

UNIVERSIDADE FEDERAL DE SERGIPE
CENTRO DE CIÊNCIAS EXATAS E TECNOLOGIA
DEPARTAMENTO DE COMPUTAÇÃO

Trabalho de Conclusão de Curso

**Avaliação de um mecanismo de zoneamento
para classificação de imagens de caracteres manuscritos**

Aluna: Sandra Eliza Fontes de Avila

Orientador: Leonardo Nogueira Matos

Co-orientadora: Cinthia Obladen de Almendra Freitas

São Cristóvão, 25/09/2006

Dedicatória

À minha mãe, aos meus irmãos e a toda minha família.

Agradecimentos

Ao professor Leonardo Nogueira Matos pela orientação, paciência e incentivo recebido durante o desenvolvimento deste trabalho.

À professora Cinthia O. de A. Freitas pela idéia do trabalho.

Aos membros da banca, professores Hendrik Teixeira Macêdo e Jugurta Rosa Montalvão Filho, pelas minuciosas revisões realizadas, pela boa vontade com que me receberam e pelas sugestões fornecidas.

À minha mãe, Maria Eliza, pelo amor, carinho, apoio e incentivo em todos os momentos de minha vida e pela grandeza de mulher que representa.

Ao meu pai, Antônio Ernane, que apesar de não poder acompanhar minha vida universitária, sempre foi um exemplo de pessoa estudiosa em quem eu pude me espelhar.

Ao meu namorado, Bruno, pelo amor, carinho, companheirismo, amizade, força e por sempre ir comigo à UFS independente do dia ou horário, principalmente nos finais de semana e feriados.

Aos meus irmãozinhos, Diego e Vinicius, por trazerem constantemente todos os seus amigos para casa, fazendo com que eu ficasse no meu quarto e mantivesse assim o foco no estudo.

A todos os amigos da UFS, professores, alunos e funcionários do DCOMP.

À minha grande família pelo afeto e pela compreensão dos meus momentos ausentes.

À família do meu namorado, Rose, Djalma, Sheilinha e Débora, por torcer com bastante força pela conclusão deste trabalho.

A todas as pessoas que rezaram por mim, especialmente minha vovó Flor.

Por fim, a todos que colaboraram direta ou indiretamente na execução deste trabalho.

Resumo

O reconhecimento automático de caracteres manuscritos tem sido um tópico de pesquisa intensivamente estudado por mais de duas décadas. Essa atividade é motivada pela grande quantidade de aplicações encontradas, dentre as quais pode-se citar o processamento de envelopes postais e cheques bancários. Entretanto, problemas complexos e de difícil solução continuam em aberto, uma vez que existem inúmeras maneiras de se representar um caracter na escrita manual. Como os seres humanos são capazes de reconhecer eficientemente padrões diferentes como manuscritos de várias formas, o presente trabalho emprega um mecanismo de zoneamento, inspirado no modelo biológico para o reconhecimento de caracteres manuscritos isolados. As bases de dados utilizada para os experimentos são a *Letter* e a IRONOFF, com caracteres manuscritos minúsculos e maiúsculos do alfabeto inglês. Neste trabalho, duas arquiteturas de redes neurais artificiais (RNA) são testadas para a classificação dos caracteres, uma RNA MLP (*Multiple Layer Perceptron*) convencional e classe-modular, ou seja, um conjunto de redes neurais MLP de forma que cada classe do problema possui uma rede específica. Resultados experimentais com as bases utilizadas demonstram uma superioridade na utilização da arquitetura classe-modular em relação à RNA MLP convencional. Para a base *Letter*, as taxas de reconhecimentos para a arquitetura convencional e classe-modular respectivamente são 83,10% e 93,67%. Para a base IRONOFF, as taxas de reconhecimento para a arquitetura convencional são 56,14% para os caracteres minúsculos e 63,49% para os caracteres maiúsculos, enquanto que para a arquitetura classe-modular, as taxas de reconhecimentos para os caracteres minúsculos são 80,75% e para os caracteres maiúsculos são 89,21%.

Sumário

1	Introdução	7
1.1	Descrição do problema	8
1.2	Objetivos	11
1.3	Contribuições	12
1.4	Organização	13
2	Reconhecimento de caracteres manuscritos	14
2.1	Reconhecimento óptico de caracteres	14
2.2	Extração de características	18
2.2.1	Transformadas matemáticas e expansão em séries	18
2.2.2	Representação estatística	19
2.2.3	Representação topológica e geométrica	21
2.3	Classificação	23
2.3.1	Classificadores estatísticos	24
2.3.2	Redes neurais artificiais	25
3	Abordagem Proposta	28
3.1	Pré-processamento	29
3.2	Extração de características	32
3.2.1	Sistema Visual Humano	32
3.2.2	Histogramas direcionais com zoneamento	34
3.3	Classificação	37
3.3.1	Redes Neurais Artificiais	37

3.3.2	Arquitetura Classe-Modular	41
4	Experimentos e Resultados	47
4.1	Base de dados	47
4.2	Simulador SNNS	52
4.3	Experimentos	53
4.3.1	Experimentos para a Arquitetura Convencional	53
4.3.2	Experimentos para a Arquitetura Classe-modular	54
4.3.3	Análise dos resultados	58
5	Conclusões	61
5.1	Perspectivas de trabalhos futuros	62
	Referências Bibliográficas	68

Lista de Figuras

1.1	Estilo de escrita de caracteres minúsculos de diversos escritores.	9
1.2	Estilo de escrita de caracteres maiúsculos de diversos escritores.	10
1.3	Similaridade entre caracteres distintos.	11
1.4	Arquitetura típica de um sistema de reconhecimento de caracteres.	12
2.1	Zoneamento aplicado a imagens retangulares	20
2.2	Histograma	21
2.3	Estruturas topológicas: máximos e mínimos nos contornos, pontos ascendentes e descendentes (ARICA; YARMAN-VURAL, 2001)	22
2.4	Direções do código de cadeia (GONZALEZ; WOODS, 2000)	23
3.1	Diagrama aplicado para o reconhecimento de caracteres manuscritos	28
3.2	Pré-processamento: limiarização	30
3.3	Pré-processamento: abertura	31
3.4	Imagens negativos após a determinação da caixa de contorno do caracter	32
3.5	Estruturas do olho humano	33
3.6	Captura de uma linha através do campo visual	34
3.7	Características extraídas do contorno da imagem (MATOS, 2004)	35
3.8	Representação da direção do vetor	36
3.9	Zoneamento aplicado a imagens retangulares	37
3.10	Arquitetura de uma rede neuronal com três camadas.	39
3.11	Modelo de um neurônio artificial.	40
3.12	Arquitetura convencional onde K classes estão misturadas (OH; SUEN, 2002)	42
3.13	Arquitetura para uma RNA-MLP classe-modular (OH; SUEN, 2002)	43

4.1	Gráficos da evolução do MSE no treinamento das redes em função do número de ciclos	59
-----	--	----

Lista de Tabelas

2.1	Quadro comparativo dos trabalhos que utilizaram a base de dados IRONOFF	17
3.1	Codificação da saída dos caracteres	44
3.2	Taxa de reconhecimento do conjunto de testes de (OH; SUEN, 2002)	44
4.1	Distribuição de caracteres - Base de dados <i>Letter</i>	49
4.2	Distribuição de caracteres minúsculos - Base IRONOFF	50
4.3	Distribuição de caracteres maiúsculos - Base IRONOFF	51
4.4	Matriz de confusão para a classificação com duas classes (REZENDE, 2003)	55
4.5	Taxas de sensibilidade e especificidade para a base <i>Letter</i>	56
4.6	Taxas de sensibilidade e especificidade de caracteres minúsculos e maiúsculos da base IRONOFF	57
4.7	Quadro comparativo do trabalho proposto e estudos relacionados	60

Lista de Siglas

HMM	<i>Hidden Markov Models</i> (Modelos Escondidos de Markov)
ICR	<i>Intelligent Character Recognition</i> (Reconhecimento Inteligente de Caracter)
k -NN	<i>k-Nearest Neighbour</i> (k vizinhos mais próximos)
MLP	<i>Multiple Layer Perceptron</i> (Perceptron Multicamadas)
MS-TDNN	<i>Multi-State Time-Delay Neural Networks</i> (Multi-Estados de Redes Neurais de Atraso Temporal)
MSE	<i>Mean Square Error</i> (Erro Médio Quadrático)
NN	<i>Neural Network</i> (Rede Neural)
OCR	<i>Optical Character Recognition</i> (Reconhecimento Óptico de Caracteres)
RNA	Rede Neural Artificial
SNNS	<i>Stuttgart Neural Network Simulator</i> (Simulador de Rede Neural de Stuttgart)
SVM	<i>Support Vector Machine</i> (Máquina de Vetor de Suporte)

Capítulo 1

Introdução

A escrita manuscrita é uma habilidade pessoal de cada indivíduo. De acordo com (PLAMONDON; SRIHARI, 2000), a escrita manuscrita consiste de marcas gráficas em uma superfície, cujo propósito na maioria das vezes é a comunicação obtida em virtude da relação dos símbolos convencionais das linguagens. A escrita manuscrita é valorizada por ter contribuído muito para o desenvolvimento das culturas e civilizações.

Cada manuscrito é um conjunto de ícones, os quais são caracteres ou letras que possuem suas formas básicas definidas. Há regras para combinação de letras para formar unidades representativas lingüística de alto nível. Por exemplo, existem regras para combinação de formas e letras individuais para formar palavras cursivas no alfabeto latino.

A razão da escrita manuscrita ter persistido ao longo dos anos na era do computador é a conveniência do papel e da caneta, comparada aos teclados, para as numerosas situações do dia-a-dia (PLAMONDON; SRIHARI, 2000). O estudo dos caracteres manuscritos está ligado ao desenvolvimento de métodos de reconhecimento voltados para aplicações do mundo real envolvendo caracteres manuscritos, tais como: processamento automático de cheques bancários, envelopes postais, formulários, textos manuscritos, entre outros.

Apesar de muitos trabalhos tratarem do reconhecimento de caracteres manuscritos, o tema continua sendo um grande desafio, por ser um problema complexo e de difícil solução (ARICA; YARMAN-VURAL, 2001). Muitos são os fatores que dificultam o reconhecimento automático de caracteres manuscritos. Documentos escritos à mão apresentam componentes de grande complexidade, devido às inúmeras possibilidades de se apresentar diversas formas e tama-

nhos quando escritos por pessoas diferentes ou pela mesma pessoa, bem como podem possuir aparência ambígua ou estarem ilegíveis. Esta diversidade pode ser causada por vários aspectos, tais como o tipo de papel e o instrumento de escrita utilizados, habilidade, origem, nível de educação, profissão, idade e estado emocional do escritor (SUEN; GUO; LI, 1994).

Para a realização automática do reconhecimento de caracteres manuscritos, os dados a serem identificados precisam ser convertidos para a forma digital através do uso de *scanners*, no caso da escrita no papel, ou provenientes de superfícies de cristal líquido, que capturam os caracteres escritos sobre ela com um bastão que assemelha-se a uma caneta. As duas abordagens são diferenciadas como digitalização *off-line* e *on-line*, respectivamente. Na digitalização *off-line*, somente o manuscrito completo é disponível em uma imagem. No caso da digitalização *on-line*, as coordenadas bi-dimensionais de pontos sucessivos são descritas como uma função do tempo e armazenadas em ordem, isto é, a ordem dos segmentos de caracteres traçados pelo escritor é prontamente disponibilizada. A abordagem *off-line* envolve análise de espaço-luminosidade da imagem, enquanto que a abordagem *on-line* fornece uma representação espaço-temporal da entrada (PLAMONDON; SRIHARI, 2000).

1.1 Descrição do problema

Há cerca de seis mil anos o homem exprime o seu pensamento através da escrita. Desde então, a escrita manual é uma das formas mais naturais de comunicação entre as pessoas. Essa forma de comunicação gera uma quantidade volumosa de dados em papel. Assim, é conveniente processar os dados contidos nestes papéis de tal maneira que os computadores tenham capacidade de “ler” e interpretar documentos manuscritos.

O reconhecimento de caracteres manuscritos tem sido um tópico de pesquisa extensivamente estudado na comunidade científica devido ao seu valor teórico no reconhecimento de padrões e ao vasto número de aplicações encontradas, tais como, o processamento de códigos postais em envelopes e numerais em cheques bancários. Embora muita pesquisa já tenha sido desenvolvida, este continua sendo um problema de difícil solução devido às variações existentes na escrita manual. Nas Figuras 1.1 e 1.2 é apresentada a diversidade de estilos de escrita do mesmo carácter entre amostras de vários escritores para caracteres minúsculos e maiúsculos da base de



Figura 1.1: Estilo de escrita de caracteres minúsculos de diversos escritores.



Figura 1.2: Estilo de escrita de caracteres maiúsculos de diversos escritores.

dados utilizada, respectivamente.

Outro problema no reconhecimento de caracteres manuscritos, é a similaridades entre caracteres distintos, por exemplo, U e V, O e Q, H e K, E e L entre outros. Na Figura 1.3 são apresentadas algumas similaridades entre caracteres na base de dados estudada.

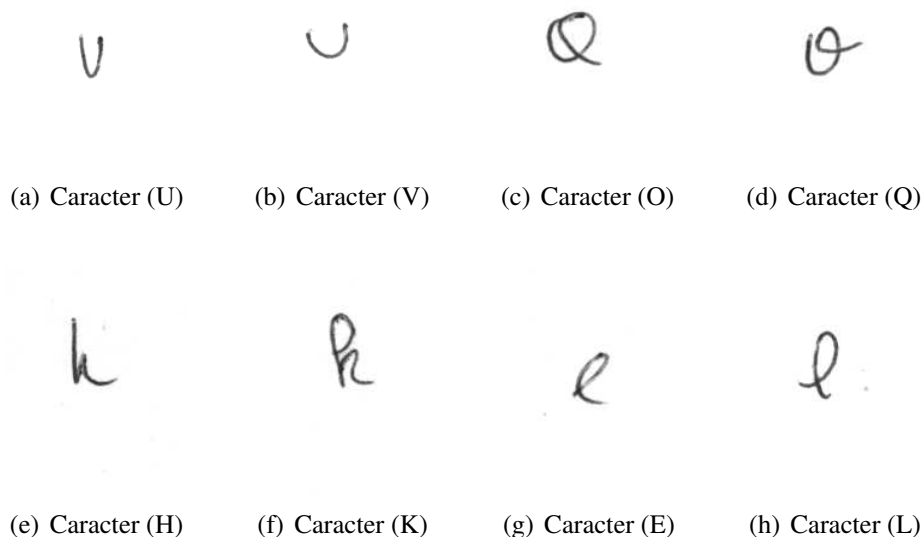


Figura 1.3: Similaridade entre caracteres distintos.

O problema abordado neste trabalho consiste no reconhecimento automático de caracteres manuscritos *off-line*.

1.2 Objetivos

Tipicamente, o trabalho refere-se ao reconhecimento de dois conjuntos da base de dados utilizada: os caracteres minúsculos e maiúsculos, correspondentes aos caracteres do alfabeto inglês.

A abordagem proposta de reconhecimento é composta por uma seqüência de três módulos principais: (1) pré-processamento, (2) extração de características e (3) classificação, como ilustrado na Figura 1.4. A etapa de pré-processamento visa melhorar a qualidade da imagem digitalizada, suavizando ruídos e reduzindo as variações presentes nos caracteres. A etapa de extração de características busca representar a imagem de cada caracter por um conjunto de

características relevantes, isto é, cada caracter é representado por um vetor de características. E a etapa de classificação tenta identificar a classe à qual o caracter analisado pertence, de acordo com as características fornecidas.

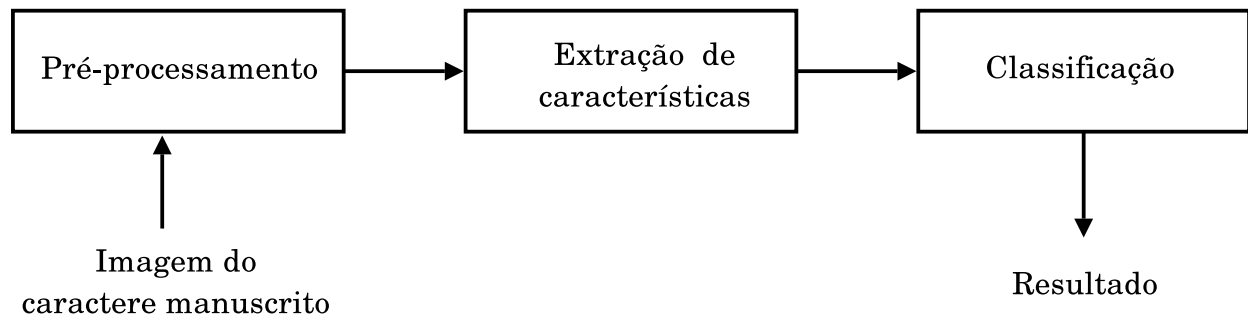


Figura 1.4: Arquitetura típica de um sistema de reconhecimento de caracteres.

1.3 Contribuições

Podem ser citadas como contribuições deste trabalho:

- Avaliação das arquiteturas de redes neurais convencional e classe-modular *Multiple Layer Perceptron* para o reconhecimento de caracteres manuscritos, possibilitando também sugerir como podem ser realizados os processos de treinamento e reconhecimento nas arquiteturas avaliadas. A diferença entre os resultados alcançados de acordo com as arquiteturas utilizadas mostra que a pesquisa sobre a arquitetura classe-modular é promissora e prova ser merecedora de investigações adicionais do paradigma de modularidade de classes;
- Avaliação de mecanismo de zoneamento para a extração de características de imagens de caracteres manuscritos isolados. No Capítulo 3 foi apresentado o algoritmo de extração de características para processamento de imagens de caracteres manuscritos inspirado no modelo biológico, conforme proposto por (MATOS, 2004). Os resultados obtidos nos experimentos foram competitivos quando comparados aos publicados na literatura especializada, o que revela a boa capacidade do extrator de características em melhorar a separabilidade entre as classes.

A principal contribuição deste trabalho está relacionada à originalidade da investigação do mecanismo de zoneamento juntamente com a classificação neural classe-modular, o que refletiu nos bons resultados atingidos neste trabalho.

1.4 Organização

Este trabalho está organizado como segue. No Capítulo 2 é apresentado um resumo de diversos trabalhos relacionados ao reconhecimento de caracteres manuscritos, também são descritas as principais técnicas de extração de características e os tipos de classificadores comumente usados. No Capítulo 3 é descrito a abordagem proposta para o reconhecimento de caracteres manuscritos. No Capítulo 4 são mostradas as bases de dados utilizadas para os experimentos, os resultados obtidos e a análise destes. No Capítulo 5 são apresentadas as conclusões do trabalho, as contribuições e as perspectivas de trabalhos futuros.

Capítulo 2

Reconhecimento de caracteres manuscritos

Neste capítulo é apresentada uma revisão teórica sobre o reconhecimento de caracteres manuscritos. O capítulo está organizado da seguinte forma: na Seção 2.1 é apresentado os trabalhos mais significativos encontrados na literatura que utilizaram a mesma base de dados empregada no presente estudo; na Seção 2.2 são descritas as principais técnicas de extração de características; e na Seção 2.3 são apresentados determinados tipos de classificadores existentes.

2.1 Reconhecimento óptico de caracteres

Os sistemas de reconhecimento óptico de caracteres (*Optical Character Recognition* - OCR) são uma das aplicações mais bem sucedidas na área de reconhecimento de padrões. Há pelo menos cinco décadas, OCR é tema de estudo de várias pesquisas e desenvolvimento (TRIER; JAIN; TAXT, 1996).

Em alguns sistemas OCRs, os caracteres a serem reconhecidos se encontram impressos em uma determinada fonte, apresentando um formato bem definido. Inicialmente, os OCRs eram capazes de reconhecer apenas tipos específicos de fontes. Posteriormente, surgiram os OCRs multi-fonte que permitiram que a capacidade de reconhecimento se estendesse para um conjunto maior de opções de fontes. Por fim, surgiram os OCRs omni-fonte, capazes de reconhecer qualquer tipo de fonte.

Entretanto, apesar de existir diversos sistemas OCR, estes somente podem reconhecer documentos de textos impressos com alta qualidade. Devido a essas limitações, surgiram os sistemas ICR (*Intelligent Character Recognition*). Diferentemente dos sistemas OCRs, os ICRs lidam com um problema bem mais complexo, uma vez que os caracteres a serem reconhecidos são manuscritos e não mais impressos. Esta complexidade é devido às variações de estilo existentes na escrita manuscrita. Esta escrita pode se apresentar de duas formas, isolada ou cursiva. No primeiro caso os caracteres estão dispostos de forma não conectada, enquanto que no segundo estão dispostos de forma completamente irrestrita, ou seja, conectados ou eventualmente desconectados.

Os dados de entrada de um sistema de reconhecimento de caracteres podem ser provenientes da digitalização *on-line* ou *off-line*. No caso *on-line*, à medida que o traço do caractere é delimitado pelo escritor, este é prontamente apresentado ao sistema, enquanto no caso *off-line*, somente a imagem completa do que foi escrito previamente é apresentada ao sistema. Os sistemas OCRs são essencialmente do tipo *off-line*, enquanto os sistemas ICRs podem ser de ambos os tipos.

Atualmente a área de pesquisa em sistemas ICRs é muito intensa, devido ao aumento do poder computacional em computadores pessoais e várias outras formas de tecnologia na área de reconhecimento de caracteres manuscritos. Existem diversos trabalhos nesta área, sendo alguns abordados no presente estudo.

(LALLICAN; VIARD-GAUDIN; KNERR, 2000) propõem um método que visa alcançar, em uma base de dados *off-line* de palavras, o desempenho de reconhecimento conseguido em uma base de dados *on-line*. Nesta abordagem, denominada “*OrdRec*”, o objetivo é encontrar o provável caminho que assemelha-se à verdadeira trajetória da caneta. O sistema é baseado na descrição de sinais dos manuscritos através de grafos e utiliza Modelos Escondidos de Markov (*Hidden Markov Models* - HMM) para o processo de reconhecimento. O treinamento é realizado com imagens *on-line* e *off-line* disponíveis na base de dados IRONOFF. A taxa de reconhecimento é de 93% empregando-se a abordagem (*Ord. on&off*), enquanto que, sem a utilização do “*OrdRec*” o sistema alcança uma taxa de reconhecimento de 90,2%. E utilizando apenas imagens *on-line*, o sistema obtém uma taxa de reconhecimento de 94,5%.

(TAY et al., 2001) descreve um sistema de reconhecimento de palavras manuscritas *off-line* que combina Redes Neurais (*Neural Network* - NN) e HMM, NN-HMM. O sistema emprega um método de corte vertical para a segmentação das palavras, gerando um grafo que descreve todas as maneiras possíveis da segmentação de uma palavra em letras. A rede computa as probabilidades para cada hipótese da letra através do grafo. E por meio de HMMs, para cada letra, soma-se a estas probabilidades todos os caminhos possíveis encontrados no grafo. A base de dados utilizada para os experimentos foi a IRONOFF. A taxa de reconhecimento alcançada foi 91,7%.

(POISSON; VIARD-GAUDIN; LALLICAN, 2002) apresentam uma arquitetura que explora o reconhecimento de caracteres manuscritos isolados. As arquiteturas são testadas separadamente e depois uma arquitetura híbrida é proposta. Os experimentos são realizados utilizando caracteres da base IRONOFF, *on-line* e *off-line*. As taxas de reconhecimento para caracteres minúsculos da base IRONOFF *on-line* é de 90,7% para a arquitetura proposta e, para uma MLP (*Multiple Layer Perceptron*) convencional é de 90,2%. No caso de caracteres maiúsculos, a taxa é de 94,2% e 93,6%. Para os caracteres minúsculos *off-line* as taxas para a arquitetura proposta e uma MLP convencional são de 80,5% e 77,8%, respectivamente, e 89,9% e 87,1% para os caracteres maiúsculos.

(TAY et al., 2003) realiza um estudo de palavras manuscritas isoladas e avalia o reconhecimento de caracteres no processo de segmentação. Para o processo de reconhecimento utiliza um sistema híbrido NN-HMM. Os experimentos são realizados na base IRONOFF *off-line*, que é particionada em IRONOFF-196 e IRONOFF-2000, contendo um léxico de 196 palavras e 2000 palavras, respectivamente. As taxas de reconhecimento na base IRONOFF-2000 são 83,1% para caracteres e 88,1% para palavras.

(AHMAD et al., 2004) propõe um sistema híbrido para o reconhecimento de manuscritos que combina Máquina de Vetor de Suporte (*Support Vector Machine* - SVM) e HMM, SVM-HMM. Os resultados experimentais são obtidos utilizando dígitos e caracteres isolados da base de dados IRONOFF *on-line*. A taxa de reconhecimento para caracteres minúsculos é 92,47% e para caracteres maiúsculos é 95,46%.

(CAILLAULT; VIARD-GAUDIN; AHMAD, 2005) analisam o comportamento de arquiteturas

híbridas baseadas em representações Multi-Estados de Redes Neurais de Atraso Temporal e HMM (*Multi-State Time-Delay Neural Networks HMM - MS-TDNN HMM*) aplicadas no reconhecimento de palavras manuscritas. O sistema utilizou a base de dados IRONOFF *on-line*. A taxa de reconhecimento obtida é 92,01%.

(AIRES, 2005) investiga mecanismos de zoneamento perceptivo para reconhecimento de caracteres manuscritos isolados. No processo de reconhecimentos utiliza-se NN MLP classe-modular. Uma metodologia para múltiplos classificadores é aplicada ao problema de reconhecimento de caracteres maiúsculos. As taxas médias de reconhecimento obtidas, para os zoneamentos avaliados, foram as seguintes: zoneamento tipo 4 = 82,89%, zoneamento tipo 5Horizontal = 81,75%, zoneamento tipo 5Vertical = 80,94%, zoneamento tipo 7 = 84,73% e zoneamento tipo 5Horizontal-5Vertical-7 = 85,9%. A base de dados utilizada é a IRONOFF *off-line*.

Um quadro comparativo das taxas de reconhecimento dos trabalhos citados na literatura que utilizaram a base de dados IRONOFF é apresentado na Tabela 2.1.

Tabela 2.1: Quadro comparativo dos trabalhos que utilizaram a base de dados IRONOFF

Autor	Caracteres				Palavras		
	<i>on-line</i>		<i>off-line</i>		<i>on-line</i>	<i>off-line</i>	<i>on&off</i>
	minúsc.	maiúsc.	minúsc.	maiúsc.			
(LALLICAN; VIARD-GAUDIN; KNERR, 2000)	-	-	-	-	94,5%	-	93%
(TAY et al., 2001)	-	-	-	-	-	91,7%	-
(POISSON; VIARD-GAUDIN; LALLICAN, 2002)	90,7%	94,2%	80,5%	89,9%	-	-	-
(TAY et al., 2003)	-	-	-	-	-	88,1%	-
(AHMAD et al., 2004)	92,47%	95,46%	-	-	-	-	-
(CAILLAULT; VIARD-GAUDIN; AHMAD, 2005)	-	-	-	-	92,01%	-	-
(AIRES, 2005)	-	-	-	85,9%	-	-	-

Os trabalhos citados nesta seção estão diretamente relacionados com os experimentos realizados no presente trabalho, eles ajudam entender a complexidade do problema de reconhecimento de caracteres manuscritos, bem como fornecem um panorama abrangente dos resultados

encontrados na literatura com a base IRONOFF, os quais poderão ser comparados ao método proposto.

2.2 Extração de características

A representação da imagem representa um dos papéis mais importantes em sistemas de reconhecimento de caracteres. Estes podem ser representados de muitas maneiras, dependendo do tipo de característica usada para a sua representação. Uma inspeção técnica minuciosa é apresentada em (TRIER; JAIN; TAXT, 1996). Nesta seção as técnicas de extração, comumente empregadas no reconhecimento de caracteres, são apresentadas em três categorias, como é proposto em (ARICA; YARMAN-VURAL, 2001): (1) transformadas matemáticas e expansão em séries, (2) representação estatística e (3) topológicas e geométricas.

2.2.1 Transformadas matemáticas e expansão em séries

Transformadas matemáticas e expansão em séries possibilitam uma análise da imagem do caractere em outro domínio de representação, que pode ressaltar certas características difíceis de serem observadas na imagem original. Seu princípio de funcionamento consiste em decompor os sinais através de uma combinação linear de funções base elementares. Um sinal contínuo geralmente contém informação a mais do que é necessário para o propósito da classificação. Os coeficientes resultantes da combinação linear fornecem uma representação compacta do sinal analisado, que pode ser usada como características em sistemas de reconhecimento. As transformadas matemáticas e as expansões em séries comumente usadas no reconhecimento de caracteres são descritas a seguir.

- **Transformada de Fourier:** Fornece a representação do sinal no domínio da frequência. Essa ferramenta tem sido aplicada no reconhecimento de caracteres de diversas maneiras, como nos trabalhos de (WANG; CHEN; LIN, 1994) *apud* (ARICA; YARMAN-VURAL, 2001) e (ZHU; SHI; WANG, 1999) *apud* (ARICA; YARMAN-VURAL, 2001).
- **Transformada de Gabor:** É uma variação da transformada de Fourier janelada, que consiste em dividir o sinal em regiões e aplicar a transformada de Fourier a cada uma

destas. No caso da transformada de Gabor, a janela usada não possui um tamanho discreto, mas é definida por uma função Gaussiana.

- **Wavelets:** A transformada wavelet é utilizada na decomposição de sinais através de um conjunto de funções elementares, denominadas *wavelets*. Tal como a transformada de Gabor, a transformada wavelet aplica um mecanismo de janelamento. As funções elementares representam o sinal em diferentes níveis de resolução, isto é, o sinal é analisado por janelas grandes e pequenas. Quando o sinal é visto através de uma janela grande, suas características gerais são observadas. Similarmente, quando o sinal é visto por uma janela pequena, suas particularidades e detalhes são detectados (CORREIA, 2005). As imagens, letras ou palavras, são representadas pelos coeficientes da transformada wavelet, que correspondem aos vários níveis de resolução. Estes coeficientes são usados como entrada no classificador para o reconhecimento (SHIOYAMA; WU; NOJIMA, 1998) *apud* (ARICA; YARMAN-VURAL, 2001).
- **Momentos:** São considerados uma forma de expansão em séries, uma vez que a imagem original do caractere pode ser completamente reconstruída a partir dos coeficientes resultantes da combinação linear com um polinômio característico. Os momentos comumente utilizados são os momentos centrais, momentos centrais normalizados, momentos de Legendre e momentos Zernike (TRIER; JAIN; TAXT, 1996).
- **Transformada de Karhunen-Loève:** Baseada em autovetores que permitem reduzir a dimensão do conjunto analisado, criando um novo conjunto de características que são combinações lineares dos dados do conjunto original. Essa transformada é considerada uma transformada ótima no sentido de compressão de informação. Tal técnica tem sido usada em diversos problemas de reconhecimento de padrões, como o reconhecimento de faces.

2.2.2 Representação estatística

A representação estatística permite uma análise da distribuição dos *pixels* que compõem a imagem do caractere. Esse tipo de representação não é capaz de reconstruir a imagem original,

mas fornece um conjunto de características de dimensão reduzida. As características, desta representação, mais empregadas no reconhecimento de caracteres são descritas a seguir.

- **Divisão em zonas:** A imagem do caracter é dividida em várias regiões retangulares (Figura 3.9), sobrepostas ou não, denominadas *zonas*. As densidades dos pontos ou outras características em cada uma das regiões são analisadas e usadas para compor o vetor de características. Tipicamente, no zoneamento as imagens de caracteres têm sido divididas em seis zonas ou regiões em uma grade retangular (SUEN; GUO; LI, 1994) (OLIVEIRA et al., 2002). O benefício desta técnica é que as características de cada zona da imagem são processadas separadamente, permitindo uma análise detalhada do caracter em regiões diferentes. Recentemente, pesquisas têm emergido no sentido de determinar métodos de zoneamento inteligentes que buscam melhorar a maneira de particionar o caracter, como o trabalho de (RADTKE et al., 2003) que emprega algoritmos genéticos na determinação do zoneamento.

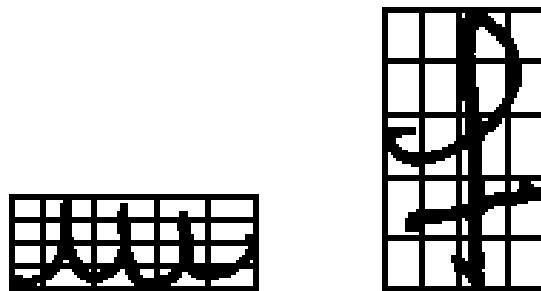


Figura 2.1: Zoneamento aplicado a imagens retangulares

- **Cruzamentos e distâncias:** A característica do cruzamento é obtida do número de vezes que o contorno do caracter é interceptado por um segmento de linha especificado em uma determinada região. A escolha da posição desse segmento pode revelar características discriminantes dos caracteres. Por exemplo, um segmento horizontal pode descrever partes ascendentes e descendentes e um segmento de linha vertical passando pelo centro de gravidade do caracter pode descrever intersecções estáveis da sua representação. Outro conjunto de características que pode ser empregado são as distâncias entre os pontos que formam o caracter e segmentos de linha especificados em determinadas posições da imagem, tais como os seus limites inferiores e superiores (OLIVEIRA et al., 2002).

- **Histograma:** Descreve a distribuição estatística dos níveis de cinza da imagem em termos do número de *pixels* em cada nível. Há diferentes tipos de histogramas na literatura, tais como histograma de projeção e histograma direcional. *Histograma de projeção* faz uso da projeção da imagem sobre cada uma das duas dimensões existentes. A cada dimensão é associado um vetor onde é armazenado o número de *pixels* com tonalidade de cinza acima de um determinado limiar (SILVA; THOMÉ, 2003). Os *histogramas direcionais* não são restritos à imagem original do carácter, podendo ser calculados a partir de imagens resultantes de um processamento, como imagens que contém a fase e a curvatura do carácter (MATOS, 2004). Na Figura 2.2 é ilustrado uma representação de um histograma.

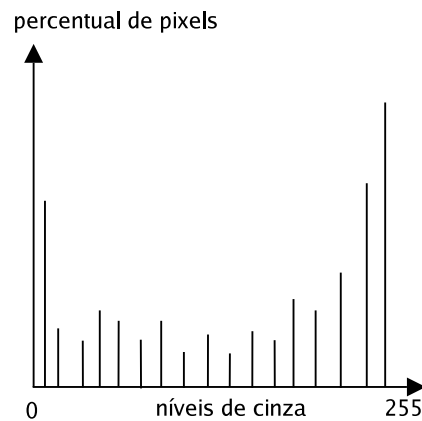


Figura 2.2: Histograma

2.2.3 Representação topológica e geométrica

Várias propriedades globais e locais dos caracteres podem ser representadas através de características topológicas e geométricas. Este tipo de representação pode fornecer algum conhecimento sobre a estrutura do carácter ou pode descrever aspectos importantes da estrutura que forma o carácter. As técnicas de extração que fornecem características topológicas ou geométricas são descritas a seguir.

- **Estruturas topológicas:** Neste tipo de representação, uma estrutura pré-definida é buscada no carácter. Estas estruturas são segmentos de reta em várias direções, concavidades, convexidades, pontos iniciais e finais, pontos de cruzamento, inserções, cavidades,

junções de linhas, pontos isolados, ascendentes, descendentes, corpo do carácter, enfim, qualquer estrutura que descreva a geometria que forma o carácter analisado. Muitas destas técnicas utilizam máscaras para a extração de características, como as máscaras de Kirsch (DONG, 1999).

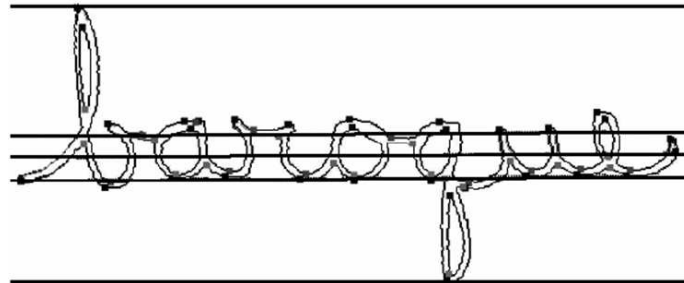


Figura 2.3: Estruturas topológicas: máximos e mínimos nos contornos, pontos ascendentes e descendentes (ARICA; YARMAN-VURAL, 2001)

- **Medida e aproximação de propriedades geométricas:** Os caracteres podem ser representados por medidas geométricas, tais como a taxa entre sua altura e largura, as distâncias relativas, horizontal e vertical, entre os primeiros e últimos pontos e distância entre dois pontos. Uma característica importante é a medida de curvatura ou a mudança de curvatura. As quantidades geométricas medidas podem ser, às vezes, aproximadas por um conjunto de características mais compacto, através de métodos que incluem aproximação poligonal da imagem do esqueleto do carácter.
- **Codificação:** Códigos de cadeia são métodos usados para codificar a curvatura do carácter. Um dos codificadores mais populares é o código de cadeia de Freeman. Esta codificação é essencialmente obtida mapeando-se os traços do carácter em uma cadeia de códigos que representa as suas direções horizontais, verticais e diagonais nos sentidos esquerda, direita, para cima e para baixo (Figura 2.4). O conjunto de características é formado pelos números que formam o código de cadeia do carácter analisado (CAI; LIU, 1999).
- **Grafos e árvores:** Os grafos são estruturas usadas para representar as primitivas topológicas, tais como traços, laços, pontos cruzados. Existem dois tipos de representação

de caracteres utilizando grafos. O primeiro tipo faz uso das coordenadas da forma do carácter. E o segundo, é uma representação abstrata na qual os nós correspondem aos traços e as arestas correspondem aos relacionamentos entre os traços. As árvores também podem ser empregadas para representar o conjunto de características dos caracteres, desde que exista uma relação hierárquica neste conjunto.

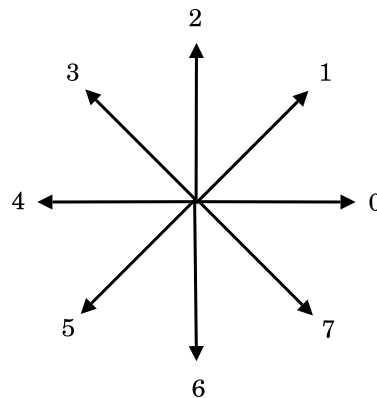


Figura 2.4: Direções do código de cadeia (GONZALEZ; WOODS, 2000)

(ARICA; YARMAN-VURAL, 2001) realizou uma cobertura ampla de métodos de extração de características concluindo que o seu principal objetivo é extrair e selecionar um conjunto de características que maximize a taxa de reconhecimento com a menor quantidade possível de elementos. Neste trabalho, o extrator de características utiliza um procedimento de zoneamento com histogramas direcionais. A etapa de extração de características é detalhada na Seção 3.2.

2.3 Classificação

Classificação é uma das mais frequentes tarefas de tomada de decisão da atividade humana. Um problema de classificação acontece quando um objeto precisa ser associado a um determinado grupo ou classe baseando-se em um número de atributos observados e relacionados àquele objeto (ZHANG, 2000). Muitos problemas em negócios, ciências e na indústria podem ser tratados como problemas de classificação, por exemplo: análise de crédito, diagnóstico médico, controle de qualidade, reconhecimento de voz e reconhecimento de caracteres manuscritos.

Nos sistemas de reconhecimento de caracteres, o classificador escolhido deve ser adequado

ao tipo e ao formato das características extraídas (TRIER; JAIN; TAXT, 1996). Os classificadores estatísticos e as redes neurais artificiais são os métodos de classificação mais utilizados (LIU; SAKO; FUJISAWA, 2002) (LIU et al., 2003) (KAPP, 2005).

2.3.1 Classificadores estatísticos

Nos classificadores estatísticos, conceitos da teoria da decisão estatística são aplicados para estabelecer as fronteiras de decisão entre as classes padrões. Técnicas usadas para a classificação como a regra dos k vizinhos mais próximos (*k-Nearest Neighbour* - k -NN), regra da decisão de Bayes, máquinas de vetores de suporte SVM (*Support Vectors Machines*), análise de agrupamento (*Clustering Analysis*) e modelos escondidos de Markov (*Hidden Markov Models* - HMM) têm sido utilizadas no reconhecimento de caracteres manuscritos (LALLICAN; VIARD-GAUDIN; KNERR, 2000) (KOERICH; SABOURIN; SUEN, 2003).

- **Regra dos k vizinhos mais próximos (k -NN):** Método no qual um exemplo x cuja classe é desconhecida é classificado na mesma classe a que pertence a maioria dos exemplos X mais próximos. Em geral, a medida de distância utilizada é a distância Euclidiana.
- **Regra da decisão de Bayes:** Calcula a probabilidade de um evento, dado que um outro tenha ocorrido. O método permite que se ajuste uma probabilidade *a priori* (conhecida) de um dado evento de acordo com novas evidências envolvendo um outro evento que apresenta relação de dependência com o primeiro.
- **Máquinas de vetores de suporte (SVM):** É baseado na teoria de aprendizagem estatística e na otimização de programação quadrática, sendo um classificador binário, de modo que vários SVM podem ser combinados para formar um sistema de classificação (LIU et al., 2003). Sistemas que empregam SVM têm superado os resultados obtidos com técnicas tradicionais como as redes neurais, entretanto, necessitam de mais memória e apresentam um maior custo computacional.
- **Análise de agrupamento (*Clustering analysis*):** É o processo de organizar uma coleção de padrões, usualmente representados por vetores de características, em grupos baseados em suas similaridades (ARICA; YARMAN-VURAL, 2001). Intuitivamente, padrões em um

mesmo agrupamento são mais similares entre eles do que um padrão em um outro agrupamento. A variedade de técnicas usadas para representar os dados, medir a similaridade entre os elementos e o modo de representar os agrupamentos produziram vários métodos de *clustering*. Esta abordagem é não-supervisionada, portanto a decisão não é baseada em um conjunto de treinamento (JAIN; DUIN; MAO, 2000).

- **Modelos escondidos de Markov (HMM):** É um processo duplamente estocástico, com um processo estocástico não visível, o qual não é observável (escondido), mas que pode ser observado através de outro processo estocástico que produz a seqüência de observações (CORREIA, 2005). Aplicações de HMM surgiram inicialmente no reconhecimento de fala (FECHINE; MADEIRO; NETO, 2003) e depois foram estendidas para modelar caracteres (CAI; LIU, 1999) e palavras (EL-YACOUBI et al., 1999) (FREITAS; BORTOLOZZI; SABOURIN, 2001).

2.3.2 Redes neurais artificiais

As redes neurais artificiais (RNA) formam uma disciplina ampla, com vários modelos existentes, dentre os quais pode-se citar: o perceptron multi-camadas (*Multiple Layer Perceptron - MLP*), as redes de funções de base radiais e as redes neurais probabilísticas (LIU et al., 2003). As entradas dos classificadores neurais são características da forma de n -tuplas ou vetores. Além de classificador, as redes neurais também podem ser usadas para a extração de características. Para tanto, coloca-se diretamente na sua entrada as imagens padrões, em vez de um vetor de características que represente estas imagens.

- **Rede perceptron multi-camadas (MLP):** A MLP é uma das RNAs mais empregada em problemas de classificação. Ela permite a solução de problemas nos quais as classes são não-linearmente separáveis. A RNA MLP é abordada na Seção 3.3.1.
- **Rede de funções de base radiais:** Na sua forma mais simples, uma *rede de funções de base radial* apresenta três camadas com papéis totalmente diferentes. A primeira camada, a *camada de entrada*, é a conexão do modelo com o ambiente externo, à qual são apresentados os vetores de entrada. A *camada intermediária* (escondida) é composta de k

neurônios totalmente conectados aos vetores de entrada. Esta camada tem a tarefa de realizar uma transformação não-linear do espaço n -dimensional da entrada em outro espaço k -dimensional. Os neurônios dessa camada são um conjunto de funções de base radial, que constitui uma base arbitrária no espaço por ele formado. As funções de base radial produzem uma resposta maior em uma determinada região (radial) Cada função requer um centro e um parâmetro escalar denominado *raio*, na qual calcula a distância entre o vetor de entrada e o centro da função de base radial associada. A última camada, a *camada de saída*, transforma o espaço vetorial interno em uma saída, através de um processo linear no qual realiza a soma ponderada das saídas das unidades radiais (HAYKIN, 1999).

- **Rede neural probabilística:** Uma *rede neural probabilística* possui três camadas (entrada, escondida e saída). Na camada de saída, a quantidade de neurônios é a mesma do número de classes a serem reconhecidas. Em uma rede neural probabilística, um vetor de entrada é apresentado à rede na primeira camada. Na camada escondida é calculada a distância existente entre o vetor de entrada e a soma ponderada das entradas, e através de uma função de base radial é produzido um vetor cujos elementos variam entre 0 e 1. A terceira camada, soma essas contribuições para cada classe de entrada e produz uma saída que constitui o vetor de probabilidades. Finalmente, uma determinada função, na saída da terceira camada, escolhe o máximo destas probabilidades, atribuindo 1 para esta classe e 0 nos outros casos (HAYKIN, 1999).

A diferença principal entre classificadores estatísticos e redes neurais é que os parâmetros da rede são otimizados em um processo discriminativo de aprendizagem supervisionado que visa separar os padrões em classes diferentes (HAYKIN, 1999). Quando a estrutura da rede é apropriadamente designada e o conjunto de treinamento possui muitos elementos, as redes neurais são capazes de fornecer uma alta precisão da classificação dos padrões desconhecidos do conjunto teste. Entretanto, o treinamento é lento e corre-se o risco da rede perder sua capacidade de generalização e tornar-se superespecializada (LIU; SAKO; FUJISAWA, 2002) (LIU et al., 2003).

Muitos sistemas utilizam uma combinação de mais de um tipo de classificador, no intuito de aumentar o reconhecimento uma vez que nesse caso, as deficiências de um classificador podem

ser suprimidas através do bom desempenho dos outros (MATOS, 2004). Outro recurso usado para aumentar a eficácia dos sistemas é a utilização de um classificador modular, que consiste em vários classificadores, sendo que cada um é especialista em uma das classes do problema (OH; SUEN, 2002). Este recurso é aplicado no presente trabalho, sendo apresentado na Seção 3.3.2.

Capítulo 3

Abordagem Proposta

Neste capítulo é descrita a abordagem aplicada no presente trabalho. Segundo (GONZALEZ; WOODS, 2000), é conceitualmente útil dividir o espectro de técnicas empregadas em análise de imagens em três áreas básicas. Essas áreas são (1) processamento de baixo nível, (2) processamento de nível intermediário e (3) processamento de alto nível. Embora as subdivisões entre os processamentos não possuam fronteiras definitivas, elas provêm uma arquitetura útil para a categorização de vários processos que são componentes inerentes de um sistema autônomo de análise de imagens. Na Figura 3.1 é ilustrado esses conceitos e as etapas aplicadas na abordagem proposta para o reconhecimento de caracteres manuscritos.

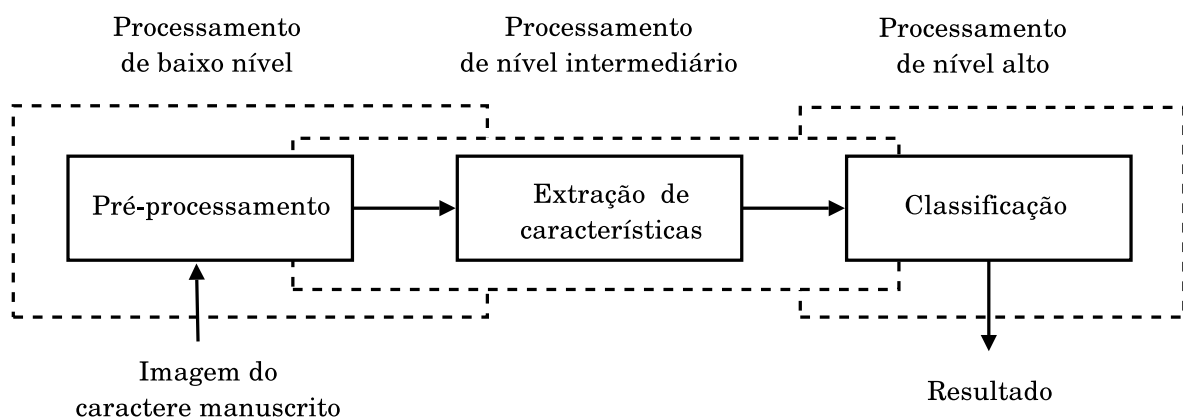


Figura 3.1: Diagrama aplicado para o reconhecimento de caracteres manuscritos

O *processamento de baixo nível* trata de funções que podem ser vistas como ações automáticas, não requerendo qualquer inteligência por parte do sistema de análise de imagens. As

tarefas que se enquadram neste nível são, em geral, aquisição de imagens e pré-processamento. Como no presente trabalho, a abordagem proposta é empregada em bases de imagens já coletadas, logo é desnecessário um processo de aquisição.

O *processamento de nível intermediário* trata da tarefa de extração e caracterização de componentes (por exemplo, regiões) em uma imagem resultante de um processo de baixo nível. Os processos de nível intermediário incluem a segmentação e a extração de características de componentes ou regiões da imagem.

E por fim, o *processamento de alto nível* envolve o reconhecimento e interpretação de padrões. Na abordagem proposta neste capítulo, o reconhecimento abrange a tarefa de classificação de imagens dos caracteres manuscritos.

A seguir são detalhadas cada uma das etapas para o reconhecimento de caracteres manuscritos apresentadas na Figura 3.1.

3.1 Pré-processamento

Na fase de pré-processamento diversas operações são aplicadas às imagens digitais para o tratamento destas, tais como limiarização, eliminação de ruídos, realce de contraste, de tal forma que as características das imagens possam ser realçadas. Neste trabalho, a base de dados utilizada encontra-se segmentada em caracteres isolados. Sendo assim, a fase de pré-processamento é composta pelas técnicas descritas a seguir:

- **Limiarização** (*thresholding*): é uma das mais importantes abordagens para a segmentação de imagens (GONZALEZ; WOODS, 2000). Nesta técnica, é analisado a similaridade dos níveis de cinza da imagem extraindo os objetos de interesse através da seleção de um limiar T que separa os agrupamentos de níveis de cinza. Uma imagem limiarizada $g(x, y)$ é definida como:

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

onde $f(x, y)$ corresponde ao nível de cinza do *pixel*, os *pixels* rotulados com 1 correspon-

dem aos objetos e os *pixels* rotulados com 0 correspondem ao fundo e T é um valor de tom de cinza predefinido denominado *limiar*.

A maior dificuldade na limiarização de uma imagem é a melhor determinação do limiar (*thresholding*), ou seja, do ponto de separação dos *pixels*. A mais simples de todas as técnicas de limiarização é a do particionamento do histograma da imagem por um limiar único T , denominada *limiarização global simples*. O histograma é uma das formas mais comuns de se representar a distribuição dos níveis de cinza de uma imagem. Este descreve a distribuição estatística dos níveis de cinza em termos do número de *pixels* em cada nível, e esta distribuição pode também ser dada em termos da percentagem do número total de *pixels* da imagem.

O sucesso desse método depende inteiramente de quão bem o histograma pode ser particionado. No presente estudo, a limiarização global simples é aplicada, sendo que o limiar T é igual a 0.8. Na Figura 3.2 é apresentada a imagem original da base de dados e a imagem resultante do método de limiarização.

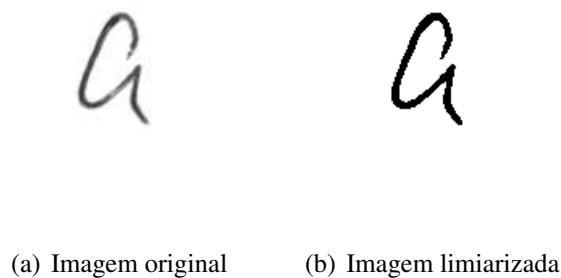


Figura 3.2: Pré-processamento: limiarização

- **Eliminação de ruídos:** O objetivo desta fase é eliminar ruídos resultantes do processo de limiarização. Uma vez que a imagem limiarizada é binária, foi utilizado neste trabalho filtros da morfologia matemática binária para realizar a remoção de ruídos. Foi utilizada a operação morfológica de abertura (GONZALEZ; WOODS, 2000), que consiste na aplicação de uma operação de erosão seguida de dilatação. A escolha do elemento estruturante utilizado é um fato muito importante, já que serão excluídos todos os objetos menores

do que o objeto estruturante determinado. Neste estudo, o elemento estruturante (EE) utilizado foi o quadrado de dimensão 3×3 .

$$EE = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}$$

O procedimento de *erosão* elimina objetos finos ou pequenos e objetos maiores tem sua área reduzida. A *dilatação*, por sua vez, realiza a operação inversa, elimina buracos finos e une objetos. Os objetos, por sua vez, têm sua área aumentada.

As operações de erosão e dilatação, isoladamente, causam distorções nas áreas dos objetos. No entanto, sua combinação como é o caso da operação de abertura, gera resultados mais interessantes. A *abertura* elimina pontos indesejáveis sem modificar o tamanho e aspecto dos objetos. Na Figura 3.3 é ilustrado passo-a-passo o processo de abertura.

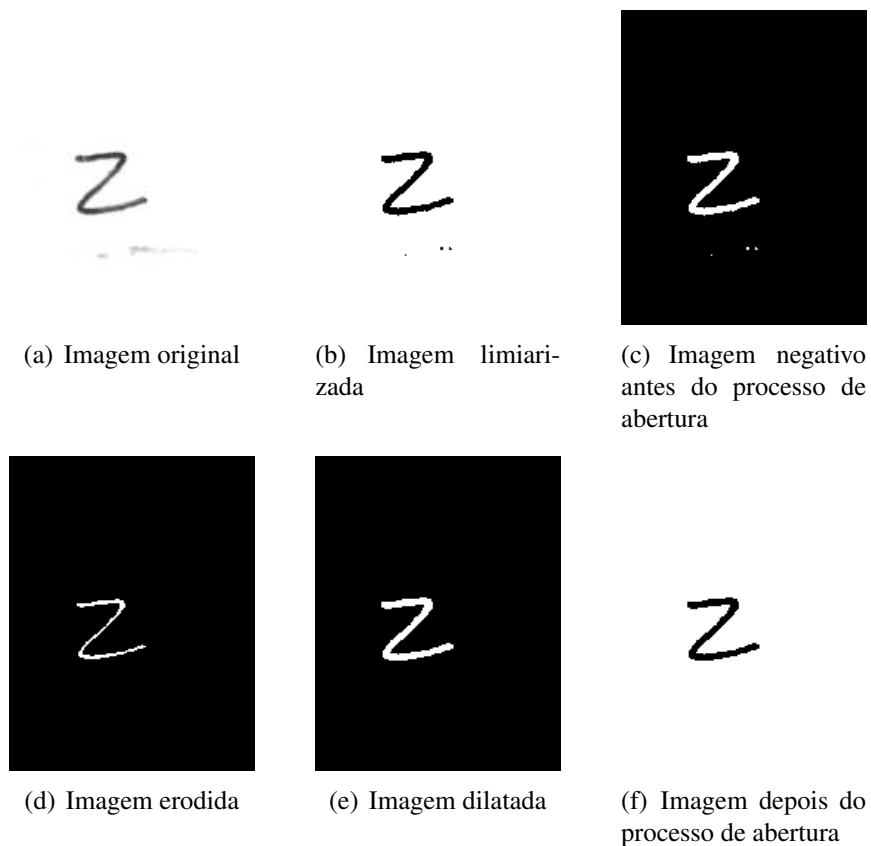


Figura 3.3: Pré-processamento: abertura

- **Caixa de contorno** (*bounding box*): este pré-processamento busca extrair da imagem os limites acima, abaixo, esquerda e direita criando ao redor do caracter uma caixa retangular. A Figura 3.4 exemplifica a criação da caixa de contorno em imagens da base de dados utilizada.

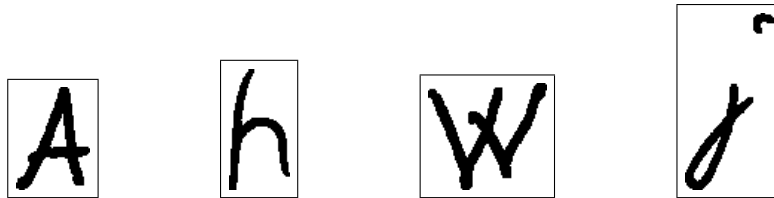


Figura 3.4: Imagens negativos após a determinação da caixa de contorno do caracter

3.2 Extração de características

A visão dos seres humanos é seletiva às componentes espaciais de alta frequência, que são fundamentais para reconhecimento de formas e do sentido de profundidade. Como estas componentes se manifestam nos pontos de contorno, então admite-se que estes pontos contêm toda informação necessária para o reconhecimento da imagem do caracter.

O método de extração de características utilizado neste trabalho foi desenvolvido (MATOS, 2004). Este método é baseado na imagem do contorno do caracter e é inspirado no modelo biológico. O objetivo do método é conciliar a informação extraída do contorno com um procedimento de zoneamento, imitando desta forma os campos receptivos do sistema visual humano.

3.2.1 Sistema Visual Humano

A visão humana é um procedimento complexo, que faz parte do sistema nervoso e que emprega diversas células neurais especializadas. Nesse contexto, o olho é o elemento frontal do sistema de visão que capta os raios luminosos e os projeta na retina, uma membrana sensorial que age como um transdutor, convertendo a luz em sinais nervosos correspondentes que são, posteriormente, conduzidos ao córtex visual do cérebro através do nervo óptico.

O olho, ilustrado na Figura 3.5, tem a forma de um globo aproximadamente esférico que constitui um órgão perfeito para o foco, seleção e absorção dos raios luminosos que chegam

até ele. A luz refletida pelo meio atinge, inicialmente, o olho através da membrana transparente que o reveste, a *córnea*, e em seguida, penetra no globo ocular por um orifício denominado *pupila*. A quantidade de luz que atravessa a pupila é regulada pela *íris* que controla o tamanho do seu diâmetro. Ao atingir o *cristalino*, a luz é finalmente direcionada para a *retina*, onde uma imagem invertida e reduzida é projetada sem mudar seu aspecto original, ou seja, a proporção entre a largura e o comprimento da imagem é mantida.

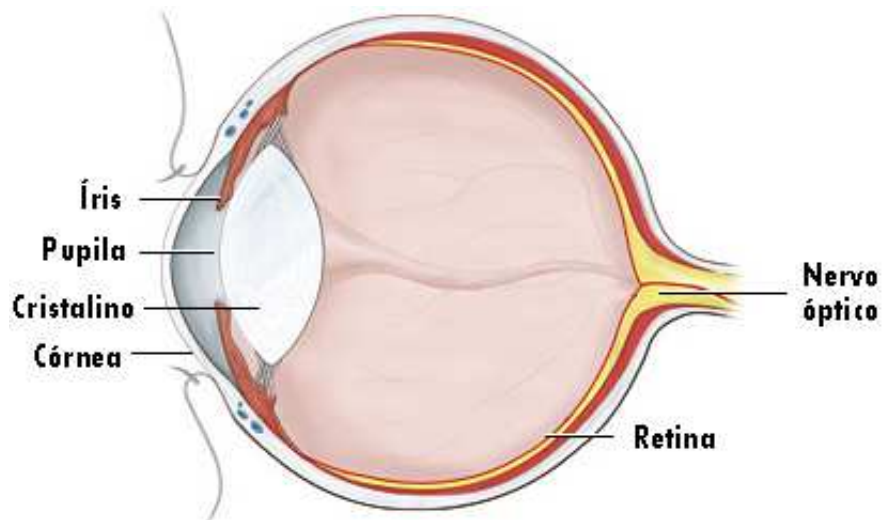


Figura 3.5: Estruturas do olho humano

A retina localiza-se na parte posterior interna do olho e é organizada em *campos receptivos*. Cada campo receptivo é composto por um conjunto de células fotorreceptores (células sensíveis à luz) de diferentes tipos, os cones e os bastonetes, que através de uma reação fotoquímica, convertem a imagem visual em impulsos nervosos equivalentes. Os *cones* são responsáveis pela visão fotóptica ou de luz clara, sensíveis às cores e permitem o discernimento de detalhes nas imagens. Os *bastonetes*, ao contrário, são responsáveis pela visão escotópica ou de luz escura, que detectam a imagem em preto e branco.

O campo visual é construído em decorrência da aglomeração de cones e bastonetes na retina. Na Figura 3.6 é ilustrado como os cones e os bastonetes capturam uma linha reta. À esquerda é representada uma linha imaginária sobreposta aos campos receptivos. À direita, as respostas destes a esta linha estão inclinadas de acordo com a cor representada, a cor mais escura significa mais resposta. Sendo assim, o resultado não é uma linha. Em vez disso, é uma coleção de

diferentes respostas. A linha é construída a partir destas respostas pelo córtex visual do cérebro.

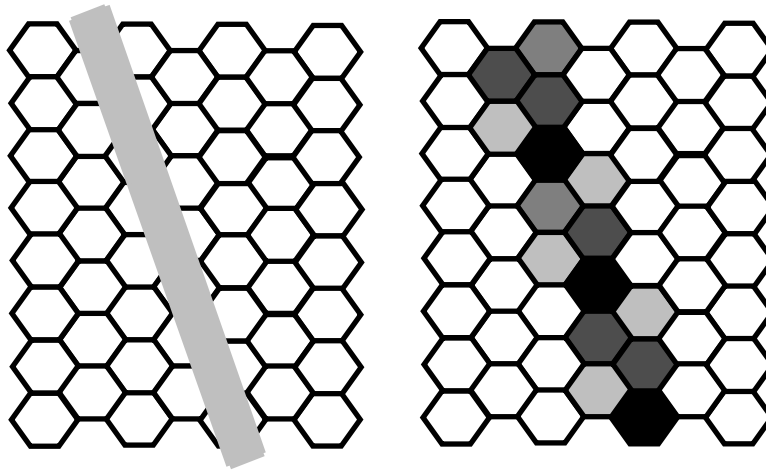


Figura 3.6: Captura de uma linha através do campo visual

Da mesma forma que o sistema visual humano age, o método de extração de características utilizado neste trabalho propõe projetar as imagens dos caracteres nas zonas sem alterar a proporção entre a largura e o comprimento das imagens, sendo que as zonas “mais escuras” significam mais resultados.

3.2.2 Histogramas direcionais com zoneamento

Um *histograma direcional* é um procedimento para realizar cruzamento de características pontuais de uma imagem através de uma distribuição de frequências. As características pontuais, isto é, pertencentes a um *pixel* isoladamente são a curvatura e o gradiente. Sendo o gradiente um vetor, esta característica é desmembrada em duas outras: a magnitude e a fase. Uma vez que a fase tem uma variação limitada ao intervalo de 0 a 2π , ela foi usada para gerar intervalos de mesma amplitude empregados como classes para a distribuição de frequência, por este motivo o método é denominado *direcional*.

Em seu estudo, Matos trabalhou com imagens do contorno por admitir que nele esteja localizada a informação mais relevante para a realização do reconhecimento, já que imagens de caracteres manuscritos são representações visuais de um traçado feito a mão. O autor desenvolveu um método de histograma que realiza a contagem da intensidade do vetor gradiente por classe de variações da fase ou da curvatura. Este procedimento de contagem adotou uma abor-

dagem mais simples, por considerar que no contorno, que é uma região de transição de nível de cinza, a intensidade do gradiente é máxima, podendo ser considerada igual em todos os pontos.

Alguns trabalhos relacionados empregam no início do processo um algoritmo de normalização em escala, cujo propósito é a padronização da dimensão das imagens para um número de linhas e colunas previamente estabelecido. O processo de normalização geralmente causa distorções geométricas. Para evitar tais distorções, são utilizadas zonas que são independentes da resolução espacial da imagem (MATOS, 2004). Estas não possuem dimensão fixa, elas mapeiam uma determinada região, como por exemplo, o canto superior direito ou o canto inferior esquerdo. Naturalmente, para imagens de dimensões distintas as zonas possuem dimensões distintas.

Entretanto, em cada uma destas zonas as proporções de *pixels* com as mesmas intensidades de fase e curvatura são aproximadamente iguais desde que elas pertençam a mesma classe. Na imagem do dígito dois, por exemplo, os pontos de curvatura costumam ocorrer nas mesmas posições, as maiores inflexões ocorrem nos cantos superior direito e inferior esquerdo e nas extremidades do traçado, como ilustrado na Figura 3.7. Como as zonas não possuem dimensão fixa então a contagem do número de ocorrências de *pixels* por classe de ângulo e curvatura em cada uma delas é bastante variável: imagens de área grande originam histogramas com frequências elevadas e imagens de área pequena, histogramas com frequências menores. Para caracterizar uma imagem sem introduzir variações nos histogramas decorrentes do tamanho, foi utilizado histogramas de frequências relativas. As etapas de extração de características e construção dos histogramas direcionais e geração do vetor de características são brevemente explicadas a seguir.

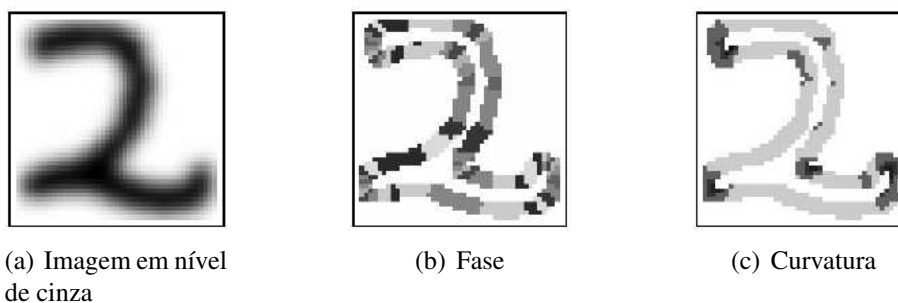


Figura 3.7: Características extraídas do contorno da imagem (MATOS, 2004)

Para a extração de características, inicialmente é aplicado na imagem binária um algoritmo de extração de bordas a fim de obter as posições dos pontos de contorno. Em seguida, é empregado o filtro da média 3×3 r vezes sobre a imagem original a fim de se obter uma imagem em nível de cinza (Matos utilizou $r = 4$), como ilustrado na Figura 3.7(a). Para limitar a variação dos níveis de cinza ao intervalo $[0, 1]$ é utilizado um algoritmo de normalização. Logo depois, é aplicado o filtro de Roberts (GONZALEZ; WOODS, 2000) para obter a magnitude e fase do vetor gradiente. Por último, é calculado a fase (Figura 3.7(b)) e a curvatura (Figura 3.7(c)) nos pontos de contorno, segundo procedimento apresentado em (SHI et al., 2002) *apud* (MATOS, 2004).

Para a construção dos histogramas direcionais e a geração do vetor de características, primeiramente é realizada a extratificação da fase do gradiente em 10 classes de amplitude $\pi/10$. Ângulos separados de π radianos são considerados iguais, portanto, é analisada apenas a direção do vetor, ignorando o sentido, como ilustrado na Figura 3.8. Em seguida, é efetuada a extratificação dos valores de curvatura em 5 classes (MATOS, 2004). Logo após, é aplicado um procedimento de zoneamento que divide a imagem em 16 blocos (4×4) para imagens aproximadamente quadráticas. No caso de imagens retangulares é realizada uma divisão em 20 blocos, 4 blocos por linha e 5 por coluna caso a imagem tenha uma disposição horizontal (Figura 3.9(a)), ou 5 blocos por linha e 4 por coluna, caso tenha uma disposição vertical (Figura 3.9(b)).

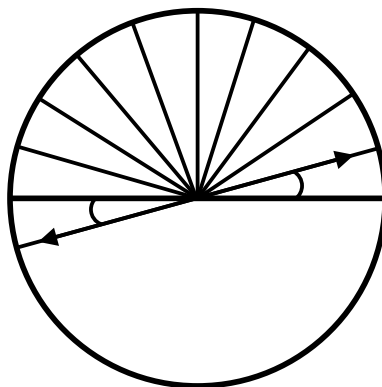


Figura 3.8: Representação da direção do vetor

Para cada bloco é feita a contagem do número de ocorrência de cada uma das classes da fase e curvatura, originando histogramas com 15 classes (MATOS, 2004). Para cada um destes histogramas, os valores são normalizados de modo que o número de ocorrências em cada classe

fique limitado ao intervalo $[0, 1]$. Então, é efetuado a concatenação dos histogramas obedecendo a seqüência apresentada a seguir. Assim, o vetor de características possui dimensão 375×1 , uma vez que a seqüência abaixo possui 25 posições e cada posição apresenta histogramas com 15 classes.

1	2	3	4	5
6	7	8	9	10
11	12	13	14	15
16	17	18	19	20
21	22	23	24	

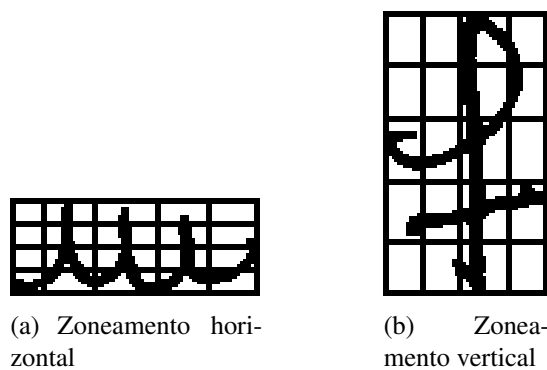


Figura 3.9: Zoneamento aplicado a imagens retangulares

3.3 Classificação

3.3.1 Redes Neurais Artificiais

As redes neurais artificiais (RNAs) emergiram como uma ferramenta importante para a classificação. As recentes atividades de pesquisa são vastas em classificação neural, estabelecendo-as como uma alternativa promissora para vários métodos de classificação convencionais (ZHANG, 2000).

Neste trabalho, a classificação é realizada com a utilização de RNAs MLP. A vantagem dessas redes fundamenta-se nos seguintes aspectos teóricos. Primeiramente, as RNAs são métodos auto-adaptativos e dirigidos pelos dados, ou seja, ajustam-se aos dados por conta própria, sem qualquer especificação explícita da forma de distribuição ou função para um dado modelo. Se-

gundo, elas são aproximadores funcionais universais, pois aproximam qualquer função com precisão arbitrária (HORNİK; STINCHCOMBE; WHITE, 1989) (HORNİK, 1991). Considerando que qualquer procedimento de classificação busca uma relação funcional entre um grupo relacionado e os atributos do objeto, a identificação precisa desta função é sem dúvida importante. Terceiro, RNAs são modelos não-lineares, o que as fazem flexíveis na modelagem de relacionamentos complexos do mundo real. Finalmente, RNAs podem calcular as probabilidades *a posteriori* que provêm a base para estabelecer regras de classificação e desempenhar análises estatísticas (RICHARD; LIPPMANN, 1991).

RNAs são sistemas paralelos distribuídos compostos por unidades de processamentos simples, denominados de neurônios, células ou nós, que calculam determinadas funções matemáticas (normalmente não-lineares). Tais unidades são dispostas em uma ou mais camadas interligadas por um grande número de conexões, geralmente unidirecionais. Na maioria dos modelos, estas conexões estão associadas a pesos, os quais armazenam o conhecimento representado no modelo e servem para ponderar a entrada recebida por cada neurônio da rede (BRAGA; LUDERMIR; CARVALHO, 2000).

A arquitetura de uma RNA MLP consiste de um conjunto de neurônios que constituem a camada de entrada, uma ou mais camadas escondidas e a camada de saída. O número de neurônios na camada de entrada é determinado pela dimensão do vetor de características, a quantidade de neurônios na saída é determinada pela resposta desejada, e a quantidade de camadas escondidas e o número de neurônios em cada uma destas camadas são determinados pelo projetista da rede (HAYKIN, 1999). Existem trabalhos (YAMAZAKI; LUDERMIR; SOUTO, 2002) que buscam determinar a arquitetura ótima de uma RNA, uma vez que se um pequeno número de neurônios for mantido, a rede corre o risco de não desempenhar a tarefa de aprendizagem devido a sua capacidade limitada. Por outro lado, se um grande número de neurônios for empregado a rede pode se tornar superespecializada e perder sua capacidade de generalização. Na Figura 3.10 é apresentado uma representação gráfica de uma RNA MLP.

Os neurônios na camada de entrada servem para distribuir a informação contida em um vetor de características para todos os neurônios da camada escondida. O valor associado a qualquer neurônio é denominado de *ativação* e é representado pela aplicação de uma função não-linear

à soma ponderada das entradas. Ou seja, para um neurônio k , o cálculo da ativação é dado por:

$$net_k = \sum_{j=1}^N x_j w_{kj} \quad (3.1)$$

onde N é o número de entradas do neurônio k ; x_1, x_2, \dots, x_N são as entradas do neurônio e $w_{k1}, w_{k2}, \dots, w_{kj}$ são pesos sinápticos associados a cada entrada.

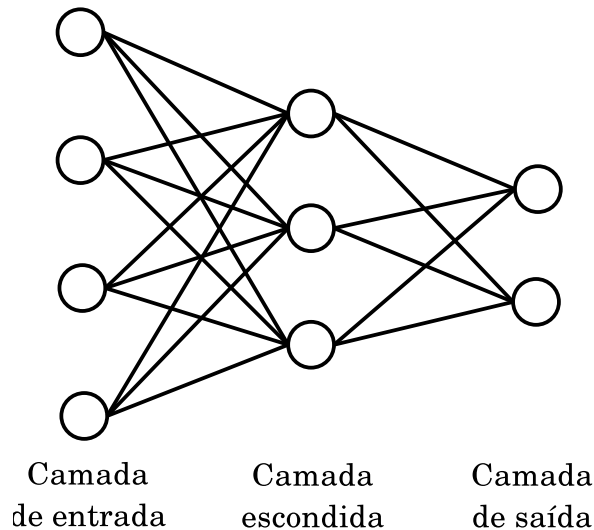


Figura 3.10: Arquitetura de uma rede neuronal com três camadas.

O resultado na saída do neurônio é dado por uma função de ativação não-linear. A função sigmóide é a forma mais comum de função de ativação utilizada na construção de RNAs (HAYKIN, 1999). Um exemplo de função sigmóide é a *função logística*, definida por:

$$\varphi(net_k) = \frac{1}{1 + e^{-Rnet_k}} \quad (3.2)$$

onde R é o parâmetro de inclinação da função sigmóide. Variando-se o parâmetro R é possível obter funções sigmóides com diferentes inclinações. Na Figura 3.11 é ilustrado o modelo do neurônio.

Todo o conhecimento necessário para classificar os padrões de entrada se concentra nos pesos das conexões entre os neurônios da rede. Inicialmente, os pesos mais adequados para um determinado problema são desconhecidos e precisam ser determinados. Caso contrário, a rede não desenvolve a habilidade para realizar o reconhecimento. Treinar a rede consiste no ajuste

dos pesos de forma que todas as saídas produzidas pela mesma representem um erro inferior a uma determinado fator de vigilância (CORREIA, 2005).

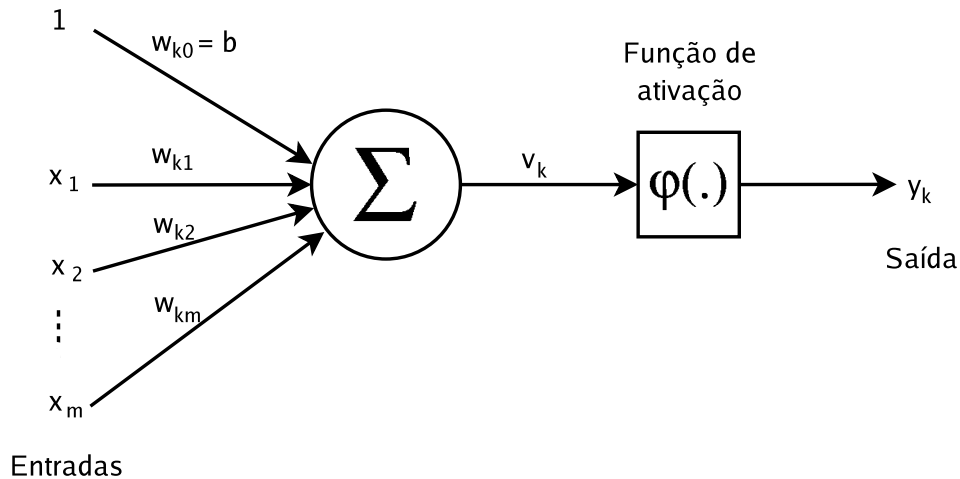


Figura 3.11: Modelo de um neurônio artificial.

O treinamento da rede neste trabalho é do tipo supervisionado (através de exemplos), que requer que cada vetor de entrada esteja associado com um vetor de saídas desejadas. A rede neural é treinada com um conjunto de pares (entrada, saída desejada) denominado *conjunto de treinamento*.

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

O algoritmo utilizado para treinar tais redes é o de retropropagação do erro (*backpropagation*). Este algoritmo é baseado na *regra delta*, proposta por Bernard Widrow e Ted Hoff. Esta regra consiste em minimizar o erro quadrático médio através de um processo de interação local, utilizando os exemplos do conjunto de treinamento (FAUSETT, 1994). A regra delta não estabelece um mecanismo de ajuste dos pesos para os neurônios da camada escondida.

O algoritmo *backpropagation* ou *regra delta generalizada* propõe uma forma de definir o erro dos neurônios das camadas intermediárias, possibilitando o ajuste de seus pesos (BRAGA;

LUDERMIR; CARVALHO, 2000). O treinamento ocorre em duas fases, em que cada fase percorre a rede em um sentido. Estas duas fases são denominadas de *forward* e *backward*. A fase *forward* é utilizada para definir a saída da rede para um dado padrão de entrada. A fase *backward* usa a saída desejada e a saída fornecida pela rede para atualizar os pesos das conexões.

Para o presente estudo foi realizada uma investigação sobre duas arquiteturas de RNAs MLP: Arquitetura Convencional e Arquitetura Classe-Modular, tal como em (OH; SUEN, 2002) e (KAPP et al., 2003). A seguir é descrito um detalhamento sobre esta última arquitetura.

3.3.2 Arquitetura Classe-Modular

Para (OH; SUEN, 2002), as RNAs convencionais, quando projetadas para classificar um grande número de classes com uma estrutura grande de rede, torna-se inevitável determinar os limites ótimos da decisão para todas as classes envolvidas em um espaço extenso de características, isto devido à complexidade do problema como é o caso do reconhecimento de caracteres manuscritos. Para tanto, introduz-se o conceito da modularidade de classes ao classificador da rede neural na tentativa de superar tais limitações. Desta maneira, um problema K -classificação é decomposto em K 2-classificação sub-problemas, um para cada uma das K classes. Assim, cada módulo é responsável por um sub-problema 2-classificação, sendo que a discriminação das amostras é feita em $K - 1$ classes.

De acordo com (OH; SUEN, 2002), a arquitetura convencional tem uma estrutura rígida composta de uma caixa preta em que todas as K classes estão juntas e misturadas. Os módulos não podem ser modificados ou otimizados localmente para cada classe. (PRINCIPE; EULIANO; LEFEBVRE, 1999) e (OH; SUEN, 2002) citam problemas de convergência quando utiliza-se uma rede grande em uma aplicação específica. Um dos problemas que pode ocorrer na convergência é, principalmente, quando um conjunto de treinamento não é grande o suficiente comparado com o tamanho do classificador, isto é, com o número de parâmetros livres no classificador (pesos). Então uma solução seria possuir um conjunto de treinamento tão grande quanto à rede, o que nem sempre é possível ou trivial de se obter.

Na arquitetura classe-modular, o módulo de classificação da linha tradicional de reconhecimento apresentado na Figura 3.12 é substituído por K subredes, M_{wi} para $0 \leq i < K$, cada uma

referente a uma classe. A tarefa específica de cada M_{w_i} é selecionar entre dois grupos de classes, conforme mostrado na Figura 3.13(a). Ω_0 e Ω_1 , com $\Omega_0 = \{w_i\}$ e $\Omega_1 = \{w_k | 0 \leq k < K \text{ e } k \neq i\}$, ou seja, com apenas duas saídas, classificando se determinado exemplo pertence a classe ou não. As redes M_{w_i} foram projetadas da mesma maneira como uma não-modular RNA-MLP mostrada na Figura 3.12.

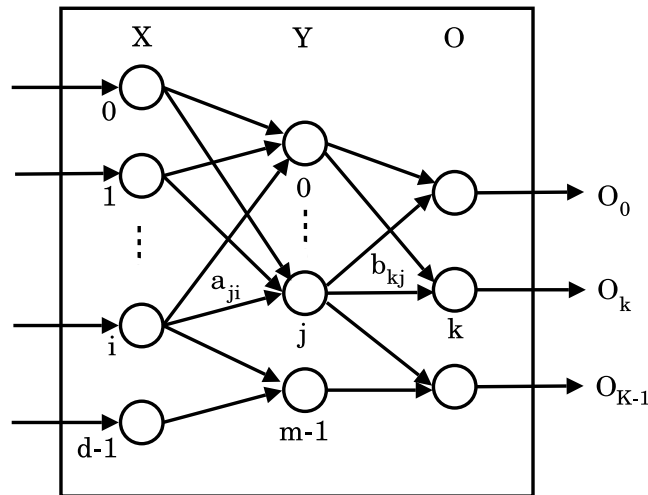
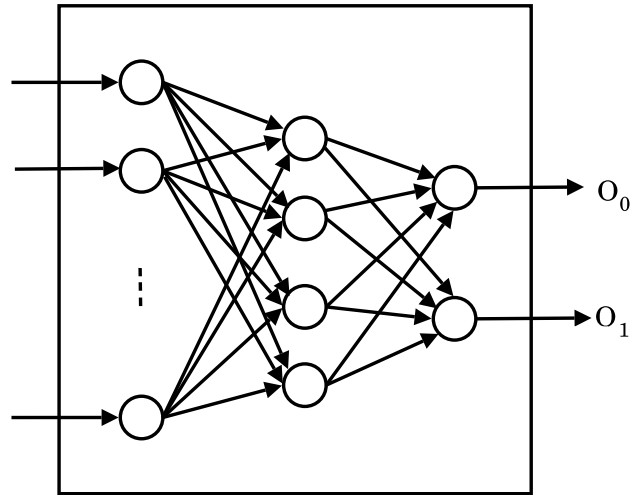


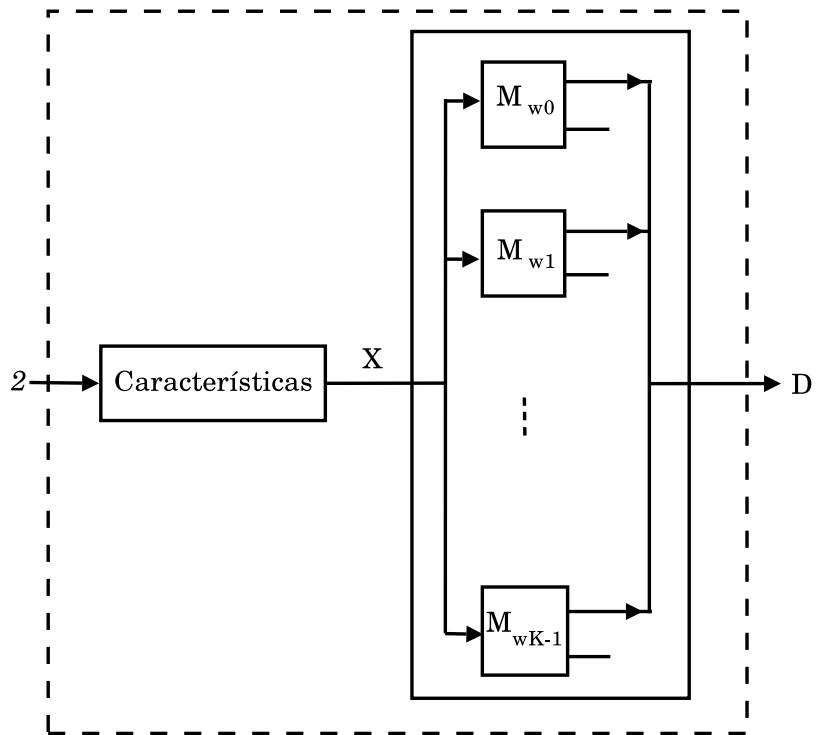
Figura 3.12: Arquitetura convencional onde K classes estão misturadas (OH; SUEN, 2002)

Em cada uma das redes, as três camadas são totalmente conectadas. A camada de entrada tem d nós de entrada para aceitar o vetor de característica d -dimensional, a camada de saída tem dois nós de saída, denotados por O_0 e O_1 para Ω_0 e Ω_1 respectivamente. A arquitetura para a rede inteira formada por K sub-redes é mostrada na Figura 3.13(b). O módulo de extração de características extrai um vetor X que será usado comumente para todas as K classes. X é aplicado para a camada de entrada de todas as sub-redes e cada M_{w_i} efetua os cálculos do processo *forward* usando seu próprio conjunto de pesos para produzir um vetor de saída $D = (O_0, O_1)$. Os valores de O_0 constituem o vetor de decisão final assim como também ocorre em (OH; SUEN, 2002) e (KAPP et al., 2003).

O treinamento e o reconhecimento na arquitetura classe-modular é realizado da seguinte maneira: cada um dos K 2-classificadores é treinado independentemente de outras classes. O algoritmo de aprendizagem retropropagação do erro é aplicado para cada um dos 2-classificadores da mesma forma como ocorre na arquitetura convencional MLP. Os conjuntos de treinamento são preparados para os K 2-classificadores, separando-os em dois grupos, Z_{Ω_0} e Z_{Ω_1} , tais que



(a) Uma subrede M_{wi}



(b) RNA-MLP classe-modular completa

Figura 3.13: Arquitetura para uma RNA-MLP classe-modular (OH; SUEN, 2002)

Z_{Ω_0} contém as amostras das classes em Ω_0 e Z_{Ω_1} contém as restantes para Ω_1 . A mesma separação é feita no conjunto de teste. No segundo estágio, os valores obtidos das saídas das sub-redes são submetidos a uma regra do tipo “vencedor-leva-tudo” (*winner-take-all*) para determinar qual é a classe final, ou seja, a classe vencedora é associada à sub-rede que responde com maior intensidade pelo reconhecimento do padrão. Desta maneira, a classe vencedora responde com o valor de ativação 1 e todas as demais classes com o valor 0. Na Tabela 3.1 é representada a codificação utilizada na saída da rede classe-modular.

Tabela 3.1: Codificação da saída dos caracteres

Classe	Saída
classe vencedora	1 0
demais classes	0 1

Os conjuntos de treinamento e teste para um 2-classificador não são balanceado entre as 2 classes Ω_0 e Ω_1 , o que acarreta em mais exemplos para Ω_1 do que para Ω_0 . Por exemplo, para as amostras da classe do caracter A , Ω_0 representa os exemplos da classe A , enquanto Ω_1 representa os exemplos das demais 25 classes (B a Z).

No trabalho de (OH; SUEN, 2002), os autores apresentam a eficácia das RNAs de arquitetura classe-modular em termos de seu poder da convergência e do reconhecimento. Os testes realizados por estes autores confirmaram o superioridade da rede neural classe-modular, conforme a Tabela 3.2.

Tabela 3.2: Taxa de reconhecimento do conjunto de testes de (OH; SUEN, 2002)

	Número de classes	Arquitetura convencional (%)	Arquitetura classe-modular (%)
Numerais	10	94, 15	97, 30
Letras maiúsculas inglesas	26	81, 03	91, 11
Pares de números	100	57, 06	75, 18
Caracteres Hangul	352	22, 46	68, 75

(KAPP et al., 2003) avalia o uso da arquitetura convencional MLP e classe-modular para o reconhecimento de manuscritos no contexto de cheques bancários. As experiências mostram que a arquitetura classe-modular é melhor do que a arquitetura convencional. As taxas médias

obtidas do reconhecimento foram 77,08% usando a arquitetura convencional e 81,75% usando a classe-modular.

(SILVA; THOMÉ, 2003) apresentam a utilização de uma rede neural para a classificação dos caracteres da base de dados do Núcleo de Computação Eletrônica da Universidade Federal do Rio de Janeiro, em um “*time*” de redes neurais, que podem ser de mesmo modelo ou de modelos diferentes trabalhando juntas ou em cascata em um mesmo problema. A finalidade do *time* é dividir o problema em sub-problemas específicos e atribuir a cada rede integrante a tarefa de resolver um sub-problema. Depois de resolver todos os sub-problemas suas respostas são combinadas de maneira a produzir a resposta para o problema original. Os autores consideram que, os sub-problemas específicos são menores que o problema original, logo, as redes neurais que atuam nos mesmos são menores. Sendo assim, espera-se que estas redes menores tenham um bom poder de generalização para seus sub-problemas e que quando unidas alcancem um poder de generalização maior do que o alcançado por apenas uma rede. Com esta estratégia foi obtido um desempenho de 91,91%.

(AIRES, 2005) investiga mecanismos de zoneamento perceptivo para reconhecimento de caracteres manuscritos. Trata-se de um mecanismo de zoneamento não simétrico baseado na análise das matrizes de confusão dos classificadores individuais (Classe-Modular). No processo de reconhecimentos utilizam-se RNAs MLP Classe-Modular. Uma metodologia para múltiplos classificadores é aplicada ao problema de reconhecimento. A metodologia contempla o estudo de medidas de diversidade e discordância para buscar uma alternativa de combinação de classificadores, sem basear-se unicamente nas taxas de reconhecimento. As taxas médias de reconhecimento obtidas, para os zoneamentos avaliados, foram as seguintes: zoneamento tipo 4 = 82,89%, zoneamento tipo 5Horizontal = 81,75%, zoneamento tipo 5Vertical = 80,94% e zoneamento tipo 7 = 84,73%. As combinações realizadas entre os classificadores individuais apresentam uma melhoria na taxa de reconhecimento, sendo de 85,9% para a rede 5Horizontal-5Vertical-7. A base de dados utilizada foi a IRONOFF, especificamente os caracteres manuscritos alfabéticos.

Deste modo, conclui-se que é evidente a eficácia das redes neurais de arquitetura classe-modular em relação a sua capacidade de reconhecimento. A fim de ratificar tal afirmação, os

experimentos e os resultados obtidos neste trabalho são discutidos a seguir.

Capítulo 4

Experimentos e Resultados

Neste capítulo são apresentados os experimentos realizados e os resultados obtidos pela abordagem proposta com o objetivo de investigar sua eficiência. O capítulo está organizado da seguinte forma: na Seção 4.1 são descritas as bases de dados utilizadas; na Seção 4.2 é apresentado o simulador SNNS, o qual foi utilizado para a implementação, treinamento e teste das redes neurais deste trabalho; e por fim, na Seção 4.3, são apresentados os experimentos realizados e são analisados os resultados alcançados.

4.1 Base de dados

A avaliação dos métodos de extração de características/classificação utilizados foi feita principalmente com a base de caracteres manuscritos IRONOFF da Universidade de Nantes na França (VIARD-GAUDIN et al., 1999). Entretanto, testes preliminares foram realizados com a base *Letter* do repositório de aprendizado de máquina da Universidade da Califórnia em Irvine (BLAKE; MERZ, 1998), a fim de investigar somente o desempenho do sistema de classificação, já que nesta base a etapa de extração de características já foi realizada.

Letter

O repositório de aprendizado de máquina da Universidade da Califórnia, em Irvine, possui dezenas de bases de dados, a maior parte com permissão de acesso pública, coletadas a partir de diversas fontes sendo associadas a muitos problemas distintos, tais como: reconhecimento de

caracteres *on-line* e *off-line*, análise de crédito, jogos (como xadrez ou jogo-da-velha), dentre outros. Todas as bases deste repositório contém padrões pré-processados e todos os atributos e classes são bem documentados, o que demanda apenas que seja realizada uma etapa preliminar para condicionar os dados ao formato de entrada desejado. Portanto, não é necessário realizar extração de características, mas é preciso converter as entradas em números, uma vez que em várias bases estas entradas são nominais.

Para a realização dos experimentos utilizou-se deste repositório a base de dados *Letter*, que permitiu desenvolver experimentos preliminares com redes classe-modular dispensando a etapa de extração de características utilizando uma base com 26 classes com muitas amostras de treinamento e teste. Esta base é formada por 20000 vetores de características extraídas de imagens de caracteres manuscritos, das quais as primeiras 16000 são usadas para treinamento e as 4000 últimas para teste. As classes dos padrões associam-se às 26 letras do alfabeto inglês, conforme a Tabela 4.1, e os atributos a 16 características extraídas de imagens da varredura do caracter. A documentação que acompanha a base referencia um artigo publicado em 1991, que obteve como resultado uma taxa de reconhecimento um pouco acima de 80%. Em seu trabalho, Matos (MATOS, 2004) alcançou para a base *Letter* a taxa de reconhecimento de 95,67%.

IRONOFF

A base de dados IRONOFF foi produzida pela IRESTE (Universidade de Nantes - França) em colaboração com VISION OBJECTS, onde foram coletadas duas bases de manuscritos. Para cada caractere e/ou palavra inseridos na base de dados, são verificados sinais *on-line* e *off-line*, que podem ser utilizados em diversos métodos de reconhecimento manuscrito *on-line* e *off-line*.

A base de dados IRONOFF contém:

- 4086 dígitos isolados
- 10685 caracteres minúsculos isoladas
- 10681 caracteres maiúsculos isoladas + 410 símbolos EURO
- 31346 palavras isoladas (28657 palavras francesas + 2689 palavras inglesas)

Tabela 4.1: Distribuição de caracteres - Base de dados *Letter*

Classe	Amostras	Treino	Teste
A	789	633	156
B	766	630	136
C	736	594	142
D	805	638	167
E	768	616	152
F	775	622	153
G	773	609	164
H	734	583	151
I	755	590	165
J	747	599	148
K	739	593	146
L	761	604	157
M	792	648	144
N	783	617	166
O	753	614	139
P	803	635	168
Q	783	615	168
R	758	597	161
S	748	587	161
T	796	644	152
U	813	646	167
V	764	628	136
W	752	613	139
X	787	628	159
Y	786	641	145
Z	734	576	158
Σ	20000	16000	4000

Os dados foram coletados a partir de aproximadamente 700 escritores diferentes, sendo a maioria de nacionalidade francesa. Cerca de $\frac{2}{3}$ dos escritores é do sexo masculino, sendo 90% destros. O escritor mais novo tem 8 anos de idade e o mais velho tem 77 anos.

Para este estudo foram utilizados arquivos da base *off-line* de caracteres minúsculos e maiúsculos. Esta base de dados está dividida em um léxico de 26 caracteres conforme apresentado nas Tabelas 4.2 e 4.3. As taxas de reconhecimento citadas na literatura para esta base foram apresentadas na Tabela 2.1.

Tabela 4.2: Distribuição de caracteres minúsculos - Base IRONOFF

Classe	Amostras	Treino	Teste
A	412	278	134
B	411	277	134
C	411	272	139
D	411	286	125
E	412	274	138
F	410	271	139
G	409	274	135
H	411	279	132
I	410	273	137
J	409	271	138
K	411	274	137
L	410	269	141
M	409	267	142
N	412	273	139
O	412	265	147
P	412	288	124
Q	412	248	164
R	412	273	139
S	412	278	134
T	410	265	145
U	410	279	131
V	411	275	136
W	412	280	132
X	411	280	131
Y	412	273	139
Z	411	282	129
Σ	10685	7124	3561

Tabela 4.3: Distribuição de caracteres maiúsculos - Base IRONOFF

Classe	Amostras	Treino	Teste
A	411	276	135
B	412	270	142
C	412	274	138
D	411	283	128
E	412	279	133
F	412	289	123
G	412	282	130
H	411	276	135
I	411	273	138
J	411	264	147
K	410	280	130
L	411	290	121
M	412	261	151
N	411	275	136
O	410	276	134
P	412	275	137
Q	408	282	126
R	410	273	137
S	409	262	147
T	412	276	136
U	411	266	145
V	410	277	133
W	409	272	137
X	411	274	137
Y	411	254	157
Z	409	270	139
Σ	10681	7129	3552

4.2 Simulador SNNS

Para a realização dos experimentos foi utilizado um simulador de redes neurais de distribuição livre¹, desenvolvido na Universidade de Stuttgart, na Alemanha, denominado *SNNS* (*Stuttgart Neural Network Simulator*).

Basicamente o projeto desta ferramenta busca alcançar eficiência e alta flexibilidade no projeto e aplicação de redes neurais, integrados em um só ambiente de simulação.

O simulador SNNS é formado por quatro componentes principais: o *kernel* do simulador, a interface gráfica com o usuário, a interface de execução em modo *batch* (*batchman*) e o compilador de redes *snns2c*. O *kernel* opera sobre uma representação interna das redes neurais e é responsável por todas as operações e pelas estruturas de dados que as compõem. A interface gráfica trabalha sobre o *kernel*, fornecendo uma representação gráfica das redes neurais e controla o *kernel* durante a simulação. O simulador permite gerenciar a implementação de uma rede neural através de um painel principal denominado *SNNS Manager*. Este painel possibilita o acesso à todas as funcionalidades disponíveis no simulador. O SNNS suporta cinco tipos de arquivos, dos quais os três mais importantes e que foram utilizados neste trabalho são:

- *.net*: arquivos que descrevem informações sobre a topologia da rede e regras de aprendizado;
- *.pat*: arquivos que contêm os padrões de treinamento e de teste;
- *.res*: arquivos de resultados que permitem ao usuário observar a saída da rede para cada tipo de padrão testado.

O simulador SNNS é utilizado neste trabalho por ser conhecido pela comunidade científica, por se tratar de um *software* livre, e principalmente por ser apropriado para realizar processamentos intensivos em modo *batch*, que são úteis para realizar eficientemente o treinamento de redes neurais complexas.

¹Disponível em <www-ra.informatik.uni-tuebingen.de/SNNS/>

4.3 Experimentos

Para ambas as bases descritas na seção 2.1 são discutidos os experimentos e os resultados em partes, ou seja, inicialmente são apresentados os experimentos e os resultados para a arquitetura convencional e logo em seguida para a arquitetura classe-modular.

4.3.1 Experimentos para a Arquitetura Convencional

Letter

O treinamento da rede neural foi realizado usando uma RNA MLP com uma camada de entrada de 16 neurônios, uma camada escondida com 64 neurônios e uma camada de saída com 26 neurônios, em que cada saída equivale a um caracter do léxico. O algoritmo de aprendizado foi o *backpropagation* padrão, com taxa de aprendizagem 0,02. Os pesos foram inicializados aleatoriamente com valores no intervalo $[-1, 1]$ e suas atualizações realizadas de forma topológica da entrada para a saída. A função de transferência utilizada para os neurônios da camada escondida e da camada de saída foi a função logística ou sigmóide. O treinamento foi efetuado em 1000 épocas. A taxa de reconhecimento obtida foi 83,10%.

IRONOFF

Para os conjuntos de caracteres minúsculos e maiúsculos foi utilizada a mesma RNA MLP. A camada de entrada é formada por 375 neurônios, que correspondem ao número de posições do vetor de características, a camada escondida possui 256 neurônios e a camada de saída 26 neurônios. O algoritmo de aprendizado foi o *backpropagation* padrão, com taxa de aprendizagem 0,02. Os pesos foram inicializados aleatoriamente com valores no intervalo $[-1, 1]$ e suas atualizações realizadas de forma topológica da entrada para a saída. A função de transferência utilizada para os neurônios da camada escondida e da camada de saída foi a função logística ou sigmóide. O treinamento foi efetuado em 1000 épocas. A taxas de reconhecimento para caracteres minúsculos foi 56,14% e para caracteres maiúsculos foi 63,49%.

4.3.2 Experimentos para a Arquitetura Classe-modular

Letter

O processo de obtenção de resultados para a arquitetura classe-modular segue como descrito na Seção 3.5, onde utiliza-se k RNA MLP, uma rede para cada caracter do léxico do problema. Cada um dos k 2-classificadores é treinado independentemente das outras classes, utilizando um conjunto de treinamento distinto para cada rede. O algoritmo *backpropagation* padrão é usado em cada um dos 2-classificadores da mesma maneira como ocorre na arquitetura convencional.

Cada um dos k 2-classificadores é composto por 16 neurônios de entrada, 64 neurônios na camada escondida e 2 neurônios de saída. Os pesos foram inicializados aleatoriamente com valores entre -1 e 1 e suas atualizações realizadas de forma topológica da entrada para a saída. A função de transferência utilizada para os neurônios da camada escondida e da camada de saída foi a função logística ou sigmóide. A taxa de aprendizagem aplicada foi $0,02$ e o treinamento foi efetuado em até no máximo 100 épocas.

As medidas utilizadas para avaliar desempenho do classificador foram a sensibilidade (*sens*) e a especificidade (*spec*). Essas medidas são extraídas a partir da matriz de confusão de um classificador binário, que é o caso, como é apresentada na Tabela 4.4. A *sensibilidade* equivale à taxa de exemplos positivos classificados corretamente e a *especificidade* à taxa de exemplos negativos classificados corretamente. Essas medidas podem ser definidas de acordo com as seguintes equações:

$$sens = \frac{T_P}{T_P + F_N} \quad (4.1)$$

$$spec = \frac{T_N}{T_N + F_P} \quad (4.2)$$

onde F_P e F_N são os dois erros possíveis denominados *falso positivo* e *falso negativo*, respectivamente; T_P é o número de exemplos positivos classificados corretamente e T_N é o número de exemplos negativos classificados corretamente do total de $n = (T_P + F_N + F_P + T_N)$ exemplos.

Na Tabela 4.5 é mostrado as taxas de sensibilidade e especificidade obtidas após a submissão do conjunto de teste à arquitetura proposta. É sobressalente na Tabela 4.5 as altas taxas da

Tabela 4.4: Matriz de confusão para a classificação com duas classes (REZENDE, 2003)

Classe	Predita C_+	Predita C_-	Taxa de erro da classe	Taxa de erro total
verdadeira C_+	Verdadeiros positivos T_P	Falsos negativos F_N	$\frac{F_N}{T_P+F_N}$	$\frac{F_P+F_N}{n}$
verdadeira C_-	Falsos positivos F_P	Verdadeiros negativos T_N	$\frac{F_P}{F_P+T_N}$	

especificidade, isto ocorre devido ao desbalanceamento das amostras por classe, ou seja, o número de exemplos negativos é extremamente superior ao número de exemplos positivos.

A saída do classificador global é obtida realizando-se a combinação das saídas de cada classificador individualmente. Isto é, um mesmo padrão de teste de cada rede e a saída de cada uma destas é apresentada ao combinador que realiza uma predição única. Neste trabalho, por não haver tempo hábil para investigar uma estratégia mais acurada, utilizou-se a regra “vencedor-leva-tudo” na qual a classe vencedora é associada à rede que responde com maior intensidade pelo reconhecimento do padrão. A taxa de reconhecimento global alcançada para a base *Letter* foi 93,67%. Observa-se que ocorreu uma melhora significativa na taxa de reconhecimento, em relação ao experimento realizado com a arquitetura convencional, mesmo tendo sido utilizado uma estratégia “fraca” no módulo de combinação. Nota-se também que o desempenho global do sistema é fortemente influenciado pela sensibilidade dos classificadores individuais que o compõe.

IRONOFF

Para a realização dos experimentos com a base IRONOFF utilizou-se uma rede MLP com uma camada de entrada com 375 neurônios, uma camada escondida com 256 neurônios e uma camada de saída com 2 neurônios. O algoritmo de aprendizagem foi o *backpropagation* padrão, com taxa de aprendizagem 0,02. Os pesos foram inicializados aleatoriamente com valores entre -1 e 1 e suas atualizações realizadas de forma topológica da entrada para a saída. O treinamento foi efetuado em até no máximo 250 épocas.

Tabela 4.5: Taxas de sensibilidade e especificidade para a base *Letter*

Classe	Sens. (%)	Espec. (%)
A	96,13	99,97
B	93,38	99,66
C	73,24	100,00
D	91,62	99,71
E	85,53	99,71
F	89,54	99,82
G	90,85	99,50
H	83,44	99,51
I	91,52	99,97
J	89,19	99,79
K	86,30	99,45
L	92,36	99,95
M	95,83	99,90
N	89,76	99,82
O	86,33	99,87
P	90,48	99,82
Q	91,07	99,71
R	85,09	99,66
S	86,34	99,84
T	93,42	99,90
U	98,20	99,79
V	89,71	100,00
W	97,84	99,97
X	93,71	99,74
Y	91,72	99,92
Z	92,41	99,97
Média	90,19	99,81

As mesmas medidas adotadas para avaliar o desempenho do classificador aplicado na base *Letter* foram utilizadas para base IRONOFF. Sendo assim, na Tabela 4.6 é mostrado as taxas de sensibilidade e especificidade obtidas após a submissão do conjunto de teste dos caracteres minúsculos e maiúsculos à arquitetura classe-modular. A taxa de reconhecimento global obtida para os caracteres minúsculos foi 80,75% e para os caracteres maiúsculos foi 89,21%.

Tabela 4.6: Taxas de sensibilidade e especificidade de caracteres minúsculos e maiúsculos da base IRONOFF

Classe	Caracteres minúsculos		Caracteres maiúsculos	
	Sens. (%)	Espec. (%)	Sens. (%)	Espec. (%)
A	76,87	99,97	78,03	100,00
B	81,34	99,59	76,76	100,00
C	65,47	100,00	81,16	100,00
D	55,20	100,00	78,91	100,00
E	52,90	100,00	76,69	100,00
F	89,93	99,06	79,67	100,00
G	65,93	99,53	89,23	99,50
H	64,39	99,39	81,48	100,00
I	88,46	99,82	86,89	99,79
J	79,71	99,56	84,35	99,76
K	64,96	99,71	79,23	99,82
L	85,29	99,12	94,21	99,88
M	51,41	100,00	86,09	99,59
N	76,98	99,12	86,09	99,59
O	59,86	100,00	85,07	100,00
P	50,81	100,00	85,07	100,00
Q	87,20	99,62	84,92	99,85
R	62,59	99,82	84,92	99,85
S	71,64	99,71	87,76	99,85
T	68,97	99,32	87,76	99,85
U	73,28	99,24	83,45	99,73
V	77,21	99,59	83,45	99,73
W	87,88	99,68	91,97	99,85
X	77,10	99,82	91,97	99,85
Y	69,57	99,35	85,90	99,82
Z	44,19	100,00	85,90	99,82
Média	70,35	95,84	84,29	99,86

4.3.3 Análise dos resultados

Conforme os resultados experimentais obtidos para a arquitetura convencional e a modular, duas conclusões podem ser feitas referentes aos desempenhos das arquiteturas:

- A arquitetura classe-modular teve superioridade em termos de término de convergência em relação à convencional. Isto porque o tempo de convergência para o aprendizado de 26 classes, como é o caso da rede convencional, é perceptivelmente maior que o tempo de convergência para o aprendizado de 2 classes, como é o caso de cada rede classe-modular.
- A arquitetura classe-modular também foi superior em termos de capacidade de reconhecimento do que a arquitetura convencional neste trabalho, devido ao fato de cada rede classe-modular possuir apenas 2 classes, positivas e negativas (por exemplo, classe A e a classe não A), sendo assim a capacidade de reconhecimento desta rede é superior a rede convencional, uma vez que esta rede possui 26 classes (A a Z).

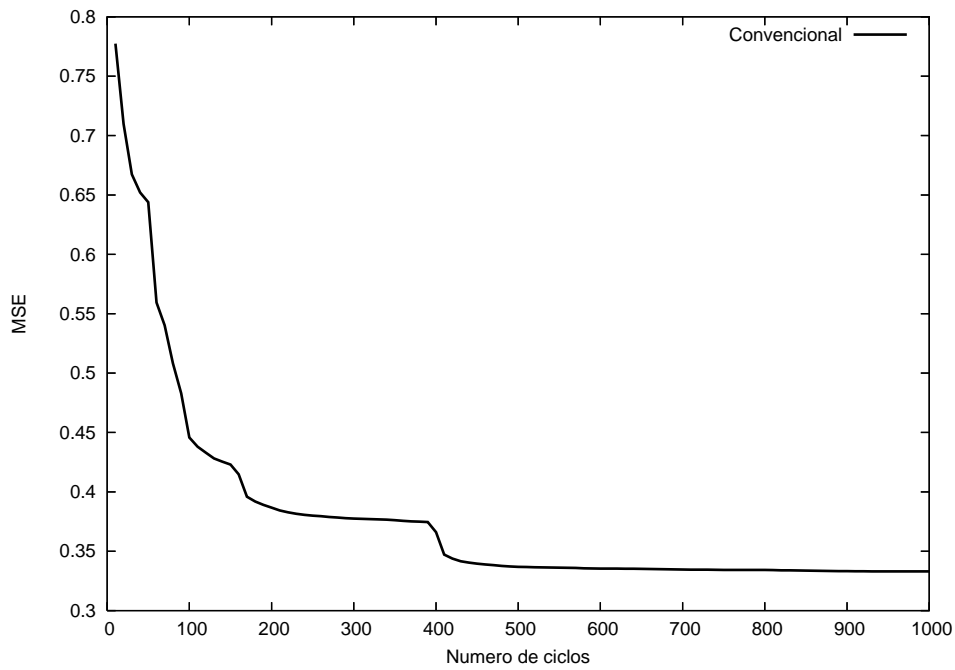
A superioridade da rede classe-modular em relação à rede convencional quanto ao término de convergência e a capacidade de reconhecimento pode ser observada nos gráficos da evolução do MSE (*Mean Square Error*) no treinamento da rede convencional da base IRONOFF (Figura 4.1(a)) e da rede classe-modular (Figura 4.1(b)), neste caso é ilustrado o gráfico do treinamento do caracter X maiúsculo.

O MSE ou *erro médio quadrático* é obtido pela diferença entre a saída produzida pela rede e a saída desejada. Matematicamente, esta medida é definida pela seguinte equação:

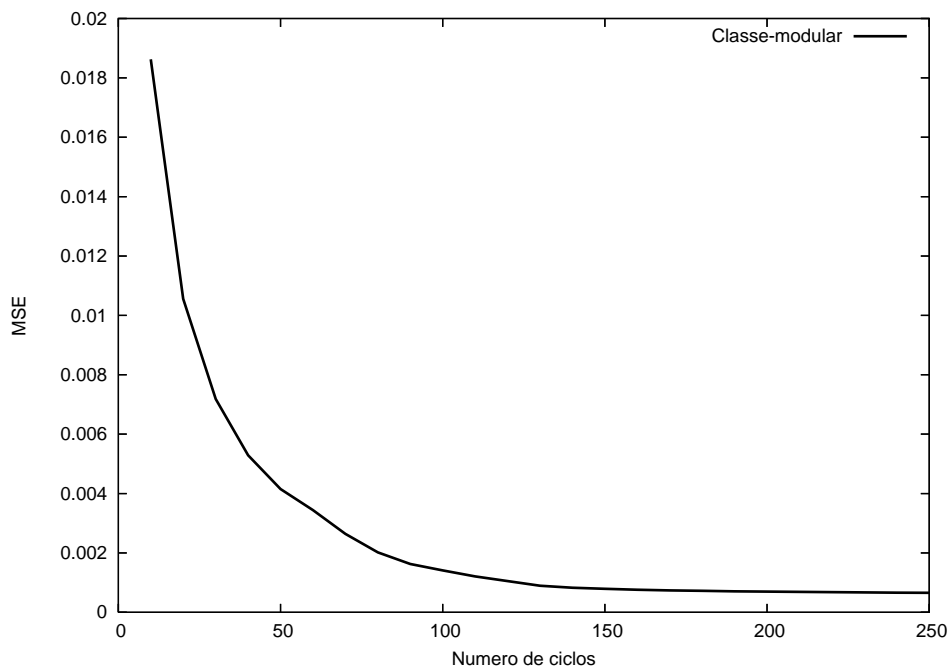
$$MSE = \frac{1}{N} \sum_{i=1}^n (d_i - y_i)^2 \quad (4.3)$$

onde d_i e y_i representam respectivamente as saídas desejadas e obtidas da rede neural e N representa o número total de padrões.

É importante também ressaltar que o treinamento na arquitetura classe-modular pode ser facilmente paralelizável, ou seja o treinamento de cada rede modular pode ser realizado de forma independente, o que não ocorre na arquitetura convencional, uma vez que existe apenas uma rede com todas as classes misturadas.



(a) Rede convencional



(b) Rede classe-modular

Figura 4.1: Gráficos da evolução do MSE no treinamento das redes em função do número de ciclos

A Tabela 4.7 resume os resultados obtidos neste trabalho e alguns outros estudos relacionados às bases de dados utilizadas.

Tabela 4.7: Quadro comparativo do trabalho proposto e estudos relacionados

Trabalhos	<i>Letter</i>	Caracteres minúsculos	Caracteres maiúsculos
(POISSON; VIARD-GAUDIN; LALLICAN, 2002)	-	80,5%	89,9%
(AIRES, 2005)	-	-	85,9%
(MATOS, 2004)	95,6%	-	-
Arquitetura convencional	83,1%	56,1%	63,5%
Arquitetura classe-modular	93,6%	80,7%	89,2%

Deve-se chamar a atenção para os dados desta tabela, uma vez esta é a parte mais “clamatória” do trabalho.

Convém mencionar que os resultados obtidos foram competitivos em relação ao apresentado na literatura. O bom desempenho se deve tanto ao fato da arquitetura classe-modular ter sido apropriada para esse tipo de problema, quanto ao bom desempenho do extrator de características.

Capítulo 5

Conclusões

A proposta deste trabalho foi avaliar a aplicação de um mecanismo de zoneamento bem como de uma arquitetura classe-modular na classificação de caracteres manuscritos. Observou-se a dificuldade da realização desta tarefa principalmente devido à grande variabilidade de estilos de escrita encontrados.

O método de histogramas direcionais com zoneamento desenvolvido por (MATOS, 2004) foi empregado para a realização da extração de características. Esta abordagem fundamenta-se nas características visuais humanas, como as singularidades existentes no contorno da forma. A visão do ser humano é um processo fortemente influenciado por características espaciais e temporais tais como contorno e pontos de singularidade. Matos trabalhou com imagens do contorno por admitir que nele esteja localizada a informação mais relevante para a realização do reconhecimento.

Em relação aos estudos das arquiteturas de RNA MLP convencional e classe-modular, os resultados obtidos indicam que esta pesquisa é promissora e prova ser merecedora de investigações adicionais do paradigma de modularidade de classes. Observou-se também que a arquitetura convencional tem uma estrutura rígida composta de uma “caixa preta” na qual todas as classes K são misturadas completamente. Os módulos não podem ser modificados ou aperfeiçoados localmente para cada classe, como pode ocorrer na classe-modular. Porém, a principal desvantagem da arquitetura classe-modular é justamente o treinamento de K redes para as classes do problema.

Os resultados alcançados mostram ser competitivos, ou seja, apresentam taxas de reconhe-

cimentos próximas a de outros trabalhos mencionados na literatura (POISSON; VIARD-GAUDIN; LALLICAN, 2002) (MATOS, 2004) (AIRES, 2005) que empregam, às vezes, vários tipos de características complementares e combinação de classificadores. As melhores taxas de reconhecimento alcançadas para base IRONOFF foi 89,21% para os caracteres maiúsculos e 80,75% para os caracteres minúsculos e para a base *Letter* foi 93,67%. Esses resultados são bem próximos, ou até superiores, às maiores taxas de reconhecimento citadas na literatura.

Para a realização do presente estudo, podem-se citar como principais dificuldades os seguintes entraves:

- Efetuar o treinamento das redes classe-modular em local favorável. Como o ambiente utilizado para o treinamento foi o mesmo em que ocorrem as aulas do Departamento de Ciência da Computação, o treinamento era impraticável durante a semana. Além da indisponibilidade do local, problemas como falta de energia e desligamentos acidentais das máquinas durante o processo de treinamento atrasaram a conclusão do trabalho.
- Desencontros em relação ao recebimento do material necessário para a realização da extração de características ocasionaram mudanças na proposta inicial do trabalho e atrasos no cronograma determinado. Este problema foi contornado com a utilização do método de extração de características desenvolvido por (MATOS, 2004).

5.1 Perspectivas de trabalhos futuros

Durante o desenvolvimento deste trabalho, não se teve a oportunidade de pesquisar ou testar alguns assuntos devido ao curto intervalo de tempo previamente estabelecido. O presente trabalho pode ser continuado explorando-se os seguintes aspectos:

- Utilizar uma camada *softmax* nas redes neurais a fim de aproximar a saída das redes de uma distribuição de probabilidades, com isso pode ser aplicada uma estratégia mais sofisticada na combinação das saídas dos classificadores.
- Empregar um método de combinação de classificadores estatístico. A combinação de classificadores baseada na regra da soma apresenta taxas de acerto superior às obtidas por

classificadores individuais. (KITTLER et al., 1998) avaliou diversas regras de combinação fixas, sendo que os melhores resultados foram obtidos com a regra da soma. Outros trabalhos também utilizaram a regra da soma com sucesso, tais como (AIRES, 2005).

- Investigar o uso da imagem afinada no processo de extração de características. A técnica de afinamento é o método de redução de uma forma para uma versão simplificada que retém as características do objeto original. Os algoritmos de afinamento excluem, de forma sucessiva, diversas camadas da extremidade (borda) de um padrão até que permaneça apenas uma representação simples do traçado, geralmente com um *pixel* de largura.
- Aplicar algoritmos genéticos para otimizar o zoneamento para caracteres manuscritos, como sugerido no trabalho de (RADTKE et al., 2003).
- Formular e testar hipóteses que expliquem a superioridade da arquitetura classe-modular.

Referências Bibliográficas

AHMAD, A. R. et al. Online handwriting recognition using support vector machine. In: *IEEE Region 10 Conference TENCN 2004*. [S.l.: s.n.], 2004. v. 1, p. 311–314.

AIRES, S. B. K. *Reconhecimento de caracteres manuscritos baseado em regiões perceptivas*. Dissertação (Mestrado) — PUCPR, PPGIA, Curitiba, 2005.

ARICA, N.; YARMAN-VURAL, F. T. An overview of character recognition focused on off-line handwriting. *IEEE Transactions on Systems, Man, and Cybernetics – Part C: Applications and Reviews*, v. 31, n. 2, 2001.

BLAKE, C.; MERZ, C. *UCI repository of machine learning databases*. [S.l.: s.n.], 1998.

BRAGA, A. de P.; LUDERMIR, T. B.; CARVALHO, A. C. P. de L. F. *Redes Neurais Artificiais: Teoria e aplicações*. Rio de Janeiro: LTC – Livros Técnicos e Científicos, 2000. 262 p. ISBN 85-216-1218-4.

CAI, J.; LIU, Z.-Q. Integration of structural and statistical information for unconstrained handwritten numeral recognition. *IEEE Trans. Pattern Anal. Mach. Intell.*, IEEE Computer Society, Washington, DC, USA, v. 21, n. 3, p. 263–270, 1999. ISSN 0162-8828.

CAILLAULT, E.; VIARD-GAUDIN, C.; AHMAD, A. Ms-tdnn with global discriminant trainings. In: *Proceedings. Eighth International Conference on Document Analysis and Recognition, 2005*. [S.l.: s.n.], 2005. v. 2, p. 856–860.

CORREIA, S. É. N. *Reconhecimento de caracteres manuscritos usando wavelets*. Tese (Doutorado) — Universidade Federal de Campina Grande, Departamento de Engenharia Elétrica, Paraíba, 2005.

- DONG, J. *Comparison of Algorithms for Handwritten Numeral Recognition*. Concordia University, Montreal Quebec H3G 1M8 Canada, 1999.
- EL-YACOUBI, A. et al. An hmm-based approach for off-line unconstrained handwritten word modeling and recognition. *IEEE Trans. Pattern Anal. Mach. Intell.*, IEEE Computer Society, Washington, DC, USA, v. 21, n. 8, p. 752–760, 1999. ISSN 0162-8828.
- FECHINE, J. M.; MADEIRO, F.; NETO, B. G. A. Reconhecimento automático da identidade vocal baseado em quantização vetorial e hmm. *Conferencia Iberoamericana em Sistemas, CISCI*, Orlando, USA, 2003.
- FREITAS, C. O. de A.; BORTOLOZZI, F.; SABOURIN, R. Handwritten isolated word recognition: An approach based on mutual information for feature set validation. In: *ICDAR*. [S.l.]: IEEE Computer Society, 2001. p. 665–669. ISBN 0-7695-1263-1.
- GONZALEZ, R. C.; WOODS, R. E. *Processamento de Imagens Digitais*. 1. ed. São Paulo: Edgar Blücher Ltda, 2000. 528 p. ISBN 85-212-0264-4.
- HAYKIN, S. *Neural networks: A comprehensive foundation*. 2. ed. New York: Prentice Hall, 1999. 528 p. ISBN 0-13-273350-1.
- HORNIK, K. Approximation capabilities of multilayer feedforward networks. *Neural Netw.*, Elsevier Science Ltd., Oxford, UK, UK, v. 4, n. 2, p. 251–257, 1991. ISSN 0893-6080.
- HORNIK, K.; STINCHCOMBE, M.; WHITE, H. Multilayer feedforward networks are universal approximators. *Neural Netw.*, Elsevier Science Ltd., Oxford, UK, UK, v. 2, n. 5, p. 359–366, 1989. ISSN 0893-6080.
- JAIN, A. K.; DUIN, R. P. W.; MAO, J. Statistical pattern recognition: A review. *IEEE Trans. Pattern Anal. Mach. Intell.*, v. 22, n. 1, p. 4–37, 2000.
- KAPP, M. N. *Reconhecimento de palavras manuscritas utilizando redes neuronais artificiais*. Dissertação (Mestrado) — PUCPR, PPGIA, Curitiba, 2005.

KAPP, M. N. et al. Evaluating the conventional and class-modular architectures feedforward neural network for handwritten word recognition. In: *SIBGRAPI*. [S.l.]: IEEE Computer Society, 2003. p. 315–322. ISBN 0-7695-2032-4.

KITTLER, J. et al. On combining classifiers. *IEEE Trans. Pattern Anal. Mach. Intell.*, IEEE Computer Society, Washington, DC, USA, v. 20, n. 3, p. 226–239, 1998. ISSN 0162-8828.

KOERICH, A. L.; SABOURIN, R.; SUEN, C. Y. Large vocabulary off-line handwriting recognition: A survey. *Pattern Analysis and Applications*, v. 6, n. 2, p. 97–121, 2003.

LALLICAN, P.; VIARD-GAUDIN, C.; KNERR, S. *From Off-line to On-line handwriting Recognition*. 2000. Disponível em: <citeseer.ist.psu.edu/lallican00from.html>.

LIU, C.-L. et al. Handwritten digit recognition: benchmarking of state-of-the-art techniques. *Pattern Recognition*, v. 36, n. 10, p. 2271–2285, 2003.

LIU, C.-L.; SAKO, H.; FUJISAWA, H. Performance evaluation of pattern classifiers for handwritten character recognition. *IJDAR*, v. 4, n. 3, p. 191–204, 2002.

MATOS, L. N. *Utilização de redes bayesianas como agrupador de classificadores locais e global*. Tese (Doutorado) — Universidade Federal de Campina Grande, Departamento de Engenharia Elétrica, 2004.

OH, I.-S.; SUEN, C. Y. A class-modular feedforward neural network for handwriting recognition. *Pattern Recognition*, v. 35, n. 1, p. 229–244, 2002.

OLIVEIRA, L. E. S. de et al. Automatic recognition of handwritten numerical strings: A recognition and verification strategy. *IEEE Trans. Pattern Anal. Mach. Intell.*, v. 24, n. 11, p. 1438–1454, 2002.

PLAMONDON, R.; SRIHARI, S. N. On-line and off-line handwriting recognition: A comprehensive survey. *IEEE Trans. Pattern Anal. Mach. Intell.*, IEEE Computer Society, Washington, DC, USA, v. 22, n. 1, p. 63–84, 2000. ISSN 0162-8828.

POISSON, E.; VIARD-GAUDIN, C.; LALLICAN, P.-M. Multi-modular architecture based on convolutional neural networks for online handwritten character recognition. In: *Proceedings of*

the 9th International Conference on Neural Information Processing. [S.l.: s.n.], 2002. v. 5, p. 2444–2448.

PRINCIPE, J. C.; EULIANO, N. R.; LEFEBVRE, W. C. *Neural and Adaptive Systems: Fundamentals through Simulations*. New York, NY, USA: John Wiley & Sons, Inc., 1999. ISBN 0471351679.

RADTKE, P. V. W. et al. Intelligent zoning design using multi-objective evolutionary algorithms. *icdar*, IEEE Computer Society, Los Alamitos, CA, USA, v. 02, p. 824, 2003.

REZENDE, S. O. *Sistemas Inteligentes: Fundamentos e Aplicações*. 1^a. ed. Barueri, São Paulo: [s.n.], 2003. ISBN 85-204-1683-7.

RICHARD, M. D.; LIPPMANN, R. Neural network classifiers estimate bayesian a posteriori probabilities. *Neural Comput.*, p. 13–16, 1991.

SHI, M. et al. Handwritten numeral recognition using gradient and curvature of gray scale image. *Pattern Recognition*, v. 35, n. 10, p. 2051–2059, 2002.

SHIOYAMA, T.; WU, H. Y.; NOJIMA, T. Recognition algorithm based on wavelet transform for handprinted chinese characters. In: *14th International Conference Pattern Recognition*. [S.l.: s.n.], 1998. v. 1, p. 229–232.

SILVA, E.; THOMÉ, A. C. G. Reconhecimento de caracteres manuscritos utilizando time de redes neurais. *IV Encontro Nacional de Inteligência Artificial*, p. 13–16, 2003.

SUEN, C. Y.; GUO, J.; LI, Z. C. Analysis and recognition of alphanumeric handprints by parts. *IEEE Transactions on Systems, Man, and Cybernetics*, v. 24, n. 4, p. 614–631, 1994.

TAY, Y. et al. Offline cursive handwriting recognition system based on hybrid markov model and neural networks. In: *IEEE International Symposium on Computational Intelligence in Robotics and Automation*. [S.l.: s.n.], 2003. v. 3, p. 1190–1195.

TAY, Y. et al. *An offline cursive handwritten word recognition system*. 2001. Disponível em: <citeseer.ist.psu.edu/article/tay01offline.html>.

TRIER, O.; JAIN, A.; TAXT, T. *Feature extraction methods for character recognition - A survey*. 1996. Disponível em: <citeseer.ist.psu.edu/trier95feature.html>.

VIARD-GAUDIN, C. et al. The ireste on/off (ironoff) dual handwriting database. In: *ICDAR '99: Proceedings of the Fifth International Conference on Document Analysis and Recognition*. Washington, DC, USA: IEEE Computer Society, 1999. p. 455. ISBN 0-7695-0318-7.

WANG, S.-S.; CHEN, P.-C.; LIN, W.-G. Invariant pattern recognition by moment fourier descriptor. *Pattern Recognition*, v. 27, n. 12, p. 1735–1742, 1994.

YAMAZAKI, A.; LUDERMIR, T. B.; SOUTO, M. C. P. de. Global optimization methods for designing and training neural networks. In: LUDERMIR, T. B.; SOUTO, M. C. P. de (Ed.). *SBRN*. [S.l.]: IEEE Computer Society, 2002. p. 136–141. ISBN 0-7695-1709-9.

ZHANG, G. P. Neural networks for classification: A survey. *IEEE Transactions on Systems, Man, and Cybernetics-Part C: Applications and Reviews*, v. 30, n. 4, p. 451–462, 2000.

ZHU, X.; SHI, Y.; WANG, S. A new distinguishing algorithm of connected character image based on fourier transform. In: *ICDAR*. [S.l.: s.n.], 1999. p. 788–791.