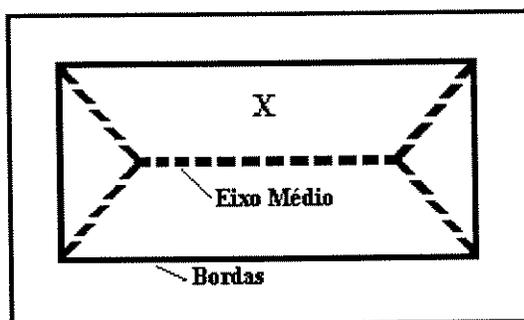


Fig. 5.3. Esqueleto da imagem X representado por linhas tracejadas.

Na literatura existem diversos algoritmos de afinamento que podem ser empregados para produzir o esqueleto do caracter. O algoritmo empregado neste trabalho é aplicado a imagens binárias [35] e utiliza a máscara de afinamento mostrada pela figura 5.4. O algoritmo assume que pixels pretos da imagem tem valor 1 e os pixels brancos valor 0, e consiste dos seguintes passos:

Primeiro passo

- a) Verificar se o número de vizinhos, denotado $N(P_1)$, do pixel $P_1 = 1$ obedece a:

$$2 \leq N(P_1) \leq 6; \quad (5.5)$$

- b) Verificar se o número de transições, $S(P_1)$, do valor 0 para o valor 1 na seqüência de pixels $P_1, P_2, P_3, P_4, P_5, P_6, P_7, P_8, P_9, P_{10}$, é igual 1;
- c) Verificar se o produto $P_2 \cdot P_4 \cdot P_6 = 0$;
- d) Verificar se o produto $P_4 \cdot P_6 \cdot P_8 = 0$.

Segundo passo

- a) Repetir os itens (a) e (b) do primeiro passo;
- b) Verificar se o produto $P_2 \cdot P_4 \cdot P_8 = 0$;
- c) Verificar se o produto $P_2 \cdot P_6 \cdot P_8 = 0$.

Terceiro passo

Se uma ou mais condições do primeiro e segundo passo não forem satisfeitas, manter o valor do pixel P_1 . Se todas as condições forem satisfeitas o pixel P_1 é marcado para ter seu valor alterado para 0. O valor de P_1 só é alterado após toda a imagem ter sido processada.

Quarto passo

Repetir os passos anteriores até que não haja mais mudanças nos valores de pixels da imagem.

P ₉	P ₂	P ₃
P ₈	P ₁	P ₄
P ₇	P ₆	P ₅

Fig. 5.4. Máscara de afinamento.

Esta técnica de afinamento, ao obter o esqueleto do caracter, obedece as seguintes regras:

- 1- Não remove pontos extremos da imagem (end points);
- 2- Não causa eliminação excessiva dos pixels pretos da região;
- 3- Preserva a conectividade da imagem.

Desta forma, a técnica usada conserva a topologia e a conectividade da imagem, e permite que todos os caracteres fiquem com a espessura de um pixel. A figura 5.7(b) exemplifica o esqueleto do numeral 4 obtido a partir da figura 5.7(a). Infelizmente, qualquer processo de afinamento gera segmentos espúrios (ramos parasitas do esqueleto) que podem provocar erros na extração de características [34]. Um modo para minimizar este efeito é descrito a seguir.

5.3.2.2 Eliminação dos Ramos Parasitas do Esqueleto

O processo de afinamento usado para obter o esqueleto do numeral gera, no esqueleto do numeral, pequenos segmentos. Estes segmentos, denominados “pés de galinha” ou “ramos

parasitas”, surgem em geral nas extremidades do caracter após um processo de afinamento. Neste trabalho um processamento da morfologia matemática, denominado pruning [34], é empregado para eliminar ramos parasitas de caracter manuscrito já afinado. O processo consiste em aplicar sobre a imagem afinada os 7 elementos estruturantes, 3x3, mostrados na figura 5.5. Os pequenos retângulos pretos, nestes elementos estruturantes, podem representar tanto pixels pretos como pixels brancos, o número “1” representa os pixels pretos e o número “0” os pixels brancos. O número “1” do centro dos elementos estruturantes 3x3 marca o pixel preto a ser eliminado quando o elemento estruturante estiver contido na imagem. Ao contrário do processo de afinamento, que é aplicado sobre a imagem até que esta não sofra mais mudanças, o processo de pruning deve possuir um número pré-determinado de ciclos. Um processo de pruning com excessivo número de ciclos pode causar redução ou até destruição da imagem afinada. Baseado na análise dos esqueletos obtidos dos caracteres, neste trabalho, o número de ciclos do processo de pruning está fixado em 2. A figura 5.7(c) ilustra o resultado final do processo de pruning aplicado sobre o esqueleto do numeral 4 da figura 5.7(b).

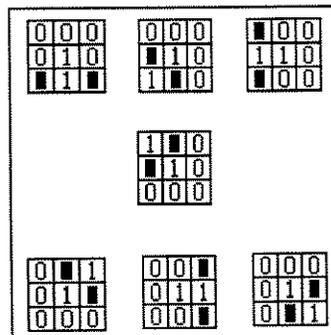


Fig. 5.5. Família de elementos estruturantes usados no processo de pruning.

5.3.2.3 Dilatação

A quantidade de pixels pretos que formam o numeral fica bastante reduzida após o processo de afinamento, o que não possibilita bons resultados no reconhecimento utilizando redes neurais, uma vez que os pixels pretos caracterizam a imagem a ser processada. A quantidade de pixels pretos fica ainda menor com o emprego do procedimento de pruning. Assim, é necessário aumentar a quantidade de pixels pretos da imagem através de um processo de dilatação. O procedimento de dilatação aplicado aqui, consiste em aumentar a espessura do caracter, de modo a se ter uma espessura de traço de 3 pixels.

A dilatação de uma imagem A por um elemento estruturante B , sendo A e B conjuntos em \mathbf{R}^2 , \emptyset denotando o conjunto vazio e $x = (x_1, x_2)$ um ponto em \mathbf{R}^2 , pode ser definida como:

$$A \oplus B = \{x \mid (\hat{B})_x \cap A \neq \emptyset\}. \quad (5.6)$$

Onde:

\hat{B} : é a reflexão do conjunto B , definida no item 4.2.1;

$(\hat{B})_x$: é a translação de \hat{B} por x , definida no item 4.2.1.

Deste modo, um processo de dilatação consiste em obter a reflexão de B sobre sua origem, e então, deslocar esta reflexão por x . A dilatação de A por B é então, o conjunto de todos os deslocamentos “ x ” que produzam a sobreposição de \hat{B} e A por no mínimo um elemento.

A figura 5.6 representa os elementos estruturantes usados para dilatar os numerais a serem reconhecidos. Logo após esta dilatação, os numerais são novamente enquadrados e

normalizados em escala, como foi descrito nas seções 5.2 e 5.3, respectivamente. Isto, porque os processos de afinamento e pruning reduzem as dimensões dos caracteres, como ilustram as figuras 5.7(b) e 5.7(c). A figura 5.7(d) mostra o resultado final do processo de dilatação seguido de enquadramento e normalização em escala aplicado a imagem da figura 5.7(c).



Fig. 5.6. Elementos estruturantes com centro em 1, usados no processo de dilatação do numeral. (a) Elemento estruturante que dilata o numeral afinado por um pixel à direita. (b) Elemento estruturante que dilata o numeral afinado por um pixel à esquerda.

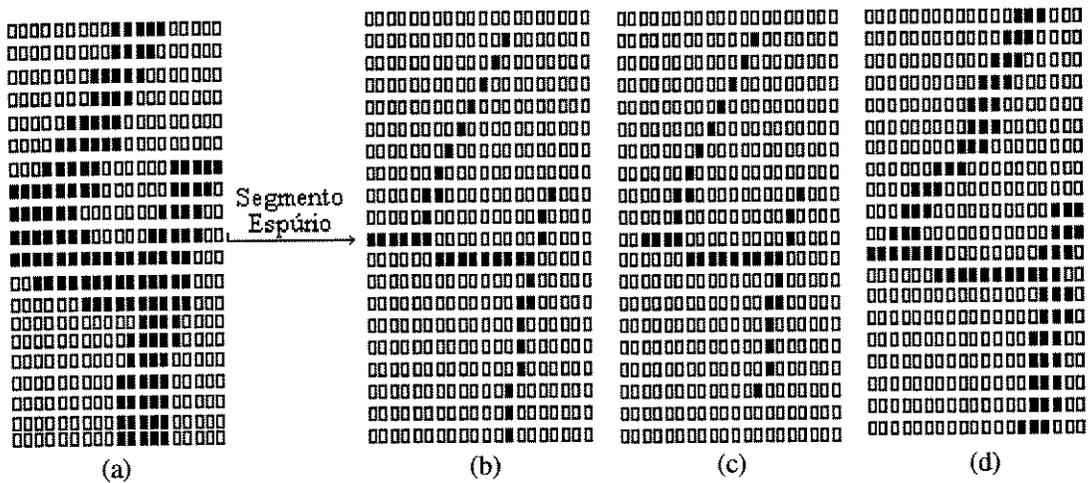


Fig. 5.7. Fases de pré-processamento para numeral 4. (a) Imagem original. (b) Esqueleto do numeral 4 após afinamento, com a presença de ramo parasita. (c) Esqueleto do numeral 4 após o processo de pruning. (d) Numeral 4 dilatado.

Com a normalização em escala, os caracteres passam a ser representados por uma matriz binária de dimensão 20x19, onde, a seguinte convenção é adotada: pixels brancos são representados pelo número “0” e pixels pretos pelo número “1”. A partir deste ponto a imagem do caracter está na forma adequada para a extração de características, processo descrito pelo capítulo 6.

CAPÍTULO 6 - EXTRAÇÃO DE CARACTERÍSTICAS

6.1 - INTRODUÇÃO

No processo de extração de características, tenta-se obter do padrão a ser reconhecido um conjunto de atributos próprios, possibilitando a representação e um bom grau de discriminação do padrão a ser reconhecido em relação a outros padrões.

O conjunto de atributos usados por nosso sistema de reconhecimento consiste de dois grupos. O primeiro grupo de atributos envolve oito parâmetros, relacionados a seguir:

1. Número de cavidades centrais do numeral;
2. Número de cavidades à direita;
3. Número de cavidades à esquerda;
4. Posição de cada cavidade no numeral, parte superior ou inferior;
5. Seqüência de cruzamento do numeral;
6. Número de interseções com o eixo principal do caracter;
7. Número de interseções com o eixo secundário do caracter;
8. Posição das interseções no numeral, parte superior ou inferior.

O segundo grupo de atributos é a própria imagem do numeral, após ser pré-processada pelo sistema.

São usados dois grupos de atributos, devido a fase de classificação ser dividida em dois estágios denominados: classificação via características extraídas e classificação via rede neural, os quais utilizam, respectivamente, o primeiro e o segundo grupo de características extraídas. A seguir, são detalhados os métodos usados para obtenção do primeiro grupo de atributos.

6.2 - CAVIDADES DA IMAGEM DOS NUMERAIS

Os parâmetros 1, 2, 3 e 4, do primeiro grupo de atributos, são obtidos através da contagem e identificação da posição das cavidades existentes em cada numeral a ser reconhecido. Cada numeral é representado por uma matriz binária de dimensão 20×19 , que também representa o menor retângulo encobrindo o numeral. Este retângulo divide o fundo da imagem em regiões, as quais foram denominadas de cavidades do carácter. A figura 6.1 mostra um exemplo destas regiões que constituem o fundo da imagem do numeral 6.

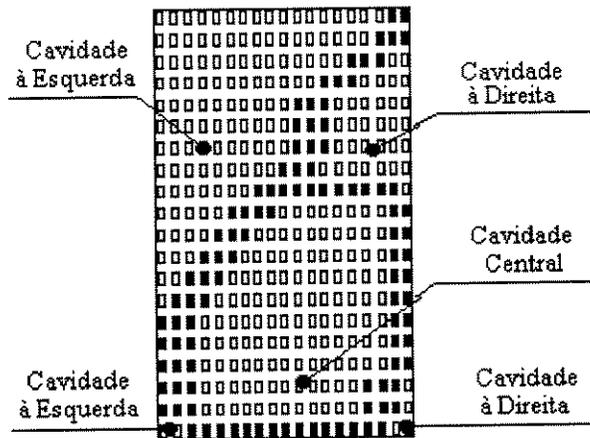


Fig. 6.1. Cavidades que constituem o fundo da imagem do numeral 6.

Para descrever o processo que realiza a contagem do número de cavidades da imagem de um carácter, é necessário que se faça as seguintes considerações:

- a) Cada pixel que forma o fundo da imagem possui uma vizinhança 4 conectada, ou seja, cada pixel do fundo da imagem tem 4 pixels vizinhos, 2 vizinhos na horizontal (um pixel à direita e outro à esquerda) e dois vizinhos na vertical (um pixel na parte superior e outro na parte inferior);

- b) A imagem processada é binária (pixels brancos são representados pelo número “1” e pixels pretos pelo número “0”, como mostra a figura 6.2(b));
- c) Denota-se p o pixel com valor 1 analisado em um momento qualquer do procedimento;
- d) Denota-se r o pixel vizinho de p localizado acima de p ;
- e) Denota-se t o pixel vizinho de p localizado abaixo de p ;
- f) Atribui-se um rótulo para cada pixel p analisado pelo procedimento. Um rótulo não representa o valor do pixel na imagem processada, pode ser por exemplo, um número ou uma letra do alfabeto usados para identificar os pixels que são conectados.

Estabelecidas as precedentes considerações, os seguintes procedimentos são realizados pelo algoritmo na contagem do número de cavidades de cada caracter [35]:

- 1) Varre-se a imagem pixel a pixel da esquerda para a direita, de cima para baixo;
- 2) Para cada pixel p a ser analisado verifica-se os valores dos pixels r e t :
 - a) Se $r = t = 0$, assinala-se um novo rótulo em p ;
 - b) Se $r = 1$ e $t = 0$, assinala-se o rótulo de r em p ;
 - c) Se $r = 0$ e $t = 1$, assinala-se o rótulo de t em p ;
 - d) Se $r = t = 1$ e seus rótulos são iguais, assinala-se este rótulo em p ;
 - e) Se $r = t = 1$ e seus rótulos são diferentes, assinala-se um dos rótulos em p . Deste modo, os pixels r e t têm rótulos equivalentes e estão conectados por p .

- 3) No final da varredura sobre a imagem, todos os pixels com valor 1 estão rotulados. Note que alguns destes rótulos podem ser equivalentes. É necessário, então, substituir os rótulos que se equivalem por um único rótulo.

A figura 6.2 abaixo mostra um exemplo de aplicação do algoritmo acima. Pode-se observar pela figura 6.2(c) que a imagem do caracter passa a ser representada por uma matriz de dimensão 20×19 , onde os elementos com rótulos iguais (um número maior do que zero identificando elementos conexos formadores de uma cavidade) pertencem a uma mesma cavidade. Os elementos com valor zero representam o caracter propriamente dito.

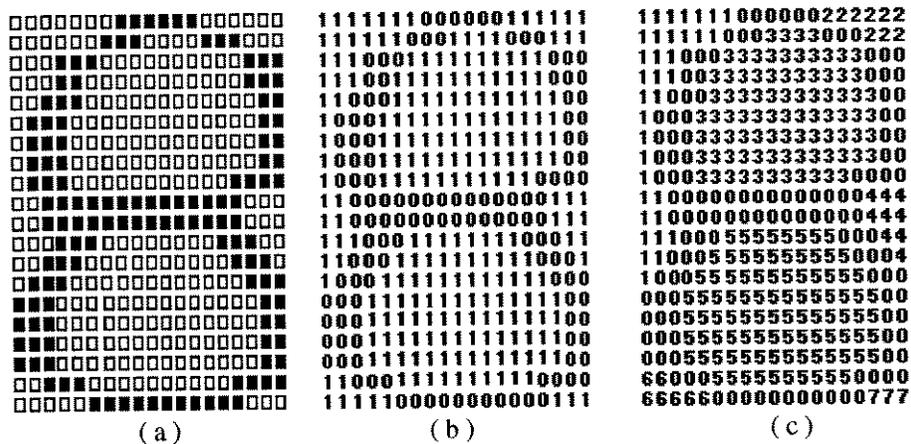


Fig. 6.2. Exemplo da aplicação do algoritmo para contagem de cavidades. (a) Numeral 8 original, os pixels pretos que formam o numeral têm valor 1 e os pixels brancos que formam o fundo da imagem têm valor 0. (b) A convenção é mudada para a aplicação do algoritmo, os pixels brancos agora têm valor 1 e os pixels pretos valor zero. (c) Numeral 8 após a contagem das cavidades.

A partir da matriz M_r representante de uma imagem rotulada, como exemplificado pela figura 6.2(c), os seguintes parâmetros são extraídos:

1. Número de cavidades centrais do numeral.

São denominadas cavidades centrais, as cavidades cujos pixels não pertençam a primeira e a última linha e a primeira e a última coluna da matriz M_r . Por exemplo, para a figura 6.2(c) temos 2 cavidades centrais com rótulos “3” e “5”, respectivamente.

2. Número de cavidades à direita.

Cavidades à direita são aquelas cujos pixels pertencem a última coluna da matriz M_r . Para a figura 6.2(c) as cavidades à direita estão rotuladas pelos números 2, 4 e 7.

3. Número de cavidades à esquerda.

As cavidades à esquerda, são aquelas cujos pixels pertencem a primeira coluna da matriz M_r , para o caso do exemplo apresentado pela figura 6.2(c), existem duas cavidades à esquerda com rótulos “1” e “6”, respectivamente.

4. Posição de cada cavidade em relação a altura do caracter, parte superior ou parte inferior do numeral.

Uma cavidade está localizada na parte superior do numeral quando a maioria de seus pixels situam-se acima da décima primeira linha da matriz M_r , caso contrário dizemos que a cavidade está localizada na parte inferior. Para a figura 6.2(c), as cavidades situadas na parte superior do caracter são as regiões rotuladas pelos números 1, 2 e 3, e as cavidades situados na parte inferior do caracter são as regiões rotuladas pelos números 4, 5, 6 e 7.

6.3 - SEQÜÊNCIA DE CRUZAMENTO DE UM NUMERAL

A figura 6.3 mostra a seqüência de cruzamento horizontal e a seqüência de cruzamento vertical localizadas no lado direito e na base da imagem do numeral 6, respectivamente. Cada

elemento destas seqüências corresponde a quantidade de blocos de pixels pretos consecutivos existentes em uma linha ou uma coluna da matriz que representa a imagem do numeral [4]. Por exemplo, para a 10ª linha da matriz na figura 6.3, o número de blocos de pixels pretos consecutivos é igual a 2, para a 17ª coluna o número é 3. Na etapa de classificação via características extraídas, é calculada a distância absoluta entre as seqüências de cruzamento de um numeral a ser reconhecido e as seqüências de cruzamento de referência de cada dígito de 0 até 9, e classifica-se o numeral a ser reconhecido de acordo com a menor distância encontrada.

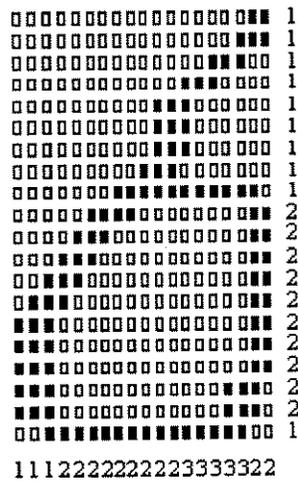


Fig. 6.3. Seqüências de cruzamento do numeral 6.

6.4. INTERSEÇÕES COM EIXO PRINCIPAL E EIXO SECUNDÁRIO DE UM NUMERAL

Define-se eixo principal de um numeral, o segmento de reta vertical que passa pelo centro de gravidade do numeral, cuja ordenada y_{cg} é calculada pela equação abaixo.

$$y_{cg} = \frac{1}{P} \sum_x^N \sum_y^M f(x, y) \cdot y, \quad (6.1)$$

onde:

P = número total de pixels pretos (valor igual a 1) na imagem $f(x, y)$ do caracter, é dado por:

$$P = \sum_x^N \sum_y^M f(x, y), \quad (6.2)$$

$f(x, y)$, especifica o valor do pixel em um ponto de coordenada (x, y) da imagem do caracter.

O eixo secundário é o segmento de reta vertical que passa no ponto médio entre a primeira coluna da matriz do numeral e o eixo principal. A figura 6.4 mostra o eixo principal e o eixo secundário interceptando a imagem do numeral 6, onde o número de interseções com ambos eixos é dois.

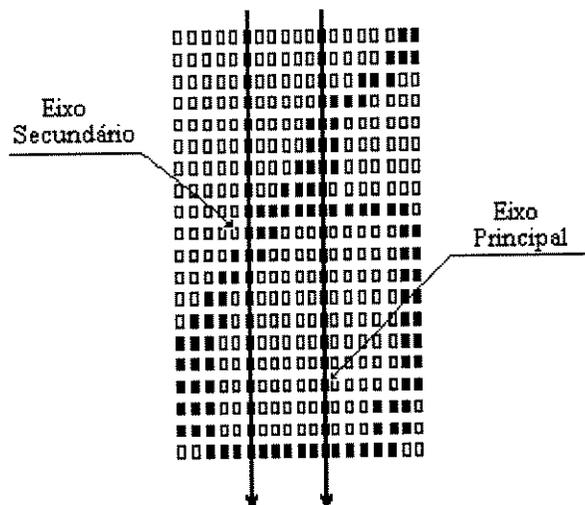


Fig. 6.4. Eixo principal e eixo secundário do numeral 6.

6.5 - DISTRIBUIÇÃO PICTORIAL

A característica denominada distribuição pictorial, diz respeito a uma quantidade ou disposição específica dos pixels pretos de uma imagem, que pode ser usada para decidir por uma determinada classe (uma classe representa um número de 0 a 9). Esta característica somente é obtida após se ter reduzido o número de classes nas quais pode ser classificado o numeral desconhecido, para 2 ou 3 classes, para evitar ambigüidade com as demais classes. A redução destas classes é feita através das características citadas anteriormente nas seções 6.2,

6.3 e 6.4. Como exemplo para distribuição pictorial, é citada a existência ou não de pixels pretos à direita ou à esquerda do eixo principal do caracter, a existência ou não de pixels pretos acima ou abaixo do ponto de interseção do caracter com o eixo principal e a quantidade de pixels pretos na primeira ou última linha da imagem do numeral. No próximo capítulo, onde será descrito o processo de classificação dos numerais, serão ilustrados vários exemplos da utilização da característica de distribuição pictorial.

CAPÍTULO 7 - CLASSIFICAÇÃO DE NUMERAIS MANUSCRITOS

7.1 - INTRODUÇÃO

Como mencionado na seção 4.3, o processo de classificação de caracteres é formado de dois estágios. O primeiro estágio denominado classificação via características extraídas e o segundo estágio denominado classificação via rede neural. O estágio de classificação via características extraídas utiliza os parâmetros do primeiro grupo de atributos, apresentado na seção 6.1, para efetuar o reconhecimento de numerais desconhecidos. Espera-se neste estágio classificar 80% do total dos numerais processados pelo sistema. Os restantes 20% dos numerais, devido a falhas ocasionadas pelo processo de digitalização ou no momento da escrita, e a grande variabilidade de forma dos caracteres manuscritos, não são classificados no primeiro estágio. Esta percentagem de caracteres não classificados são examinados no segundo estágio da fase de classificação, que utiliza a rede de Hopfield discreta como classificador. A rede de Hopfield para o segundo estágio de classificação é implementada de duas maneiras, a primeira maneira utiliza o conjunto de pesos calculado pelo método da projeção (ver 3.3.5) e a segunda maneira utiliza pesos calculados pelo processo de síntese para os sistemas LSSM (ver 3.4.2). São descritos a seguir, cada um dos estágios da fase de classificação dos numerais manuscritos.

7.2 - CLASSIFICAÇÃO VIA CARACTERÍSTICAS EXTRAÍDAS

O algoritmo usado neste estágio de classificação utiliza o primeiro grupo de atributos, obtidos pelo processo de extração de características descrito no capítulo 6, para classificar o numeral desconhecido dentre os números de 0 a 9.

A figura 7.1 mostra que, primeiramente, são analisados os numerais de acordo com o número de cavidades centrais. Assim, o diagrama do processo de classificação via características extraídas foi dividido em duas partes, sendo a primeira parte apresentada pela figura 7.1, que mostra o processo de classificação sob o aspecto da existência de cavidades centrais nos numerais desconhecidos. A segunda parte do diagrama é apresentada pela figura 7.8, onde é ilustrada a classificação dos numerais que não apresentam cavidades centrais. A seguir, é analisado o processo de classificação sob o ponto de vista da existência de cavidades centrais e, após, o processo de classificação para numerais que não apresentam cavidades centrais.

7.2.1 - CLASSIFICAÇÃO DE NUMERAIS OBSERVANDO A EXISTÊNCIA DE CAVIDADES

CENTRAIS

Analisando as várias formas que os numerais manuscritos podem apresentar, verifica-se que os numerais mais prováveis em apresentar 1 cavidade central são os numerais 0, 2, 4, 6 e 9. Por outro lado, os numerais mais prováveis em apresentar 2 cavidades centrais são os numerais 2 e 8. Assim sendo, conforme ilustrado na figura 7.1, os numerais são classificados de acordo com o número e a posição das cavidades centrais apresentadas, tendo como fator de decisão final a característica de distribuição pictorial ou suas seqüências de cruzamento. Os numerais que podem ser reconhecidos através das características representadas pelos losangos na figura 7.1, são apresentados pelos retângulos menores na parte inferior desta figura.

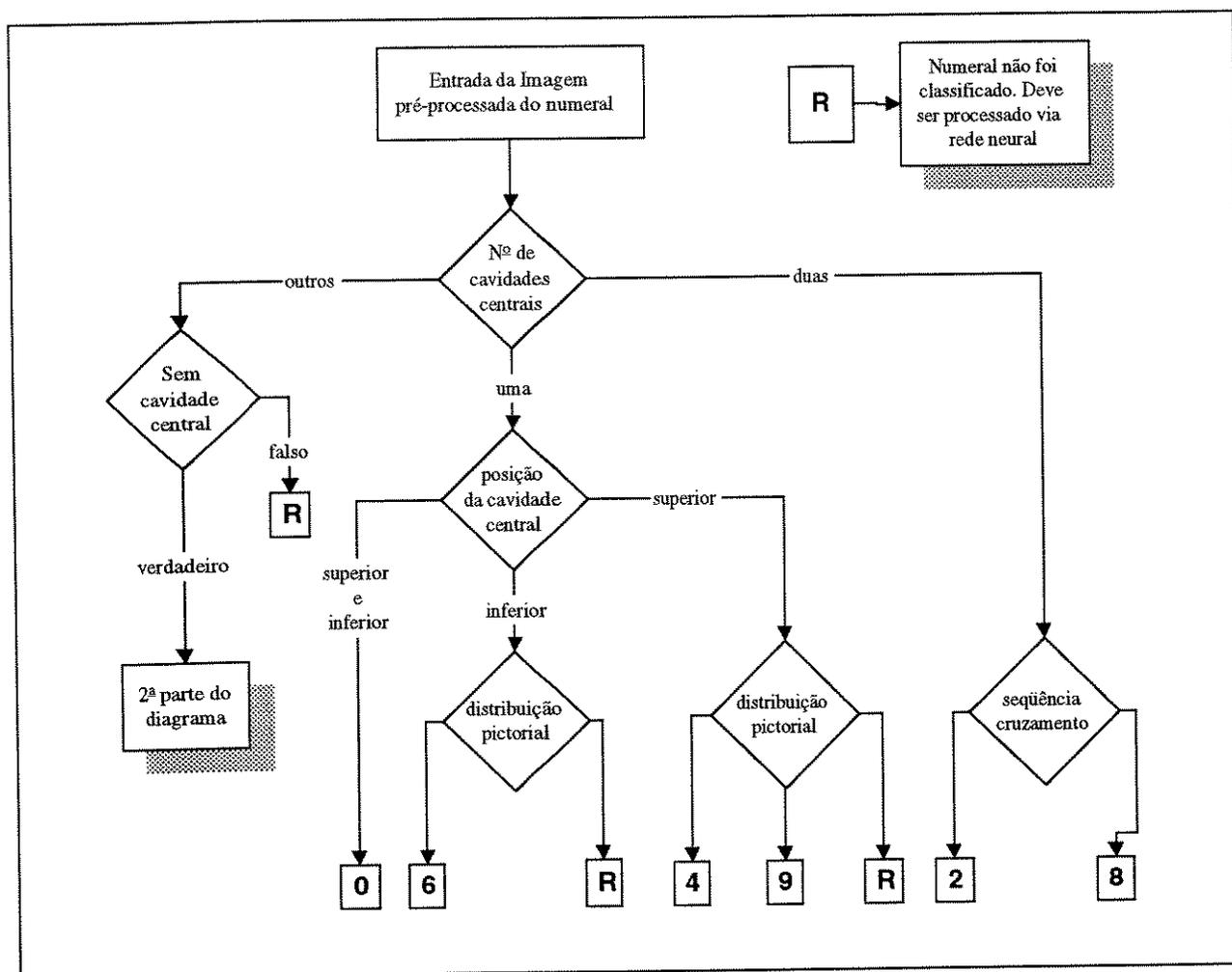


Fig. 7.1. Diagrama de classificação de numerais desconhecidos de acordo com suas cavidades centrais.

Para classificar os numerais, no caso da existência de 2 cavidades centrais, são utilizadas as seqüências de cruzamento da imagem, como explicado na seção 6.3. Para os numerais que apresentam apenas 1 cavidade central, primeiramente é analisada a posição em que se encontra esta cavidade no numeral:

- Se a cavidade se estender para parte superior e para parte inferior do caracter, o numeral é classificado como numeral 0;
- Se a cavidade ocupa a parte inferior do caracter, o numeral pode ser classificado como 2, 6 ou 8. Neste caso, o numeral desconhecido é classificado como 6, se verificada entre a 1ª e a

2ª interseção do caracter com seu eixo principal, a ausência de pixels pretos no lado direito do eixo principal, como ilustra a figura 7.2. De outro modo o numeral desconhecido será examinado pelo segundo estágio de classificação. Note que os numerais 2 e 8 possuem pixels pretos no lado direito do eixo principal, entre a 1ª e a 2ª interseção do caracter com seu eixo principal. A figura 7.3 ilustra exemplos desta característica.

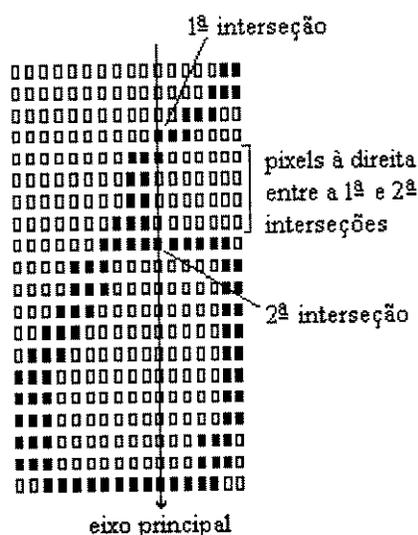


Fig. 7.2. Ausência de pixels pretos à direita do eixo principal entre a 1ª e 2ª interseções no numeral 6.

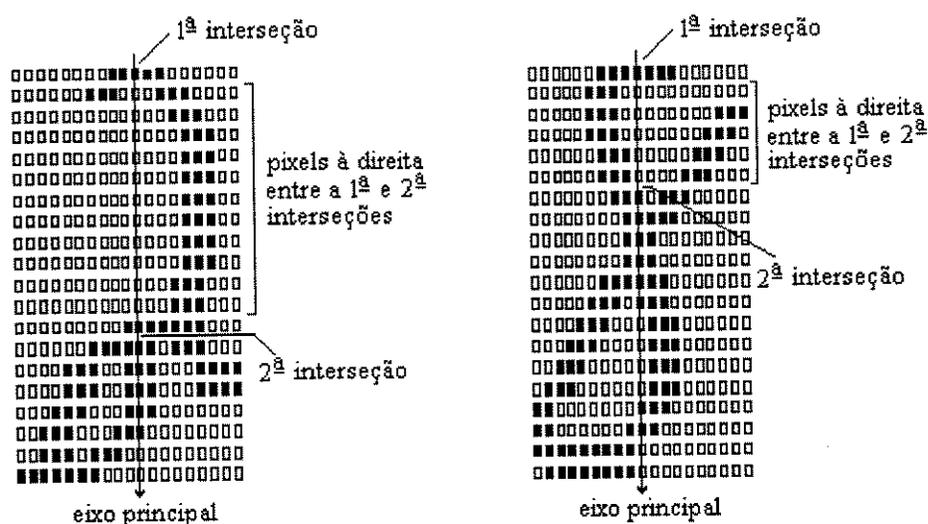


Fig. 7.3. Numerais 2 e 8 com pixels pretos à direita do eixo principal entre a 1ª e 2ª interseções.

- Se a cavidade encontra-se na parte superior do caracter, o numeral pode ser classificado como 4 ou 9. Distinguir o numeral 4 do numeral 9 não é uma tarefa simples, devido suas semelhanças. O processo para classificar os numerais 4 e 9 consiste dos seguintes passos:

1- Verifica-se o número de elementos com valor 1 (pixels pretos) existentes na última linha da matriz que representa o numeral:

- a) Se o número de pixels pretos é igual a 3 passa-se para o 2º passo;
- b) Se o número de pixels pretos é maior que 3 conclui-se o numeral desconhecido como numeral 9;

A figura 7.4 ilustra o procedimento do passo número 1.

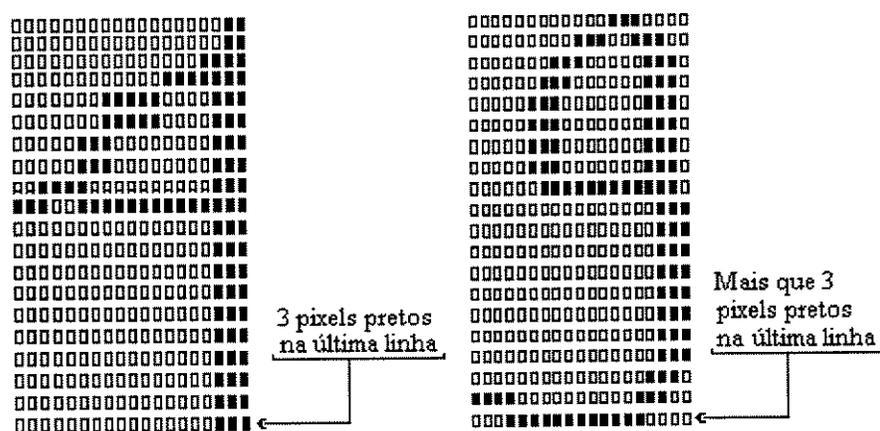


Fig. 7.4. Número de pixels pretos apresentados na última linha das matrizes representativas dos numerais 4 e 9. O numeral 4 deverá ter 3 pixels pretos na última linha.

2- Determina-se o segmento de reta exemplificado na figura 7.5. Este segmento é geralmente apresentado pelo numeral 4 possuindo uma cavidade central. Os seguintes pontos determinam o segmento de reta na matriz representativa do caracter:

- i) ponto do 1º elemento com valor 1(pixel preto), localizado na linha da 2ª interseção do caracter com seu eixo principal (fig. 7.5);

- ii) ponto de coordenadas do último elemento com valor 1 da 1ª linha da matriz do caracter (fig. 7.5).

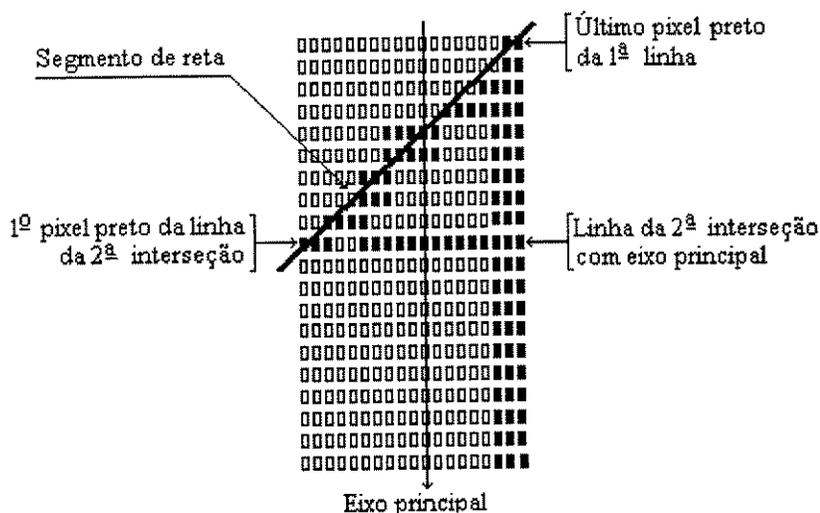


Fig. 7.5. Segmento de reta apresentado pelo numeral 4 possuindo uma cavidade central.

- 3- Calcula-se, para cada linha da matriz do caracter cortada pelo segmento de reta obtido no passo 2, a distância absoluta entre os pontos do segmento de reta e o pixel preto mais próximo do segmento de reta (caso houver mais de um pixel, somente um é escolhido);
- 4- O numeral é classificado como 4 se o somatório das distâncias calculadas no passo 3 for menor que 70% do número de pixels que formam o segmento de reta. De outra forma conclui-se o numeral desconhecido como numeral 9.

Ainda, deve-se observar que os numerais 2 e 8 podem ser classificados erroneamente como 4 ou 9. Isto pode acontecer ao numeral 2 se sua parte superior formar uma cavidade central e no caso do numeral 8 se sua cavidade central inferior for incompleta, conforme ilustrado na figura 7.6. Este problema pode ser evitado verificando-se a existência de pixels pretos no lado

esquerdo da matriz do caracter, em toda a extensão, entre a 2ª interseção do caracter com seu eixo principal e sua 18ª linha. Os numerais 4 e 9 em geral não apresentam pixels pretos à esquerda nestas posições, a figura 7.7 exemplifica estas características. Caso os numerais desconhecidos obedeam as características referentes aos numerais 4 e 9, os passos descritos acima são efetuados para classificar estes numerais, de outro modo os numerais são processados pelo 2º estágio de classificação.

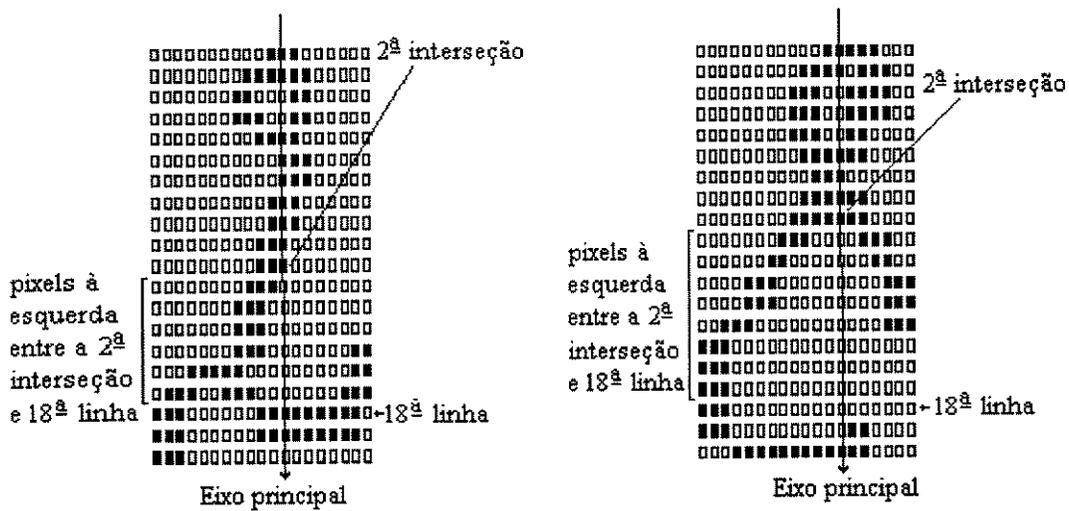


Fig. 7.6. Numerais 2 e 8 com cavidades centrais superiores e pixels pretos à esquerda do eixo principal entre a 2ª interseção e a 18ª linha das matrizes representativas destes numerais.

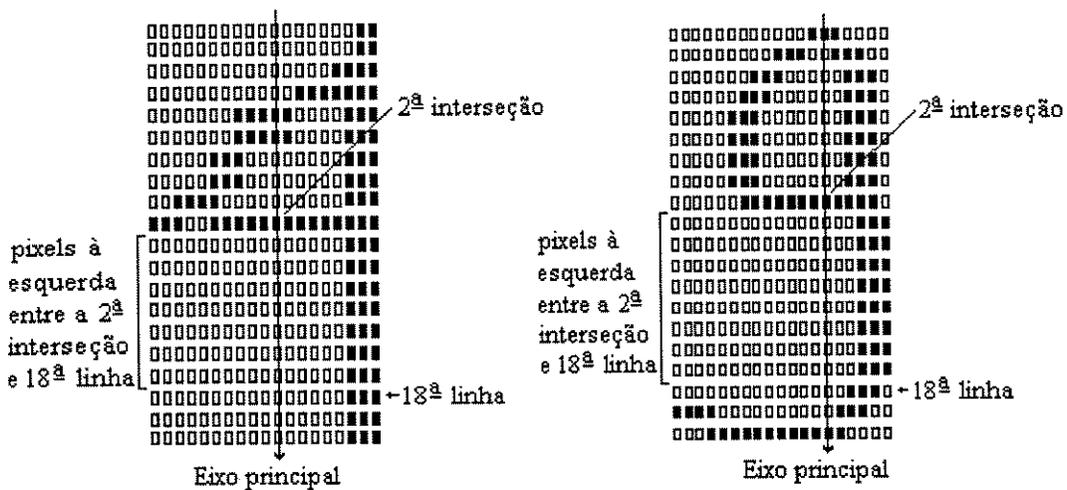


Fig. 7.7. Pixels pretos à esquerda dos numerais 4 e 9. Em geral, não há pixels pretos à esquerda do eixo principal, entre a 2ª interseção e a 18ª linha das matrizes representativas destes numerais.

7.2.2 - CLASSIFICAÇÃO DE NUMERAIS QUE NÃO APRESENTAM CAVIDADES CENTRAIS

Na seção anterior foram analisados os numerais que apresentam cavidades centrais. Nesta seção será mostrado como classificar os numerais que não apresentam esta característica, via um algoritmo ilustrado na figura 7.8. Inicialmente, estes numerais são divididos em dois grupos: numerais que apresentam 3 interseções com o eixo principal e numerais que possuem outro número qualquer de interseções com o eixo principal. São descritos abaixo estes dois grupos.

7.2.2.1 - Grupo de Numerais que Apresentam 3 Interseções com o Eixo Principal

Com base na análise das várias formas de numerais manuscritos, é observado que em situações normais os numerais 2, 3 e 5 apresentam três interseções com seu eixo principal. As seguintes condições são usadas para classificar estes numerais:

- 1- Existem pixels pretos à direita entre a 1^a e a 2^a interseção do numeral com seu eixo principal;
- 2- Existem pixels pretos à direita entre 2^a e 3^a interseção do numeral com seu eixo principal;
- 3- Existem pixels pretos à esquerda entre a 1^a e 2^a interseção do numeral com seu eixo principal;
- 4- Existem pixels pretos à esquerda entre a 2^a e 3^a interseção do numeral com seu eixo principal.

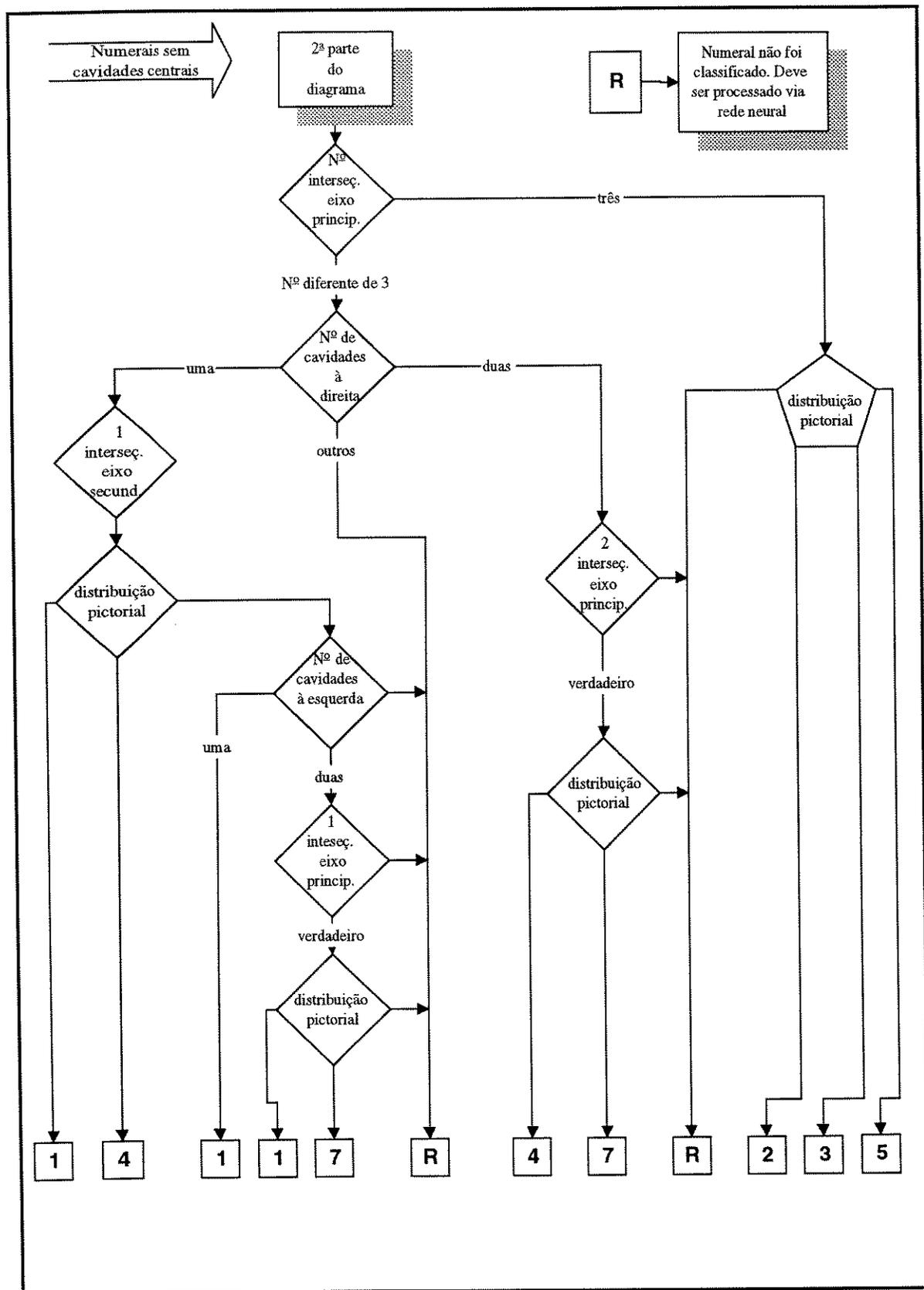


Fig. 7.8. Diagrama de classificação de numerais que não apresentam cavidades centrais.

O numeral é classificado como 5 quando somente as condições 2 e 3 são satisfeitas;

O numeral é classificado como 3 quando somente as condições 1 e 2 são satisfeitas;

O numeral é classificado como 2 quando somente as condições 1 e 4 são satisfeitas;

O numeral 7 pode ser classificado erroneamente como 3 quando a posição de seus pixels ocasionarem 3 interseções com seu eixo principal. Para evitar este erro de classificação, a distribuição pictorial do numeral desconhecido é analisada. Em geral, o numeral 7 deve possuir apenas 3 pixels em sua última linha, devido a espessura de traço obtida no processo de dilatação do numeral. O numeral 3 geralmente apresenta uma quantidade maior de pixels em sua última linha, conforme ilustrada na figura 7.9. Somente classificamos o numeral desconhecido, se ele possuir as características do numeral 3, caso contrário, o numeral não será classificado neste estágio da classificação.

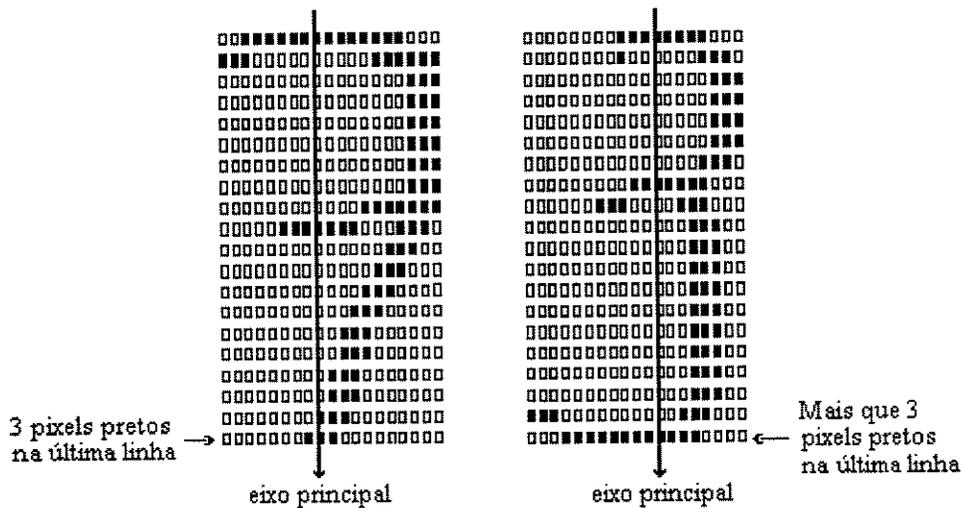


Fig. 7.9. Número de pixels pretos apresentados na última linha das matrizes representativas dos numerais 7 e 3. O numeral 7, em geral, deve apresentar 3 pixels pretos na última linha.

7.2.2.2 - Grupo de Numerais que Apresentam Outro Número de Interseções com o Eixo

Principal

O objetivo neste grupo é classificar os numerais como sendo 1, 4 ou 7. Para isto procurou-se recorrer as seguintes características de caracteres manuscritos:

- 1- Numerais com uma cavidade à direita;
- 2- Numerais com uma cavidade à direita e uma cavidade à esquerda;
- 3- Numerais com uma cavidade à direita e duas à esquerda;
- 4- Numerais com duas cavidades à direita.

A característica apresentada pelo item 1 é usada para classificar os numerais desconhecidos como sendo 1 ou 4. O numeral é classificado após ser verificada a existência de uma única interseção com o eixo secundário. Esta interseção deve estar localizada abaixo das três primeiras linhas do carácter. Se isto acontecer e, ao mesmo tempo, existirem pixels pretos acima e à esquerda da interseção, o numeral é classificado como 4, caso contrário tem-se o numeral 1. Por outro lado, se a interseção ocorrer em uma das três primeiras linhas, o número de cavidades à esquerda do numeral é analisado.

Os numerais desconhecidos que apresentarem a característica do item 2, são classificados como sendo 1. Estes numerais devem apresentar uma espessura de até 3 pixels por linha para serem classificados como numeral 1.

Os numerais desconhecidos que apresentarem a característica referenciada pelo item 3 podem ser classificados como sendo o numeral 1 ou 7. Para que isto seja possível, deve ocorrer apenas uma interseção com o eixo principal do carácter. O número de pixels por linha, após o ponto de interseção, deve ser de no máximo 3. O carácter é classificado como 7 se

existir o segmento de reta que caracteriza este numeral na sua parte superior. De outro modo, o numeral é classificado como sendo o numeral 1.

Os numerais que apresentarem a característica do item 4, podem ser classificados como sendo o numeral 4 ou 7. Esta classificação é realizada após verificada as seguintes situações:

- a) Ocorrência de somente duas interseções com o eixo principal do numeral analisado;
- b) Existência de pixels pretos acima e à esquerda da segunda interseção com o eixo principal;
- c) Existência de no máximo 3 pixels pretos em cada linha abaixo da segunda interseção com o eixo principal.

Para o numeral ser classificado como sendo 4, deve ocorrer a situação especificada no item b. Caso contrário o numeral é classificado como 7. Note que as condições impostas nos itens “a” e “c” devem ser obedecidas para ambos numerais, 4 ou 7.

7.3 - CLASSIFICAÇÃO VIA REDE NEURAL

Nesta seção será descrito o estágio de classificação via rede neural, sendo este o segundo e último estágio do processo de classificação de numerais manuscritos. O primeiro estágio de classificação, denominado classificação via características extraídas, foi relatado na seção anterior. O segundo estágio de classificação utiliza a rede de Hopfield discreta e é implementado de dois modos:

1º modo) Emprego do método da projeção para cálculo dos pesos da rede de Hopfield;

2º modo) Emprego do método de síntese para sistemas LSSM para cálculo dos pesos da rede de Hopfield.

A seguir é descrito o segundo estágio de classificação implementado pelo primeiro modo relacionado acima.

7.3.1 REDE DE HOPFIELD EMPREGANDO O MÉTODO DA PROJEÇÃO

Os numerais classificados no segundo estágio são somente aqueles que não foram classificados no primeiro estágio de classificação. O primeiro estágio procura classificar numerais desconhecidos através das propriedades obtidas na fase extração de características. Muitas vezes, devido a falhas na digitalização ou escrita de um numeral, não é possível classificá-lo através destas propriedades. Assim, o numeral é examinado no segundo estágio de classificação, onde a rede neural de Hopfield discreta é usada para realizar o reconhecimento dos caracteres.

Por ser uma rede autoassociativa [36], a rede de Hopfield é capaz de recuperar padrões armazenados como modelos, na fase de treinamento, a partir de padrões distorcidos apresentados à entrada da rede. Isto é possível, se os padrões distorcidos forem suficientemente similares a um dos padrões armazenados como modelos. A figura 7.10 mostra o diagrama de blocos deste estágio de classificação via rede neural, que também é dividido em dois passos. O primeiro passo, representado pelo bloco 1, tem a finalidade de reconhecer e agrupar os numerais em conjuntos de acordo com certas características comuns, por exemplo, segmentos de retas e curvas. Assim, um numeral é reconhecido e classificado dentro de um dos seguintes grupos:

- Grupo I {1, 4, 7};
- Grupo II {2, 3, 8, 9};
- Grupo III {0, 5, 6}.

Estes grupos foram formados com base em resultados experimentais utilizando a rede de Hopfield no reconhecimento de numerais manuscritos. Por exemplo, o numeral 7 é freqüentemente classificado como 1, devido ao segmento de linha vertical que estes dois numerais apresentam. Isto poderia ser evitado se a rede fosse treinada com um maior número de modelos, porém, como foi explicado no capítulo 3, há uma limitação para o número de modelos usados para treinamento da rede de Hopfield. Para contornar esta limitação, os numerais foram divididos em grupos como explicado acima, e então, é efetuado o segundo passo deste estágio de classificação, representado pelo bloco 2 da figura 7.10.

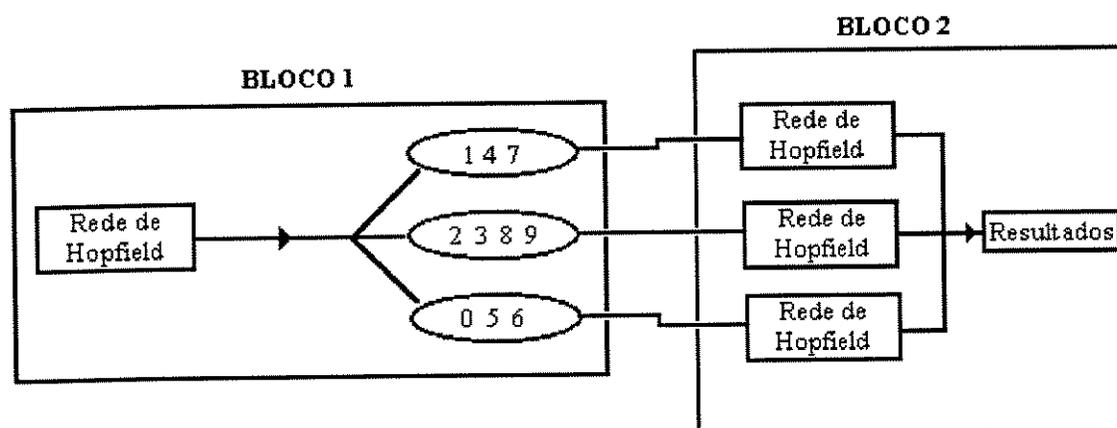


Fig. 7.10. Diagrama de classificação dos caracteres manuscritos via rede neural.

No bloco 2, três redes de Hopfield são utilizadas. Cada rede é treinada com modelos de um dos três grupos de numerais, {1, 4, 7}, {2, 3, 8, 9} e {0, 5, 6}. A quantidade reduzida de elementos por grupo, permite treinar cada rede com um maior número de variações de um

mesmo numeral, o que não acontece no bloco 1, onde a rede foi treinada com diferentes variações do conjunto de numerais de 0 até 9. Em outras palavras, no bloco 2 tem-se no máximo 4 classes de numerais por conjunto, por exemplo $\{2, 3, 8, 9\}$, o que nos possibilita armazenar um maior número de modelos com diferentes estilos de escrita para cada conjunto. Isto proporciona um melhor índice de classificações corretas.

A imagem de cada numeral desconhecido é apresentada à entrada da rede de Hopfield na forma de um vetor de 380 componentes. O reconhecimento do numeral se dá através de sucessivas interações, até que a rede recupere um numeral armazenado como modelo no seu período de treinamento. O reconhecimento é atingido quando um estado de estabilidade é alcançado, isto é, a imagem produzida na saída da rede permanece inalterada de interação a interação. Em muitos casos, a imagem recuperada pela rede não é exatamente idêntica ao modelo armazenado, como explicado na subseção 3.3.5. Ou seja, a rede pode ter atingido um ponto de estabilidade próximo daquele onde está localizado um numeral usado como modelo. Devido a isto, necessitamos de uma regra de decisão que efetue o mapeamento da imagem recuperada pela rede, para uma das imagens de numerais usadas como modelo, de acordo com certo critério de otimização. Para descrever esta regra de decisão, denota-se $\mathbf{G} = [\mathbf{G}_{ij}]$, uma matriz representativa da imagem recuperada pela rede. Uma matriz representando a imagem de um dos N numerais armazenados como modelos é denotada $\mathbf{F}^k = [\mathbf{F}_{ij}^k]$ com $k = 1, \dots, N$ e $D(\mathbf{G}, \mathbf{F}^k)$ é a distância entre as imagens \mathbf{G} e \mathbf{F}^k . A regra de decisão que classifica o numeral desconhecido dentre os numerais de 0 a 9 consiste dos seguintes passos:

1. Calcular a distância $D(\mathbf{G}, \mathbf{F}^k)$;
2. Achar k que produza menor $D(\mathbf{G}, \mathbf{F}^k)$;
3. Classificar \mathbf{G} na classe a que pertence a imagem de índice k .

A distância $D(\mathbf{G}, \mathbf{F}^k)$ é obtida da seguinte maneira:

I) Varrer \mathbf{G} e \mathbf{F}^k linha por linha;

II) Para cada linha i ($i = 1, \dots, I$):

- a) Encontrar para cada pixel preto $g(i, j) \in \mathbf{G}$, um pixel preto $f^k(i, m) \in \mathbf{F}^k$ de tal forma que $|j - m|$ seja mínimo. Matematicamente tem-se:

$$d_g(i, j) = \min_{\{m: 1 \leq m \leq J, f^k(i, m) = 1\}} |j - m|;$$

- b) Encontrar para cada pixel preto de $f^k(i, j) \in \mathbf{F}^k$, não envolvido nos cálculos do item b):

$$d_{f^k}(i, j) = \min_{\{m: 1 \leq m \leq J, g(i, m) = 1\}} |j - m|;$$

- c) Fazer $d_g(i, j) = 0$ ou $d_{f^k}(i, j) = 0$ para pixels brancos;

- d) Verificar os valores calculados de $d_g(i, j)$ e $d_{f^k}(i, j)$ e fazer:

- i) $d_g(i, j) = 0$ ou $d_{f^k}(i, j) = 0$ quando os valores calculados para estas distâncias forem menor do que 3;
- ii) $2 \cdot d_g(i, j)$ ou $2 \cdot d_{f^k}(i, j)$ quando os valores calculados para estas distâncias forem iguais a 5;
- iii) $3 \cdot d_g(i, j)$ ou $3 \cdot d_{f^k}(i, j)$ quando os valores calculados para estas distâncias forem maiores do que 5.

III) Calcular a distância $D(g, f_k)$ do seguinte modo:

$$D(g, f_k) = \sum_i \sum_j [d_g(i, j) + d_{f_k}(i, j)]$$

A precedente regra de decisão é baseada em resultados experimentais. Os valores de $d_g(i, j)$ e $d_{f_k}(i, j)$, no item d, evidenciam a distância entre duas imagens \mathbf{G} e \mathbf{F}^k quando não há similaridade entre ambas. Quando existe similaridade, os pixels pretos de ambas as imagens possuem aproximadamente as mesmas posições, deste modo, o valor calculado para $D(\mathbf{G}, \mathbf{F}^k)$ é pequeno. $D(\mathbf{G}, \mathbf{F}^k)$ mostra o quanto duas imagens diferem uma da outra, ou em outras palavras, quanto menor o valor de $D(\mathbf{G}, \mathbf{F}^k)$ mais similares são as imagens.

7.3.2 REDE DE HOPFIELD EMPREGANDO O PROCEDIMENTO DE SÍNTESE PARA SISTEMAS LSSM

A divisão do segundo estágio de classificação é somente adotado para a rede de Hopfield com pesos calculados pelo método da projeção, devido limitações citadas na subseção 3.3.5, referentes a capacidade de armazenamento da rede. Para a implementação do sistema de reconhecimento utilizando a rede de Hopfield com pesos calculados pelo processo de síntese para os sistemas LSSM (ver 3.4.2), não é feita a divisão do segundo estágio de classificação, é utilizada uma única rede de Hopfield treinada com diferentes variações dos numerais de 0 a 9. A mesma regra de decisão descrita na subseção anterior, para classificar a imagem produzida na saída da rede neural, é utilizada aqui. É esperado, com este procedimento para cálculo dos pesos da rede de Hopfield, obter um bom desempenho do sistema empregando uma única rede neural, treinada com o mesmos padrões da rede

apresentada pelo 1º bloco da figura 7.10, a qual é implementada segundo o método da projeção.

O capítulo 8 apresenta os resultados referentes aos 2 métodos de implementação da rede de Hopfield discreta, apresentados nesta seção.

CAPÍTULO 8 - RESULTADOS EXPERIMENTAIS E DISCUSSÃO FINAL

8.1 - INTRODUÇÃO

Este capítulo é dedicado a apresentação dos resultados dos testes experimentais do sistema de reconhecimento de numerais manuscritos, análise e conclusões do trabalho apresentado por esta dissertação, e a uma discussão sobre o que pode ser feito para melhorar a performance do sistema. A apresentação dos resultados obtidos refere-se ao sistema de reconhecimento implementado de dois modos. O primeiro modo utiliza, no segundo estágio de classificação, a rede de Hopfield discreta com pesos calculados pelo método da projeção (ver 3.3.5). O segundo modo emprega a rede de Hopfield, para o segundo estágio de classificação do sistema, com pesos calculados pelo processo de síntese para os sistemas LSSM (ver 3.4.2). Na última seção, finalizando este capítulo, são propostas algumas técnicas que podem melhorar a performance de nosso sistema, com base na análise dos resultados obtidos.

8.2 - RESULTADOS OBTIDOS

Como relatado no capítulo 7, a fase de classificação dos numerais manuscritos é dividida em dois estágios, sendo o primeiro estágio denominado classificação via características extraídas e o segundo, denominado classificação via rede neural. O segundo estágio de classificação utiliza a rede de Hopfield discreta, cujos pesos podem ser calculados pelo método da projeção ou pelo processo de síntese para os sistemas LSSM. Inicialmente, testou-se o sistema implementado com a rede de Hopfield com pesos calculados pelo método da projeção. Após, o sistema foi testado com a rede de Hopfield com pesos calculados pelo

processo de síntese para os sistemas LSSM. Os resultados obtidos através destes dois testes e o processo de treinamento da rede de Hopfield são apresentados a seguir.

8.2.1 - TESTE DO SISTEMA IMPLEMENTADO COM A REDE DE HOPFIELD COM PESOS CALCULADOS PELO MÉTODO DA PROJEÇÃO

Para este teste do sistema de reconhecimento utilizou-se 1500 numerais escritos por 80 pessoas. Destes 1500 numerais, 76% foram analisados no primeiro estágio de classificação e 24% foram analisados no segundo estágio de classificação.

Como explicado no capítulo 7, para a rede de Hopfield com pesos calculados pelo método da projeção, o segundo estágio de classificação foi dividido em dois blocos, um primeiro bloco formado por uma rede neural de Hopfield e um segundo bloco formado por três redes neurais de Hopfield. A rede neural do primeiro bloco foi treinada com 58 diferentes numerais de 0 a 9, enquanto que cada uma das redes neurais do segundo bloco utilizaram 60 numerais para treinamento, de acordo com os grupos especificados na seção 7.2. Note que os numerais usados para treinar as redes são diferentes daqueles usados para testar o sistema. O sistema classificou corretamente 85% dos numerais analisados, sendo 71% classificados pelo primeiro estágio de classificação e 14% classificados pelo segundo estágio de classificação. A tabela 8.1 especifica a quantidade total de padrões analisados e corretamente classificados pelo sistema para cada numeral de 0 até 9. A tabela 8.2 mostra a quantidade de padrões analisados e corretamente classificados somente no 1º estágio de classificação do sistema, a tabela 8.3 apresenta os padrões analisados e corretamente reconhecidos no 2º estágio do sistema, método da projeção.

Tabela 8.1 - Total de Padrões Analisados e Classificados pelo Sistema

NUMERAIS	TOTAL DE PADRÕES ANALISADOS PELO SISTEMA	PERCENTAGEM DE PADRÕES CORRETAMENTE CLASSIFICADOS PELO SISTEMA			
		Primeiro Modo		Segundo Modo	
		Total	%	Total	%
0	134	129	96%	131	98
1	158	113	71%	115	73
2	151	110	73%	106	70
3	175	152	87%	152	87
4	133	104	78%	100	75
5	166	160	96%	159	96
6	138	132	96%	129	93
7	151	125	83%	121	80
8	135	114	84%	111	82
9	159	143	90%	144	91
<i>Total</i>	1500	1282	85%	1268	84%

Tabela 8.2 - Total de Padrões Analisados e Classificados pelo 1º Estágio de Classificação

NUMERAIS	TOTAL DE PADRÕES ANALISADOS PELO 1º ESTÁGIO DO SISTEMA	PADRÕES CLASSIFICADOS PELO 1º ESTÁGIO DO SISTEMA			
		Corretos		Errados	
		Total	%	Total	%
0	116	114	98%	2	2%
1	85	70	82%	15	18%
2	77	68	88%	9	12%
3	146	137	94%	9	6%
4	102	84	82%	18	18%
5	147	144	98%	3	2%
6	121	119	98%	2	2%
7	109	100	92%	9	8%
8	101	97	96%	4	4%
9	140	132	94%	8	6%
<i>Total</i>	1144	1065	93%	79	7%
<i>% em 1500 numerais analisados</i>	76%	71%		5%	

**Tabela 8.3 - Padrões Analisados e Classificados pelo 2º Estágio de Classificação
Método da Projeção**

NUMERAIS	TOTAL DE PADRÕES ANALISADOS PELO 2º ESTÁGIO DO SISTEMA	PADRÕES CLASSIFICADOS PELO 2º ESTÁGIO DO SISTEMA - MÉTODO DA PROJEÇÃO			
		Corretos		Errados	
		Total	%	Total	%
0	18	15	84%	3	16%
1	73	43	59%	30	41%
2	74	42	57%	32	43%
3	29	15	52%	14	48%
4	31	20	65%	11	35%
5	19	16	84%	3	16%
6	17	13	75%	4	25%
7	42	25	60%	17	40%
8	34	17	50%	17	50%
9	19	11	58%	8	42%
<i>Total</i>	356	225	61%	139	39%
<i>% em 1500 numerais analisados</i>	24%	14%		10%	

8.2.2 - TESTE DO SISTEMA IMPLEMENTADO COM A REDE DE HOPFIELD COM PESOS CALCULADOS PELO PROCESSO DE SÍNTESE PARA SISTEMAS LSSM

Para este segundo teste do sistema foram utilizados os mesmos 1500 numerais do primeiro teste, como mostra a tabela 8.1. Portanto o primeiro estágio de classificação efetuou a análise de 76% do total dos numerais desconhecidos e o segundo estágio efetuou a análise de 24%.

Com os pesos da rede de Hopfield calculados através do processo de síntese para os sistemas LSSM, o segundo estágio de classificação utiliza uma única rede de Hopfield, diferente da técnica usada para o método da projeção, para a qual foram utilizadas 4 redes de Hopfield. A rede foi treinada com os 58 diferentes numerais de 0 a 9 empregados pelo primeiro teste do sistema, que utiliza o método da projeção no 2º estágio de classificação. O objetivo disto é fazer uma comparação de desempenho entre as redes neurais implementadas para o primeiro teste do sistema e a rede neural implementada para o segundo teste do sistema. O total de numerais corretamente reconhecidos pelo sistema foi de 84,4%, sendo 71% classificados pelo primeiro estágio de classificação e aproximadamente 13% classificados pelo

segundo estágio de classificação. A tabela 8.4 mostra a quantidade de padrões analisados e reconhecidos corretamente no 2º estágio de classificação do sistema, neste teste.

Tabela 8.4 - Padrões Analisados e Classificados pelo 2º Estágio de Classificação
Método de Síntese para Sistemas - LSSM.

NUMERAIS	TOTAL DE PADRÕES ANALISADOS PELO 2º ESTÁGIO DO SISTEMA	PADRÕES CLASSIFICADOS PELO 2º ESTÁGIO DO SISTEMA - MÉTODO PARA SISTEMAS LSSM			
		Corretos		Errados	
		Total	%	Total	%
0	18	17	95%	1	5%
1	73	45	62%	28	38%
2	74	38	52%	36	48%
3	29	15	50%	14	50%
4	31	16	52%	15	48%
5	19	15	79%	4	21%
6	17	10	59%	7	41%
7	42	21	50%	21	50%
8	34	14	42%	20	58%
9	19	12	63%	7	37%
<i>Total</i>	356	203	57%	153	43%
<i>% em 1500 numerais analisados</i>	24%	13%		11%	

8.2.3 - ANÁLISE DOS RESULTADOS OBTIDOS

Pode-se verificar, analisando a tabela 8.2, que os atributos usados pelo primeiro estágio de classificação do sistema (ver 6.1) para classificar os numerais desconhecidos proporcionaram um bom índice de reconhecimento para o sistema. Os erros de classificação neste estágio foram causados por falhas e elevações na forma do numeral, resultantes do processo de digitalização da imagem, e também pelos segmentos espúrios que surgiram no esqueleto do caracter devido ao processo de afinamento, os quais não foram totalmente eliminados pelo processo de pruning, sob pena de serem eliminados traços importantes do contorno do numeral.

As tabelas 8.3 e 8.4 apresentam, respectivamente, os resultados obtidos pela rede de Hopfield discreta utilizando o método da projeção e os resultados para a rede de Hopfield utilizando o processo de síntese para os sistemas LSSM. A percentagem de padrões corretamente reconhecidos, apresentados por estas duas tabelas é bom, tendo em vista que os caracteres processados pelo segundo estágio do sistema são aqueles numerais que possuem

grandes distorções em suas formas, e por isso não apresentam as características necessárias às suas classificações no 1º estágio do sistema, sendo de difícil classificação. O desempenho da rede de Hopfield utilizando o processo de síntese para os sistemas LSSM foi semelhante ao desempenho da rede de Hopfield empregando o método da projeção como regra de aprendizagem. No entanto, deve-se notar que foram utilizados apenas 58 numerais para treinar a rede de Hopfield utilizando o processo de síntese para os sistemas LSSM. Enquanto para a regra de aprendizagem empregando o método da projeção foram utilizadas 4 redes neurais, cada uma treinada com aproximadamente 60 numerais. A seção 8.4 finaliza esta dissertação propondo, como continuação deste trabalho, a pesquisa de algumas técnicas com vistas a melhorar a performance do sistema. A seguir são feitas as conclusões finais a respeito da técnica de reconhecimento apresentada nesta dissertação.

8.3 - CONCLUSÕES FINAIS SOBRE O MÉTODO DE RECONHECIMENTO EMPREGADO

O reconhecimento automático de caracteres manuscritos pode ser tratado de diferentes maneiras, as quais envolvem uma expressiva quantidade de ferramentas e conhecimentos de várias áreas tais como, estatística, ciência da computação, redes neurais e processamento de imagem. Muitos sistemas de reconhecimento de caracteres são apresentados pela literatura, no entanto, para que se possa comparar seus desempenhos, é necessário realizar testes que sejam comuns a todos [37]. Acreditamos que o sistema apresentado nesta dissertação tem uma boa performance, visto que o número de caracteres usados para treinamento da rede neural é baixo, se comparado as grandes variações existentes nos 1500 numerais usados para teste. É importante salientar que muitos dos numerais testados são difíceis de reconhecer sem informações contextuais. Mesmo assim, estes numerais foram incluídos em nossas estatísticas, pois acreditamos que um sistema deve ser avaliado de acordo com situações reais.

Em nossa opinião, os objetivos propostos na seção 1.4 foram atingidos, procuramos apresentar nosso trabalho de maneira clara e precisa, demonstrando todos os passos utilizados no desenvolvimento do sistema, e resultados obtidos. Pode-se notar através deste trabalho, que o reconhecimento automático de caracteres manuscritos é realmente um assunto complexo, se observado o estilo de escrita de cada pessoa e as diversas formas de representar um caracter. Sob este aspecto, acreditamos que nossa principal contribuição baseia-se no fato de se ter utilizado a rede de Hopfield discreta como classificador, tendo em vista as limitações referentes

a sua capacidade de armazenamento. E também, por se ter procurado aplicar técnicas para reduzir estas limitações, o que pode ser notado com o emprego do processo de síntese para os sistema LSSM e com a divisão do 2º estágio de classificação para a rede de Hopfield com pesos calculados pelo método da projeção.

Finalizando pode-se concluir, que características topológicas extraídas do carácter tais como número de cavidades e suas posições, e características referentes a distribuição de pontos da imagem tais como interseções com segmentos de reta, proporcionam ao sistema um bom índice de classificações corretas e grande velocidade de processamento. Porém, devido as falhas causadas pela digitalização e pelos instrumentos de escrita, há a necessidade de ser utilizado, em conjunto, um outro método capaz de reconstruir padrões distorcidos e classificá-los. Baseamo-nos nestes fatos para desenvolver o sistema aqui apresentado, cuja principal característica é o uso de dois diferentes tipos de classificadores, um classificador baseado em características topológicas e distribuição de pontos da imagem, e um segundo classificador que procura recuperar imagens armazenadas a partir de imagens distorcidas.

Para encerrar este capítulo são apresentadas, na última seção, sugestões para um futuro trabalho com a finalidade de melhorar o desempenho do sistema de reconhecimento apresentado.

8.4 - PROPOSTAS PARA APERFEIÇOAMENTO DO SISTEMA

Futuras pesquisas, para aperfeiçoamento do sistema com base nos resultados aqui apresentados, incluem o desenvolvimento de técnicas que consigam eliminar da forma do carácter, falhas e elevações resultantes do uso do scanner, da pressão exercida para escrever, do tipo de superfície de escrita e do tipo de instrumento de escrita. Isto permitiria melhor interpretação das características extraídas, apresentadas no capítulo 6 desta dissertação, e por conseguinte, uma percentagem maior de caracteres reconhecidos corretamente no primeiro estágio de classificação do sistema.

Um outro ponto, que possibilitaria extração de características mais precisas, seria o emprego de técnicas que permitissem um afinamento seletivo do carácter [38], ou seja, produzissem o esqueleto do numeral e eliminassem unicamente os segmentos espúrios de determinado comprimento.

Uma última questão a ser analisada seria o armazenamento de um maior número de padrões pela rede de Hopfield, através da utilização do método de síntese para os sistemas LSSM ou o desenvolvimento de um outro tipo de sistema com características semelhantes a rede de Hopfield. Em outras palavras, desenvolver um sistema capaz de recuperar padrões previamente memorizados a partir da apresentação de padrões distorcidos, porém, este sistema teria uma capacidade maior de armazenamento, conseqüentemente maior quantidade e variações de padrões de treinamento.

REFERÊNCIAS BIBLIOGRÁFICAS

- [1] MANTAS, John. "An Overview of Character Recognition Methodologies". *Pattern Recognition*, Vol. 19, nº 6, pp. 425-430, 1986.
- [2] SCHURMANN, J. "Reading Machines". *Proceedings 6^a Int. J. Conf. on Pattern Recognition*, Munique, pp. 1031-1044, 1982.
- [3] ROCHA, Jairo e PAVLIDIS, Theo. "Character Recognition Without Segmentation". *IEEE -Transactions on Pattern Recognition and Machine Intelligence*, Vol. 17, nº 9, pp. 903-909, set. 1995.
- [4] WANG, Gwo-En e WANG, Jhing-Fa. "A New Hierarchical Approach for Recognition of Unconstrained Handwritten Numerals". *IEEE - Transactions on Consumer Electronics*, Vol. 40, nº 3, pp. 428-437, ago. 1994.
- [5] FUKUSHIMA, Kunihiko e MIYAKE, Sei. "Neocognitron: A Neural Network Model for a Mechanism of Visual Pattern Recognition". *IEEE - Transactions on Systems, Man, and Cybernetics*, Vol. SMC-13, nº 5, set./out. 1983.
- [6] FUKUSHIMA, Kunihiko e MIYAKE, Sei. "Neocognitron: A New Algorithm for Pattern Recognition Tolerant of Deformations and Shifts in Position". *Pattern Recognition*, Vol. 15, pp. 445-469, 1982.
- [7] SHRIDHAR, M. e BADRELDIN, A. "A High Accuracy Syntactic Recognition Algorithm for Handwritten Numeral". *IEEE - Transactions on Systems, Man, and Cybernetics*, Vol. SMC-15, nº 1, jan./feb. 1985.

- [8] MANTAS, John. "Methodologies in Pattern Recognition and Image Analysis - A Brief Survey". *Pattern Recognition*, Vol. 20, n° 1, pp. 1-6, 1987.
- [9] TSIRIKOLIAS, Kostas e MERTZIOS, Basil G. "Statistical Pattern Recognition Using Efficient Two-Dimensional Moments with Applications to Character Recognition". *Pattern Recognition*, Vol. 26, n° 6, pp. 877-882, 1993.
- [10] SUEN, Ching Y. "Automatic Recognition of Handprinted Characters - The State of the Art". *Proceedings of the IEEE*, Vol. 68, n° 4, abr. 1980.
- [11] GOLSHAN, N. e HAU, C. "A Recognition Algorithm for Handprinted Arabic Numerals". *IEEE - Transactions on Systems, Man, and Cybernetics*, Vol. 6, pp. 246-250, jul. 1970.
- [12] KOSLAY, D. "Feature Extraction in An Optical Character Recognition Machine". *IEEE - Trans. Comput.*, Vol. 20, pp. 1063-1067, set. 1971.
- [13] MACLAUGHLIN, J. A. e RAVIV, J. "Nth-Order Autocorrelations in Pattern Recognition". *Inform. Contr.*, Vol. 12, pp. 121-142, 1968.
- [14] MILOSLAVSKAYA, N. K. e POLIAKOV, S. T. "Minsk-1 Digital Computer Simulation of An Algorithm for Handwritten Character Recognition". *Sov. Automat. Cont.*, Vol. 14, pp. 42-45, 1969.
- [15] SHIMURA, M. "Multicategory Learning Classifiers for Character Reading". *IEEE - Transactions on Systems, Man, and Cybernetics*, Vol. 3, pp. 74-85, jan. 1973.
- [16] IMPEDOVO, S., MARANGELLI, B. e FANELLI, A. M. "A Fourier Descriptors for Recognizing Nonstylized Numerals". *IEEE - Transactions on Systems, Man, and Cybernetics*, Vol. SMC-8, n° 8, pp. 640-645, ago. 1978.

- [17] SHRIDHAR, M. e BADRELDIN, A. "High Accuracy Character Recognition Algorithm using Fourier and Topological Descriptors". Pattern Recognition, Vol. 17, nº 5, pp. 515-524, 1984.
- [18] GRANLUND, G. H. "Fourier Preprocessing for Handprint Character Recognition". IEEE - Trans. Comput., vol. 21, pp.195-201, fev. 1972.
- [19] ANDREWS, H. C. "Multi-Dimensional Rotations in Feature Selection". IEEE - Trans. Comput., Vol.20, pp. 1045-1051, set. 1971.
- [20] WENDLING, S., GAGNEUX e STAMON, G. "A Set of Invariants Within the Power Spectrum of Unitary Transformations". IEEE - Trans. Comput., Vol.27, pp. 1213-1216, dec. 1978.
- [21] WENDLING, S., GAGNEUX e STAMON, G. "Hadamard and Harr Transforms and Their Power Spectra in Character Recognition". In Proc. Joint Workshop Pattern Recognition and Artificial Intelligence, pp. 103-112, jun. 1976.
- [22] NIEMANN, H. "A Comparison of Classification Results in Character Recognition by Man and by Machine". Proc. 3rd Int. Joint Conf. Pattern Recognition, pp. 144-147, nov. 1976.
- [23] OTT, R. "On Feature Selection by Means of Principle Axis Transform and Nonlinear Classification". Proc. 2nd Int. Joint Conf. Pattern Recognition, pp. 220-222, ago.1974.
- [24] JAMES, Mike. "Pattern Recognition". pp. 37, British Library, 1987.
- [25] MICHAEL, M. e LIN, W. C."Experimental Study of Information Measure and Inter-Intra Class Distance Ratios of Feature Selection and Orderings." IEEE - Transactions on Systems, Man, and Cybernetics, Vol. 3, pp. 172-181, mar. 1973.

- [26] KWON, S. K. e LAI, D. C. "Recognition Experiments with HandPrinted Numerals". Proc. 2nd Int. **Joint** Conf. Pattern Recognition, pp. 215-219, jun.1976
- [27] HUANG, J. S. e CHUANG, K. "Heuristic Approach to Handwritten Numeral Recognition". **Pattern Recognition**, Vol. 19, nº 1, pp. 15-19, 1986.
- [28] FAUSETT, Laurene. "Fundamentals of Neural Network". pp. 103-135, Prentice Hall, 1994.
- [29] WASSERMAN, Philip D. "Neural Computing: Theory and Practice". pp. 27-11, Van Nostrand Reinhold, 1989.
- [30] ALEKSANDER, Igor. "Neural Computing Architectures: The Design of Brian-Like Machines". pp. 239-255, Mit Press, 1989.
- [31] PAO, Yoh Han. "Adaptive Pattern Recognition and Neural Networks". pp. 158-167, Addison - Wesley, 1989.
- [32] ABU-MOSTAFA, Y. S. e ST. JAQUES, J. "Information Capacity of The Hopfield Model". **IEEE** - Transactions on Information Theory, Vol. 31, nº 4, pp. 461 - 464, 1985.
- [33] LI, Jian-Hua, MICHEL, Abthony N. e POROD, Wolfgang. "Analysis of A Class of Neural Networks: Linear Systems Operating on A Closed Hypercube". **IEEE** - Transactions on Circuits and Systems, Vol.36, nº 11, pp. 1405-1422, nov. 1989.
- [34] FACON, Jaques. "Processamento e Análise de Imagens". VI Escuela Brasileño Argentina **de** Informatica, jul 1993.
- [35] GONZALEZ, Rafael C. e WOODS, Richard E. "Digital Image Processing". Addison - Wesley, 1992.
- [36] FREEMAN, James A. "Neural Networks: Algorithms, Applications and Programming Techniques". pp. 127-168, Addison Wesley, 1991.

- [37] BOTTOU, Léon e CORTES, Corina. "Comparison of Classifier Methods: A case Study in Handwritten **D**igit Recognition". Proceedings of The International Conference on Document **A**nalysis and Recognition, pp.77-82, 1994.
- [38] SERRA, Jean. "**I**mage Analysis and Mathematical Morphology". Vol.2, Academic Press, 1988.
- [39] MALKI, H. A. e MOGHADDAMJOO, A. "Using The Karhunen-Loe've Transformations in The Back-Propagation Training Algorithm". IEEE - Transactions on Neural Networks, Vol. 2, nº 1, jan. 1991.
- [40] YÜCER, Cem e OFLAZER, Kemal. "A Rotation, Scaling, and Translation Invariant Pattern Classification System". Pattern Recognition, Vol.26, nº 5, pp. 687-710, 1993.
- [41] HERTZ, John, KROGH, Anders e PALMER, Richard G. "Introduction to The Theory of Neural **C**omputation". pp. 11-63, Addison - Wesley, 1990.
- [42] LI, Jian-Hua, MICHEL, Abthony N.e POROD, Wolfgang. "Qualitative Analysis and Synthesis of A Class of Neural Networks". IEEE - Transactions on Circuits and Systems, Vol.35, pp. 976-986, ago.1988.
- [43] LI, Jian-Hua, MICHEL, Abthony N.e POROD, Wolfgang. "Qualitative Analysis and Synthesis of A Class of Neural Networks: Variable Structure Systems with Infinite Gains". IEEE - Transactions on Circuits and Systems, Vol.36, pp. 713-731, mai.1989.
- [44] STRANG, Gilbert. "Linear Algebra and Its Applications". pp. 144-162, Harcourt Brace Javanovich, Publishers, 1988.