

PONTIFÍCIA UNIVERSIDADE CATÓLICA DO PARANÁ
Programa de Pós-Graduação em Informática Aplicada
Ciência da Imagem

Simone Bello Kaminski Aires

RECONHECIMENTO DE CARACTERES
MANUSCRITOS BASEADO EM REGIÕES
PERCEPTIVAS

Dissertação apresentada ao Programa de Pós-Graduação em Informática Aplicada da Pontifícia Universidade Católica do Paraná como requisito parcial para obtenção do título de Mestre em Informática Aplicada.

CURITIBA - PR
JULHO / 2005

SIMONE BELLO KAMINSKI AIRES

**RECONHECIMENTO DE CARACTERES
MANUSCRITOS BASEADO EM REGIÕES
PERCEPTIVAS**

Dissertação de Mestrado apresentada ao Programa de Pós-Graduação em Informática Aplicada da Pontifícia Universidade Católica do Paraná como requisito parcial para obtenção do título de Mestre em Informática Aplicada.

Área de Concentração: Computação Forense e Biometria

Orientador: Prof. Dr. Flávio Bortolozzi

Co-orientadora: Profa. Dra. Cinthia O. de A. Freitas

Co-orientador: Prof. Dr. Robert Sabourin

CURITIBA - PR

JULHO/ 2005

Aires, Simone Bello Kaminski

Reconhecimento de Caracteres Manuscritos Baseado em Regiões Perceptivas.

Curitiba:, 2005. 97 f.: il.

Dissertação (Mestrado) – Pontifícia Universidade Católica do Paraná. Programa de Pós-Graduação em Informática Aplicada, Curitiba, BR-PR, 2005. Orientador: Flávio Bortolozzi; Co-Orientadora: Cinthia O. de A. Freitas; Co-Orientor: Robert Sabourin.

1.Reconhecimento. 2. Caracteres manuscritos. 3. Percepção. 4. Redes Neurais. 5. Múltiplos Classificadores. 6. Discordância. 7. Meta-classes. I.Pontifícia Universidade Católica do Paraná. Centro de Ciências Exatas e de Tecnologia. Programa de Pós-Graduação em Informática Aplicada II-t.

Ao meu marido e minha
filha, João Paulo e Isabella,
e a toda minha família
com grande amor ...

“Nas grandes batalhas da vida, o primeiro passo para a vitória é o desejo de vencer !”

Mahatma Gandhi

Agradecimentos

Agradeço a Deus pela força e proteção durante toda esta tarefa.

A Professora Dr^a. Cinthia O. de A. Freitas pela orientação, esclarecimento, paciência, incentivo constante e valiosa amizade conquistada durante o desenvolvimento deste trabalho.

Gostaria de agradecer aos Professores Prof. Dr. Flávio Bortolozzi e Prof. Dr. Robert Sabourin pelo interesse no meu trabalho, dando importantes contribuições no desenvolvimento.

Ao professor Dr. Júlio César Nievola pelos esclarecimentos e contribuições em Redes Neurais e pela amizade desenvolvida.

Aos meus pais, Luiz e Marilda, pelo carinho, força e incentivo nos momentos difíceis.

Ao meu querido marido, João Paulo, pelo amor, dedicação e infinita compreensão.

A minha filha, Isabella, pelos carinhos e sorrisos que trouxeram motivação e força para continuar.

Minhas irmãs, Sandra e Sabrina, pelo incentivo e amor.

Aos meus amigos do CEFET-PR – Unidade Ponta Grossa, EM ESPECIAL ao Prof. Geraldo, a Prof. Simone Nasser, e ao Prof. Flávio Madalosso Vieira pelo incentivo e colaboração.

Ao colega José Josemar de Oliveira, pelas contribuições e discussões importantes.

A todos que direta ou indiretamente colaboraram na execução deste trabalho.

Sumário

CAPÍTULO 1	1
INTRODUÇÃO	1
1.1 DESCRIÇÃO DO PROBLEMA	3
1.2 OBJETIVOS	5
1.3 JUSTIFICATIVA	6
1.4 CONTRIBUIÇÕES	6
1.5 ORGANIZAÇÃO DA DISSERTAÇÃO	7
CAPÍTULO 2	8
REVISÃO BIBLIOGRÁFICA	8
CAPÍTULO 3	16
METODOLOGIA	16
3.1 PRÉ-PROCESSAMENTO	16
3.2 EXTRAÇÃO DE PRIMITIVAS	17
3.3 REPRESENTAÇÃO DAS PRIMITIVAS	21
3.3.1 <i>Formas de Representação</i>	21
3.3.2 <i>Percepção Humana das Formas</i>	22
3.3.3 <i>Mecanismos de Zoneamento</i>	32
3.4 CLASSIFICADOR NEURAL	36
3.4.1 <i>Redes Neurais</i>	36
3.4.2 <i>Arquitetura Classe-Modular</i>	37
3.4.3 <i>Múltiplos Classificadores</i>	39
3.5 DIVERSIDADE VERSUS DISCORDÂNCIA	47
3.5.1 <i>Medidas de Diversidade e Discordância</i>	47
3.5.2 <i>Matrizes de Confusão</i>	49
3.5.3 <i>Hipótese: Soft-Correlation</i>	51
3.5.4 <i>Definição de Meta-Classes</i>	51
CAPÍTULO 4	55
EXPERIMENTOS REALIZADOS	55
4.1 BASE DE DADOS – IRONOFF	55
4.2 EXPERIMENTOS	57
CAPÍTULO 5	74
CONCLUSÃO	74
REFERÊNCIAS BIBLIOGRÁFICAS	77

Lista de Figuras

Figura 1.1	Metodologia de reconhecimento de caracteres manuscritos	02
Figura 1.2	Estilos de escrita de diversos escritores	04
Figura 1.3	Similaridade entre caracteres distintos	05
Figura 2.1	Exemplo de palavras manuscritas	09
Figura 2.2	(a) Palavra <i>off-line</i> . A imagem da palavra é convertida em <i>pixels</i> de níveis de cinza usando um scanner. (b) Palavra <i>on-line</i> . As coordenadas x e y da caneta são gravadas como uma função de tempo com um digitalizador	10
Figura 3.1	Pré-processamento: limiarização	17
Figura 3.2	Determinação da “caixa” do caractere	17
Figura 3.3	Rotulação do fundo da imagem (<i>background</i>)	20
Figura 3.4	Divisão do caractere em 4 regiões	21
Figura 3.5	Representa parte de um vetor de características – 1º região da imagem.	22
Figura 3.6	Placa de sinalização	23
Figura 3.7	Princípio Gestalt da proximidade	24
Figura 3.8	Princípio Gestalt da similaridade	24
Figura 3.9	Princípio Gestalt da continuidade	24
Figura 3.10	Princípio do Fechamento: a) triângulo de Kanizsa e b) palavras incompletas	25
Figura 3.11	Princípio da Gestalt: fechamento	26
Figura 3.12	Princípio da Gestalt: figura-fundo	26
Figura 3.13	Princípio de área relata	27
Figura 3.14	Princípio da <i>pregnância</i>	28
Figura 3.15	Movimentos oculares durante a percepção visual.	30
Figura 3.16	Arquitetura da percepção humana no sistema de reconhecimento de caracteres	31
Figura 3.17	Exemplos de pontos de fixação e suas respectivas janela de atenção	32
Figura 3.18	Subdivisão do caractere	32

Figura 3.19	Partições, onde 1,2,3,4,5 e 6 são o número de posições das partes	33
Figura 3.20	Divisão do caractere K conforme padrão de zoneamento	33
Figura 3.21	Representa a formação das caixas	34
Figura 3.22	Divisão das caixas de código	35
Figura 3.23	Mecanismo de Zoneamento: Z = 4, 5H, 5V e 7 regiões	36
Figura 3.24	Modelo de neurônio	37
Figura 3.25	Arquitetura Classe Modular. a) Módulo Mi da rede e b) configuração completa com K módulos	38
Figura 3.26	Representação de classificadores, sendo que suas distâncias ótimas preservam suas discordâncias. Resultado para o problema Highleyman com 10+10 objetos.	49
Figura 3.27	Exemplos de meta-classes	52
Figura 3.28	Meta-classe “CDQRSZ”	53
Figura 3.29	Meta-classe “AB”	53
Figura 4.1	Visão geral da construção da base IRONOFF	56
Figura 4.2	Zoneamento baseado em regiões perceptivas	58
Figura 4.3	Gráfico que apresenta Taxas de Reconhecimento dos conjuntos de Validação e Teste para Classificadores Individuais.	61
Figura 4.4	Gráfico que representa a meta-classe AB projetada no espaço, conforme a distância entre os classificadores	66
Figura 4.5	Gráfico que representa a meta-classe CDQRSZ projetada no espaço, conforme a distância entre os classificadores	66
Figura 4.6	Gráfico que representa a meta-classe EJM projetada no espaço, conforme a distância entre os classificadores	67
Figura 4.7	Gráfico que representa a meta-classe FPT projetada no espaço, conforme a distância entre os classificadores	67
Figura 4.8	Gráfico que representa a meta-classe GX projetada no espaço, conforme a distância entre os classificadores	68
Figura 4.9	Gráfico que representa a meta-classe HKLNO projetada no espaço, conforme a distância entre os classificadores	68
Figura 4.10	Gráfico que representa a meta-classe IY projetada no espaço, conforme a distância entre os classificadores	69

Figura 4.11	Gráfico que representa a meta-classe UVW projetada no espaço, conforme a distância entre os classificadores	69
Figura 4.12	Meta-classes geradas pela mediana das distâncias do conjunto de validação	72
Figura 5.1	Exemplo de caractere da base IRONOFF	75
Figura 5.2	Imagens da base IRONOFF	76

Lista de Tabelas

Tabela 2.1	Comparativo de trabalhos que utilizaram a base IRONOFF	15
Tabela 3.1	Extração de Primitivas	19
Tabela 3.2	Rótulos do Vetor de Características	20
Tabela 3.3	Interseções e prioridades	35
Tabela 3.4	Distâncias com base nas matrizes de confusão para as classes: a)caractere “A”, b)caractere “B”	54
Tabela 4.1	Apresenta a distribuição de caracteres – Base IRONOFF	56
Tabela 4.2	Matriz de Confusão para conjunto de Teste, zoneamento $Z=4$	57
Tabela 4.3	Matriz de Confusão para conjunto de Teste, zoneamento $Z=5H$	58
Tabela 4.4	Matriz de Confusão para conjunto de Teste, zoneamento $Z=5V$	59
Tabela 4.5	Matriz de Confusão para conjunto de Teste, zoneamento $Z=7$	60
Tabela 4.6	Taxas de Reconhecimento para conjunto de Validação - Classificadores individuais e combinados	62
Tabela 4.7	Matriz Geral que apresenta as Taxas de Reconhecimento para conjunto de Teste Classificadores individuais e combinados	63
Tabela 4.8	Matriz Geral que apresenta as distâncias entre as matrizes de confusão para conjunto de Validação - Classificadores individuais e combinados	64
Tabela 4.9	Matriz Geral que apresenta as distâncias entre as matrizes de confusão para conjunto de Teste - Classificadores individuais e combinados	65
Tabela 4.10	<i>DD-based</i> e taxas de reconhecimento para os múltiplos classificadores - Conjunto de Validação	70
Tabela 4.11	<i>DD-based</i> e taxas de reconhecimento para os múltiplos classificadores - Conjunto de Teste	71
Tabela 4.12	Matriz que apresenta os acertos, erros e rejeições das meta-classes definidas, validadas no conjunto de Teste.	73

Lista de Símbolos

P	Espaço de padrões
M_j	Conjuntos mutuamente exclusivos
C_l	Conjunto pertencente ao espaço de padrões
X	Amostra do espaço de padrões
Z	Objeto que se deseja classificar
L_i	Conjunto de Classificadores
w_j	Classes do problema
$P(.)$	Probabilidade
α_i	Conjunto de pesos dos classificadores
ζ	Conjunto de Treinamento
$d_j(C_1, C_2)$	Discordância entre classificadores C_1 e C_2

Lista de Abreviaturas

MLP	<i>Multiple Layer Perceptron</i>
CPS	Espaço de Projeção de Classificadores
DD-based	Discordância baseada no critério da Distância
HMM	<i>Hidden Markov Models</i>
K-NN	<i>K-Nearest Neighbor</i>
MCS	<i>Multiple Classifiers System</i>
NN	<i>Neural Network</i>
NSLO	Norte, Sul, Leste, Oeste
PD	Programação Dinâmica
PE	Elemento de processamento
RNA	Redes Neurais Artificiais

Resumo

Este trabalho investiga mecanismos de zoneamento perceptivo para reconhecimento de caracteres manuscritos. Propõe-se um mecanismo de zoneamento não simétrico baseado na análise das matrizes de confusão dos classificadores individuais (Classe-Modular). Zoneamento é um método de análise de informações locais em um dado padrão particionado. A extração de características é baseada em Concavidades e Convexidades extraídas através da rotulação dos *pixels* do fundo (*background*) da imagem de entrada. Este procedimento tem por base o enquadramento do caractere em uma caixa (*bounding box*) e na divisão do mesmo em Z partes, sendo $Z = 4, 5H$ Horizontal, $5V$ Vertical e 7 . A base de dados utilizada para os experimentos é a IRONOFF, com caracteres manuscritos do alfabeto. No processo de reconhecimentos utiliza-se um comitê de rede neurais artificiais MLP (*Multiple Layer Perceptron*) Classe-Modular, ou seja, um comitê de redes neurais MLP de forma que cada classe do problema possui uma rede específica. Uma metodologia para múltiplos classificadores é aplicada ao problema de reconhecimento, podendo ser utilizada para a fusão (combinação) de classificadores. 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: $4 = 82,89\%$, $5H = 81,75\%$, $5V = 80,94\%$ e $7 = 84,73\%$. As combinações realizadas entre os classificadores individuais apresentam uma melhoria na taxa de reconhecimento, sendo de $85,9\%$ para a rede $5H-5V-7$. O resultado global considerando uma arquitetura composta por 2 níveis de classificação (meta-classe e classe) atinge uma taxa média de reconhecimento de $84,15\%$, com rejeição de $11,95\%$ e erro de $3,90\%$.

Palavras-chave: Caracteres Manuscritos, Reconhecimento, Percepção, Múltiplos Classificadores, Discordância.

Abstract

This work investigates the perceptual zoning mechanism for handwritten character recognition. It is proposed a non-symmetrical zoning mechanism as the baseline on the analysis of the confusion matrix for each individual classifier (Class-Modular). Zoning is a method for local information analysis on partitions of a given pattern. The feature extraction is based on Concavities/Convexities deficiencies, which are obtained by labeling the background pixels of the input images. Therefore, circumscribes the letter by a rectangle and partition it into Z parts, such as: $Z = 4$, 5H(horizontal), 5V (Vertical) and, 7 parts. The base of data used for the experiments is IRONOFF, with handwritten characters of the alphabet. For the recognition problem a Neural Network team is proposed, where the K -classification problem is decomposed into K 2-classification sub problems, each for one of the K classes. A methodology for multiple classifiers system (MCS) is applied to the recognition problem, could be used for the fusion (combination) of classifiers. The methodology defines an alternative approach instead of using the recognition rate criterion, which can be used to evaluate *a priori* classifiers combination in MCS. The obtained recognition rate for the evaluated zonings are the following: 4 = 82,89%, 5H = 81,75%, 5V = 80,94% and 7 = 84,73%. The combinations accomplished among the individual classifiers present an improvement in the rate recognition, being the best result of 85.9% for the network 5H-5V-7. The global result considering a composed architecture for 2 classification levels (meta-class and class) reaches an average recognition rate of 84,15%, with rejection of 11,95% and error of 3,90%.

Keywords: Handwritten Character Recognition, Perceptual Concepts, Neural Network, Multiple Classifiers System, Disagreement and Distance Measures.

Capítulo 1

Introdução

De acordo com Plamondon e Srihari [PLA00], a escrita manuscrita consiste de marcas gráficas em uma superfície, com o propósito na maioria das vezes de comunicação, sendo 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, existindo regras para combinação de letras para formar unidades representativas lingüística de alto nível. Por exemplo, há regras para combinação de formas e letras individuais para formar palavras cursivas no alfabeto latino.

Documentos em papel parecem relíquias, principalmente quando se fala em manuscritos. Para [OLI04], este pré-julgamento é falho, uma vez que o papel utilizado como meio de comunicação tem suas vantagens em relação a outros meios:

- O papel é um meio padronizado, que não possui problema de interface com o escritor e o leitor;
- Papel é portátil e seu transporte é bem estabelecido, mesmo sendo mais lento que uma transferência eletrônica;
- A escrita de um recado, de um endereço ou o preenchimento de um formulário à mão não necessita de condições especiais, a menos da habilidade do escritor, da necessidade do papel e de um instrumento se escrita.

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 [PLA00]. O estudo das palavras manuscritas está ligado ao desenvolvimento de métodos de reconhecimento voltados para aplicações do mundo real envolvendo palavras e caracteres manuscritos, tais como: processamento automático de cheques bancários,

envelopes postais, formulários, textos manuscritos, entre outros. Esses são sistemas de leitura automática cuja tarefa é servir de ponte entre o *mundo* do papel e da escrita convencional e o *mundo* dos computadores e do processamento eletrônico [OLI04].

O desenvolvimento de sistemas de reconhecimento automático de palavras manuscritas tem desafiado os pesquisadores devido ao alto grau de dificuldade em reproduzir a capacidade humana de ler.

De acordo com Freitas [FRE01], documentos manuscritos apresentam componentes de grande complexidade, tais como: diferentes estilos de escrita, diferentes tipos de números manuscritos, contexto da escrita, contexto da aquisição do documento (*on-line* - dinâmico ou *off-line* - estático) e ainda, o tamanho do léxico a ser reconhecido e o número de escritores.

O tema central desta dissertação consiste no reconhecimento de caracteres manuscritos *off-line* utilizando regiões perceptivas, sendo a metodologia de reconhecimento apresentada na Figura 1.1, composta das seguintes tarefas, :

- Pré-processamento
 - Limiarização (*thresholding*): esta tarefa consistem em converter a imagem original em níveis de cinza em uma imagem binária.
 - Caixa (*Bounding Box*): este pré-processamento busca extrair da imagem os limites acima, abaixo, esquerda e direita criando ao redor do caractere uma “caixa”.
- Extração de primitivas: visa a obtenção de um conjunto de características dos caracteres manuscritos;
- Classificação: efetua o reconhecimentos dos padrões de caracteres manuscritos.

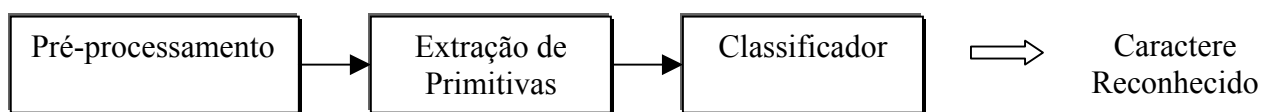


Figura 1.1: Metodologia de reconhecimento de caracteres manuscritos

Assim, o presente trabalho visa contribuir para o reconhecimento de caracteres, buscando incorporar aspectos da percepção humana no processo de reconhecimento, utilizando uma metodologia de múltiplos classificadores para auxiliar no processo, e considerando um léxico de 26 classes, correspondentes aos caracteres do alfabeto.

1.1 Descrição do Problema

A escrita manual é uma das formas mais naturais de comunicação entre as pessoas, constatando-se a geração de uma quantidade de dados em papel muito volumosa. Muitas vezes é necessário processar os dados contidos nestes papéis por máquinas, sendo extremamente desejável, que computadores tenham capacidade de ‘ler’ e interpretar documentos em papel.

O reconhecimento de caracteres manuscritos tem sido uma preocupação da comunidade científica. As aplicações para um sistema que faça tal tipo de reconhecimento são muitas, podendo citar leitoras automáticas de cheques bancários, máquinas automáticas de processamento de códigos postais, máquinas automáticas para processar qualquer tipo de formulário preenchido manualmente, entre outros.

O fato dos caracteres estarem na forma manuscrita torna a tarefa de reconhecimento complexa. Isto ocorre devido as variações de estilos existentes na escrita manuscrita, ou seja, o estilo pessoal de cada escritor, onde para cada escritor a forma de um caractere pode ser feita de diferentes maneiras e até o mesmo escritor pode representar o caractere de maneiras diferentes. Além disto, a qualidade do documento e do dispositivo de digitalização devem ser considerados. A Figura 1.2 apresenta a diversidade de estilos de escrita do mesmo caractere entre amostras de vários escritores.

Outro problema no reconhecimento de caracteres manuscritos, é a similaridades entre caracteres distintos, por exemplo, U e V, Q e O, entre outros. A Figura 1.3 apresenta algumas similaridades entre caracteres na base estudada.

O presente estudo situa-se no contexto de caracteres manuscritos isolados. Um fator determinante para um bom desempenho do reconhecimento é a seleção do conjunto de características a serem extraídas dos caracteres. Os caracteres possuem diferentes particularidades, por exemplo: laços, traços verticais e horizontais, entre outros. Entretanto, o método de extração deve identificar as características particulares de cada um, salientando suas diferenças em relação aos outros caracteres.

A	A	A	ct	#	Œ
B	B	B	B	CB	Œ
C	C	C	C	C	C
D	D	D	D	D	D
E	E	E	E	E	E
F	F	F	F	F	F
G	G	G	G	G	G

Figura 1.2: Estilos de escritura de diversos escritores

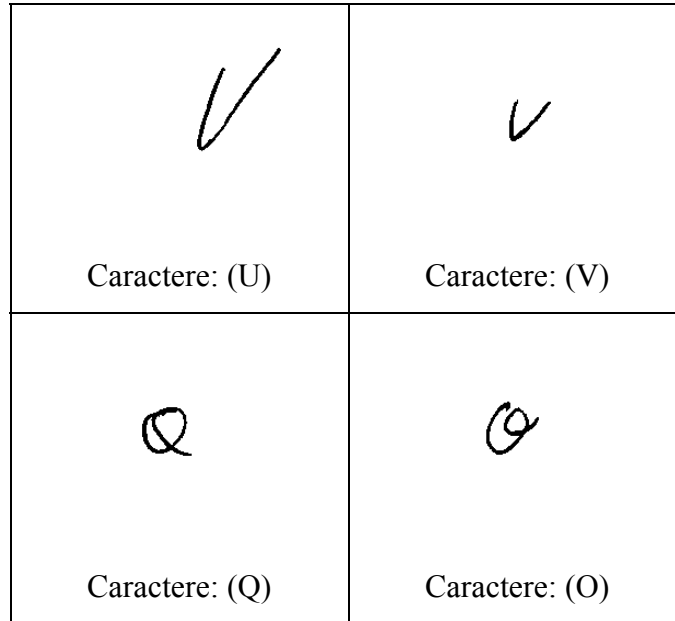


Figura 1.3: Similaridade entre caracteres distintos

1.2 Objetivos

A tarefa de leitura de manuscritos envolve habilidades humanas e o conhecimento do domínio é essencial. A escrita manuscrita pode ter duas formas: manuscritos isolados e manuscritos cursivos. No primeiro caso os caracteres estão dispostos na imagem de forma não conectada, e no segundo caso de forma completamente irrestrita, ou seja, conectados e eventualmente desconectados.

Este trabalho está inserido no contexto de caracteres manuscritos isolados, composto por letras do alfabeto latino (maiúsculas), no qual pretende-se explorar uma abordagem Analítica (*Local Approach*), visto que trabalhos focados na abordagem Global (*Global Approach*) já foram desenvolvidos por [FRE01] [KAP04] [OLI02]. Além disso pretende-se complementar a abordagem Global através da abordagem Analítica tendo-se em vista um léxico de pequena dimensão. Para isso, o estudo investiga regiões perceptivas para o processo de reconhecimento com base na percepção humana de formas [FRE02] [YAN98] [COR02], com objetivo de dar continuidade em aspectos já estudados por [FRE01], contribuindo com um conjunto de primitivas, múltiplos classificadores e definindo meta-classes, buscando auxiliar nas soluções para problemas de reconhecimento de palavras manuscritas já estudados por [KAP04] [OLI02] [OLI04]. Um mecanismo de zoneamento perceptivo para reconhecimento de caracteres manuscritos é proposto, utilizando-se um zoneamento não

simétrico baseado na análise das matrizes de confusão dos classificadores individuais (Classe-Modular).

Para a classificação dos caracteres apresenta-se um estudo da aplicação das redes neurais artificiais com arquitetura MLP (*Multiple Layer Perceptron*) Classe-Modular, ou seja, um comitê de redes neurais MLP de forma que cada classe do problema possui uma rede específica, assim como descrito em [OH_02] e [KAP03]. Propõe-se ainda uma metodologia para múltiplos classificadores aplicada para o problema de reconhecimento, que pode ser utilizada para a fusão (combinação) de classificadores.

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 ou utilizando-se de mecanismos mais complexos de definição de conjuntos (*ensembles*) de classificadores, tais como: computação evolutiva, algoritmos genéticos [OLS05].

1.3 Justificativa

O presente estudo se justifica diante da complexidade da tarefa de reconhecimento, tendo em vista a descrição do problema (Figura 1.1), bem como as confusões entre as letras (Figura 1.2), sejam estas durante o desenvolvimento de sistemas computacionais ou através do próprio ser humano. Este trabalho insere-se no estado da arte uma vez que se relaciona com outros já realizados por [FRE01] [KAP04] [FRE02] [OLI02] [OLI04].

1.4 Contribuições

Este trabalho contribue com o estudo de regiões perceptivas para o processo de reconhecimento das formas. Trata-se de uma abordagem analítica do problema, tendo em vista um léxico de pequena dimensão: caracteres manuscritos maiúsculos, ou seja, 26 letras do alfabeto. Enfatiza-se a necessidade de incorporar a percepção humana em novos sistemas para contribuir no processo de reconhecimento, tal qual sugerido por Suen [SUE94].

Uma metodologia para múltiplos classificadores é proposta buscando definir uma abordagem alternativa ao processo de combinação de classificadores, ao invés da utilização do critério da taxa de reconhecimento. Esta abordagem define uma Discordância baseada na medida da distância (*Disagreement based on Distance - DD-based*), aplicando uma distância

euclidiana computada entre as matrizes de confusão dos classificadores e uma regra de *soft-correlation* para indicar o melhor conjunto de classificadores.

Neste trabalho, o interesse é prover uma enfoque alternativo que permita avaliar a priori os conjuntos de classificadores para determinar a melhor combinação entre eles, sem combina-los diretamente. A razão principal para isto é a dificuldade de executar uma busca exaustiva dentro do espaço de combinação dos classificadores quando se tem um grande número de classificadores a verificar. A abordagem usa informação contida nas matrizes de confusão para cada classificador individual, e computa distâncias que representam a discordância entre os classificadores a serem combinados.

1.5 Organização da Dissertação

Esta dissertação está organizada em 5 capítulos. No capítulo 2 apresenta-se uma revisão sobre o estado da arte. O capítulo 3 descreve o método proposto para o reconhecimento de caracteres manuscritos isolados. Os experimentos realizados para a validação da metodologia proposta são apresentados no Capítulo 4, bem como a análise dos resultados. No Capítulo 5 são apresentadas as conclusões e os trabalhos futuros.

Capítulo 2

Revisão Bibliográfica

De acordo com [PLA00] a escrita manuscrita é pessoal a cada indivíduo, consiste em marcas gráficas em uma superfície com a finalidade de comunicação na maioria das vezes. A escrita é valorizada pela contribuição no desenvolvimento das culturas e civilizações.

Cada texto manuscrito é um conjunto de ícones, que são caracteres ou letras com formas básicas definidas. Existem regras para combinar as letras, de maneira a representar a forma de uma unidade lingüística de um nível mais elevado. Por exemplo, as regras para combinar as formas de letras individuais para dar forma a palavras cursivas escritas no alfabeto latino.

Para [PLA00], a razão da escrita ter persistido ao longo dos anos na era do computador digital é a conveniência do papel e da caneta comparados aos teclados em numerosas situações do dia-a-dia. A tarefa de leitura de manuscritos envolve habilidades humanas e o conhecimento do domínio é essencial, por exemplo, em prescrições médicas o farmacêutico utiliza seu conhecimento *a priori* sobre os medicamentos.

Os métodos de reconhecimento de manuscritos são utilizados em aplicações do mundo real envolvendo palavras manuscritas, como: textos, cheques bancários, envelopes postais, formulários, entre outros. Alguns exemplos destas aplicações são apresentados na Figura 2.1.

Vários tipos de análise, interpretação e reconhecimento podem estar associadas com o processamento de documentos manuscritos. O reconhecimento é a transformação de uma linguagem de marcas gráficas para sua representação simbólica. A interpretação determina o significado de uma palavra manuscrita, por exemplo, um endereço postal. A identificação é o processo de determinar o autor de um manuscrito em um conjunto de escritores, assumindo que cada escritor possui uma forma de manuscrito individual [PLA00].

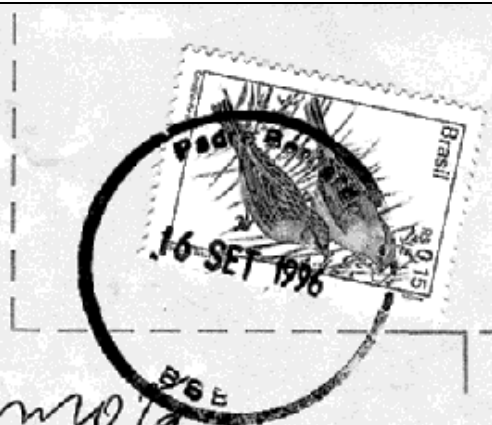
setecentos e oito mil seiscentos e oitenta e nove
reais e cinquenta centavos

NOVECENTOS e sessenta e seis mil cento e
DEZESSETE REAIS e sessenta centavos

quatorze reais e cinquenta centavos.

a) cheques bancários

esta vai para
nomato Honorio
em muquezinho
do Cuidado de
Evilazio Lerre
das Santo em Paramoia
Ceará



b) envelope postal

Figura 2.1: Exemplo de palavras manuscritas

2.1 Aquisição dos dados

Os dados manuscritos são digitalizados pela varredura da escrita no papel ou por uma caneta especial ou superfície eletrônica, tal como um digitalizador associado a uma tela de cristal líquido. Estas abordagens são diferenciadas como manuscrito *off-line* e *on-line*, respectivamente. No caso *on-line*, as coordenadas bidimensionais de pontos sucessivos são descritas em uma função de tempo e são armazenadas seqüencialmente, sendo assim, a ordem dos segmentos de palavras realizados pelo escritor está prontamente disponível. Na

abordagem *off-line* somente o manuscrito completo está disponível na imagem. Estes manuscritos são concebidos por um indivíduo, através de um lápis ou caneta sobre um papel, os quais são posteriormente “escaneados” e gerados na forma digital. A abordagem *on-line* trata-se de uma representação espaço-temporal da entrada, ao passo que a abordagem *off-line* envolve análise de espaço-luminosidade da imagem [PLA00].

Entradas típicas de sinais *on-line* e *off-line* podem ser observadas na Figura 2.2. As taxas de armazenamento dos dados requeridas pelos sistemas são extremamente diferentes. Os dados requeridos para uma palavra escrita de maneira cursiva são: no caso *on-line* (Figura 2.2b), poucas centenas de *bytes*, capturados em 100 amostras por segundo, e no caso *off-line* (Figura 2.2a), poucas centenas de *kilobytes*, amostrados em x pontos por polegada (*dpi – dots per inches*).

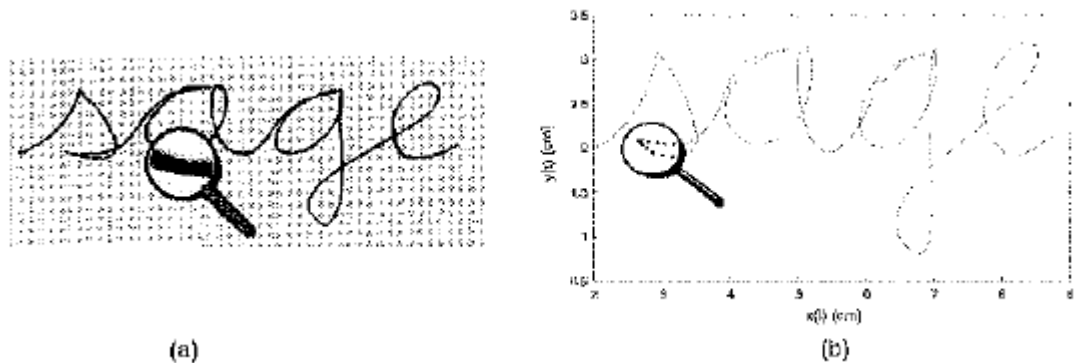


Figura 2.2: (a) Palavra *off-line*. A imagem da palavra é convertida em *pixels* de níveis de cinza usando um scanner. (b) Palavra *on-line*. As coordenadas x e y da caneta são gravadas como uma função de tempo com um digitalizador

Fonte: Adaptado de [PLA00]

No reconhecimento manuscrito *off-line* as principais tarefas são o reconhecimento de palavras e caracteres. [KAP04]. De acordo com [PLA00], a leitura de documentos manuscritos conta com conhecimentos básicos sobre percepção. Experimentos psicológicos humanos no reconhecimento de caracteres apresentam dois resultados: 1) um caractere que ocorre freqüentemente, ou tem uma estrutura simples, é processado em unidades simples sem qualquer decomposição da estrutura original do caractere, 2) os caracteres que não ocorrem freqüentemente ou possuem estruturas complexas necessitam de uma decomposição (segmentação) e, ainda, apresentam um aumento do tempo de processamento de acordo com a quantidade de segmentos. Assim, de um modo geral, o reconhecimento de palavras manuscritas pode ser tratado por duas abordagens, a saber:

- a) *Local (analytical approach)*: esta abordagem é realizada em nível dos caracteres. Este método necessita da segmentação das palavras em caracteres ou pseudo-caracteres. Esta abordagem, no entanto apresenta dificuldades em se estabelecer as fronteiras entre os caracteres. Desta forma, o método de reconhecimento dependerá do sucesso do processo de segmentação utilizado. Esta abordagem atende as necessidades de léxicos desconhecidos *a priori* e de grandes dimensões, por exemplo, envelopes postais [KAP04].
- b) *Global (global approach)*: esta abordagem é realizada em nível das palavras. Este método evita a etapa de segmentação das palavras, extrai-se primitivas globais sem a necessidade de segmentação explícita de caracteres/letras. Esta abordagem explora informações do contexto das palavras. É uma abordagem restrita a aplicações de léxicos com pequenas dimensões, por exemplo, cheque bancários [KAP04].

Sistemas de reconhecimento com desempenho de alto nível consideram a combinação de ambas as abordagens [PLA00]. Os métodos podem diferir na utilização das restrições específicas do domínio da aplicação, entretanto sua estrutura básica é a mesma. A metodologia de um modo geral envolve as seguintes fases: pré-processamento, segmentação (que pode ser omitida se utilizado uma abordagem Global), reconhecimento e pós-processamento. Os métodos utilizados para a extração de primitivas definem uma das principais etapas na obtenção de um sistema robusto para reconhecimento de palavras [KAP04].

2.2 Reconhecimento de padrões

No trabalho apresentado por [JAI00] é realizado uma sumarização dos métodos utilizados em várias fases de um sistema de reconhecimento de padrões.

Para a etapa de reconhecimento de manuscritos, ressaltam-se as seguintes metodologias:

a) Métodos Estatísticos: o padrão é representado por um conjunto de d características, ou atributos, visto como um vetor de características d -dimensional. Conceitos da teoria de decisão estatística são utilizados para estabelecer fronteiras de decisão entre as classes, sendo o sistema operado em dois modos: treinamento (aprendizado) e classificação

(teste) [JAI00]. Esses métodos exploram as propriedades métricas das formas a serem reconhecidas efetuando-se uma modelagem em termos de componentes ligados a natureza de cada forma. Dividem-se em: métodos paramétricos e não-paramétricos.

a.1) Métodos Paramétricos (*bayesianos*): estes métodos operam na hipótese de que as classes seguem uma distribuição de probabilidades de uma determinada forma. Supõe-se que são conhecidas as leis de probabilidade das classes, ou seja, os métodos não podem ser aplicados de maneira realista a não ser sobre hipóteses gaussianas com estimativa de parâmetros correspondentes. A decisão busca determinar a classe para a qual a forma desconhecida apresenta a probabilidade de pertencer, considerando-se um máximo de pertinência [KAP04]. Estes métodos utilizam o Teorema de Bayes que é tratado como um instrumento de atualização de conhecimento quando novos fatos são apresentados aos sistema, pois relaciona probabilidade *a priori*, com probabilidade *a posteriori*. O Teorema de Bayes é apresentado na Equação 2.1.

$$p(w_i | x) = \frac{p(x | w_i)}{\sum_j p(x | w_j)} \quad (2.1)$$

Duas técnicas utilizadas em reconhecimento de formas com base em métodos estatísticos paramétricos são:

- **Distância de Mahalanobis:** esta técnica mede a distância entre o ponto que representa a forma desconhecida e os pontos que representam as classes conhecidas.
- **Métodos Neurais:** Estes métodos foram estudados com a intenção de realizar funções próximas às do cérebro humano. Estes modelos são compostos de vários elementos (ou neurônios) de cálculo não lineares operando em paralelo e organizados de maneira a imitar redes de neurônios biológicos. Este método será abordado com maiores detalhes na Seção 3.4.

a.2) Métodos Não-Paramétricos: para estes métodos, as probabilidades são desconhecidas em cada classe. O problema está na decisão das regiões de fronteira entre as classes. Para este método três são as técnicas utilizadas:

- **Método dos k -Vizinhos-Mais-Próximos:** A forma desconhecida é afetada pela classe dos seus k -vizinhos-mais-próximos do conjunto de treinamento. Este método utiliza a classificação através de funções de cálculo da Distância Euclidiana clássica entre a forma analisada e as classes pré-definidas, podendo ser aplicada em todo espaço munido de uma distância qualquer.
- **Janelas de *Parzen*:** Este método visa dar uma estimativa de probabilidade da forma desconhecida para uma classe conhecida. Para cada ponto do espaço, conta-se o número de pontos das classes conhecidas próximas da forma desconhecida e atribui-se a estes pontos um valor ponderado, considerando um valor mais forte quanto a proximidade do ponto da forma desconhecida.
- **Método de Correspondência por Máscaras:** Esta técnica chamada de *Template Matching* define cada classe por um conjunto padrão de protótipos armazenados em memória. A comparação se faz diretamente, sendo que a classe reconhecida é aquela que possui o protótipo o mais correlacionado a forma desconhecida. Esta comparação pode ser realizada por uma comparação simples *pixel a pixel*, ou por uma análise de árvore de decisão nas quais apenas alguns *pixels* selecionados são testados.

b) Métodos Estruturais: estes métodos descrevem características intrínsecas de uma forma, realizando a modelagem em termos de componentes estruturais, ligados a natureza de cada forma. A técnica mais conhecida entre os métodos estruturais é a Programação Dinâmica (PD), que é um paradigma comumente utilizado para unir caracteres candidatos em potencial para palavras candidatas. Um exemplo de método que utiliza PD é o de [PAR02] que combina heurísticas com PD para desclassificar certos grupos de segmentos de primitivas só sendo avaliados se eles possuem uma complexidade de um caractere.

2.3 Reconhecimento de caracteres

Existem muitos trabalhos que exploram tecnologias com o propósito de fazer o reconhecimento de caracteres manuscritos. Atualmente a área de pesquisa é muito intensa

considerando-se o aumento do poder computacional em computadores pessoais, e também as aplicações de leitores automáticos de endereços postais, cheques bancários, e várias outras formas de tecnologia que instigam um rápido desenvolvimento na área de reconhecimento de palavras manuscritas nos últimos anos.

De acordo com Suen et al. [SUE94] com muita prática, todo humano alfabetizado tem habilidades para o reconhecimento de caractere desde a infância. Investigando o processo de reconhecimento humano, pode conduzir a diferentes maneiras de identificar os caracteres. Olhando de relance os padrões inteiros normalizados pelo tamanho, os humanos podem identificar corretamente, mas dada só uma parte do padrão, sua identificação não é tão óbvia. Uma maneira de reconhecer é verificar o pedaço do padrão particionado e passar pelos caracteres armazenados no cérebro da pessoa, escolhendo um possível candidato que contém a mesma parte, e então tentar acrescentar outras partes a isto para formar o possível caractere. Em alguns trabalhos [SUE91] [SUE92] um modelo hierárquico é criado para avaliar a taxa de reconhecimento das distintas partes dos caracteres. São propostas partições nos caracteres em Z partes, sendo, $Z = 6, 4, 2LR, 2UD$. Em [SUE94] é proposto um modelo avançado aos trabalhos [SUE91] [SUE92], buscando maior eficiência no reconhecimento através das probabilidades de ocorrência do padrão, definições como partes cruciais, relações da eficiência, grau de confusão, pares similares do caractere, foram verificadas buscando facilitar a análise do teste padrão e o reconhecimento do caractere. Este trabalho será abordado no item 3.3.3.

No trabalho apresentado por [LAL00], o autor propõem uma metodologia que visa alcançar em uma base *off-line* de **palavras**, o desempenho de reconhecimento conseguidos em uma base *on-line*. Sua abordagem denomina-se “*OrdRec*” e o objetivo da mesma é encontrar o provável caminho que corresponde idealmente à verdadeira trajetória da caneta. Neste sistema utiliza-se *HMM (Hidden Markov Models)* para o processo de reconhecimento. Apresenta-se que sistemas treinados com as imagens *on-line* e *off-line* disponíveis na base IRONOFF[VIA01], na abordagem (Ord. ***On&Off***) alcança uma taxa de reconhecimento de 93%, comparada a 90.2% sem a utilização do “*OrdRec*”. O autor ressalta que sistemas de reconhecimento *on-line* alcançam 94.5%.

Tay [TAY01] apresenta um estudo com **palavras** manuscritas *off-line*, propondo um sistema híbrido combinando HMM e Redes Neurais (*Neural Network - NN*), NN-HMM. A

base utilizada para os experimentos foi a IRONOFF. Neste estudo uma taxa de reconhecimento de 91.7% foi alcançada.

Em [TAY03] o autor apresenta um estudo com *palavras* manuscritas, bem como, avalia também o reconhecimento de caracteres no processo de segmentação. Para o processo de reconhecimento utiliza um sistema híbrido NN-HMM. Nos resultados o processo é testado com uma base pequena que ele chama de IRONOFF-196, depois o sistema é avaliado em uma base maior IRONOFF-2000. As taxas de reconhecimento na base IRONOFF-2000 são de 83.1% para **caracteres** e 88.1% para **palavras**.

[POI02] apresenta uma arquitetura que investiga o reconhecimento de caracteres manuscritos isolados. Em seu trabalho são realizados estudos utilizando caracteres da base IRONOFF, on-line e off-line. As arquiteturas são testadas separadamente e depois uma arquitetura híbrida é proposta. Para caracteres maiúsculos (*uppercase*) da base IRONOFF *on-line* a taxa de reconhecimento é de 94,2% para a arquitetura proposta e, para uma MLP convencional é de 93,6%. Para os caracteres *off-line* as taxas são de 89,9% e 87,1% respectivamente para a arquitetura proposta em uma MLP convencional.

Um comparativos das taxas de reconhecimento dos trabalhos apresentados na literatura que utilizam a base de dados IRONOFF é apresentado na Tabela 2.1.

Tabela 2.1: Comparativo de trabalhos que utilizaram a base IRONOFF

Autor	Palavras			Caracteres		
	<i>on-line</i>	<i>off-line</i>	<i>on-line+off-line</i>	<i>on-line</i>	<i>off-line</i>	<i>on-line+off-line</i>
Lallican (2000)			93%			
Tay (2001)		91,70%				
Tay (2003)		88,10%			83,10%	
Poisson					87,10%	

Os trabalhos citados neste capítulo contribuem na elaboração do presente trabalho, principalmente ajudam a entender a complexidade do problema de reconhecimento de caracteres e palavras manuscritas. No capítulo seguinte, são descritos: o pré-processamento, a extração de primitivas, a representação das primitivas, os mecanismos de zoneamento e os classificadores utilizados na metodologia do presente trabalho para o reconhecimento de caracteres manuscritos baseado em regiões perceptivas.

Capítulo 3

Metodologia

Este capítulo apresenta a seqüência de etapas necessárias ao desenvolvimento de um método de reconhecimento, como mostrado na Figura 1.1, sendo estas as seguintes:

- Pré-processamento: limiarização, e construção de uma “caixa” (*bouding Box*);
- Extração de primitivas: Geométricas;
- Reconhecimento dos caracteres utilizando um classificador neural classe-modular;

3.1 Pré-Processamento

Na etapa de pré-processamento realiza-se diversas operações para tratamento da imagem digitalizada, tais como realce de contraste, eliminação de ruídos, segmentação, limiarização, de maneira que as características da imagem possam ser realçadas. No presente estudo, a base de dados utilizada já se encontra segmentada em caracteres isolados, sem ruído, sendo a etapa de pré-processamento composta pelas técnicas descritas a seguir:

- a) **Limiarização** (*thresholding*): esta tarefa consiste em converter a imagem original em níveis de cinza em uma imagem binária. Na imagem original, as variações de nível de cinza ocorrem em um intervalo que vai do preto (valor mínimo) ao branco (valor máximo). O extremo maior corresponde ao fundo branco do papel, enquanto que o extremo menor representa a imagem que está em primeiro plano. A tarefa de limiarização consiste em determinar o valor para um limiar de escala de cinza e todos os valores abaixo deste limiar escolhido associa-los ao preto. No presente trabalho

utiliza-se o método de OTSU [OTS79]. A Figura 3.1 apresenta a imagem original da base de dados e a imagem resultante do método de limiarização.



(a) Imagem original (b) Imagem limiarizada

Figura 3.1: Pré-Processamento: Limiarização

- b) **Caixa** (*Bounding Box*): este pré-processamento busca extrair da imagem os limites acima, abaixo, esquerda e direita criando ao redor do caractere uma “caixa”. Este procedimento visa auxiliar o mecanismo de zoneamento (*zoning*) do caractere, de acordo com as regiões perceptivas a serem estudadas. A Figura 3.2 exemplifica a criação da “caixa” em imagens da base de dados.



Figura 3.2: Determinação da “caixa” do caractere

3.2 Extração de Primitivas

As primitivas mais empregadas na representação global das propriedades dos caracteres são as primitivas perceptivas. Inclui-se nesta classe os ascendentes, descendentes, laços, traços, barras verticais, horizontais e inclinadas, segmentos de linhas, propriedades regulares, entre outros [FRE01]. A Tabela 3.1 apresenta um resumo de trabalhos relacionados a extração de primitivas no contexto de palavras e caracteres manuscritos.

A extração de primitivas perceptivas não se aplica ao estudo de caracteres manuscritos isolados, uma vez, que em caracteres não cabe a determinação das zonas de ascendentes, corpo e descendentes. Aplica-se as primitivas perceptivas em métodos de reconhecimento de palavras manuscritas pelo fato que estas apresentam tolerância às distorções e às variações de estilo e podem descrever as propriedades globais das palavras [FRE01].

Uma análise das partes côncavas e convexas das palavras permite evidenciar as propriedades topológicas e geométricas das formas a reconhecer. Para [FRE01] existem dois tipo de métodos que permitem extrair as concavidades, a saber:

- métodos com base na rotulação das áreas brancas das imagens ou do fundo das imagens (*background*),
- métodos com base no tratamento do contorno das formas.

O método composto pela rotulação dos *pixels* do fundo da imagem dos caracteres está sendo utilizado no presente trabalho de maneira a extrair para cada “caixa” da imagem um vetor de características (*feature vector*).

A composição do vetor de característica é baseada na rotulação de cada *pixel* do fundo da imagem, mapeando 24 situações diferentes. Considera-se dois tipos de verificação, a saber:

- a) verificação dos 4 vizinhos, norte-sul-leste-oeste (NSLO) para cada *pixel* branco no caractere,
- b) Verificação de falsos laços com a possibilidade de escape em 4 direções de sondagem para os *pixels* determinados como fechados (laço) para as 4 direções principais (NSLO).

A Figura 3.3 exemplifica a rotulação dos *pixels* do fundo da imagem do caractere, retirada da base de dados de treinamento. A Tabela 3.2 apresenta a rotulação do vetor de características do caractere.

Tabela 3.1: Extração de Primitivas

Autores	Características
[SUE94]	Baseado na probabilidade de ocorrência de padrões. Recomenda um reconhecimento por partes, subdividindo regiões em 2,4 e 6 partes.
[YAN98]	Recomenda imitar a psicologia humana no processo de reconhecimento.
[HUN00]	Propõem que cada caractere seja assinalado em uma caixa de código de n ou menos dígitos. Cada dígito da caixa de código representa um único tipo de movimento de um dado caractere. Este esquema independe do estilo da escrita e da seqüência de movimentos.
[SIL03]	Os vetores de características são compostos por um conjunto de valores de distâncias calculados através de polígonos de referência até o contorno da imagem. Para a correção de falhas internas dos caracteres, devido ao fato de que o método do polígono faz análise do contorno, sugere uma complementação a criação de um vetor de características internas do caractere. Estas características são extraídas partindo do centro da imagem e terminando no contorno interno do caractere. Sugere ainda extração utilizando quadrado rotacionado.
[OLI02]	Primitivas perceptivas (posição ascendente, tamanho do ascendente, posição e tamanho do descendente, tamanho <i>loop</i> fechado, localização do <i>loop</i> fechado, concavidade, estimativa do tamanho da palavra), direcionais (medidas de concavidade) e topológicas. Um mecanismo de <i>zoning</i> divide cada palavra em 8 regiões com 3 zonas: ascendentes, corpo e descendentes.
[COR02]	Propõe uma abordagem de reconhecimento de manuscritos baseado na percepção humana. As características para extração são: pontos de fixação e detalhes da imagem nas posições vertical, horizontal e diagonal.
[PAR02]	Recomenda que um dado caractere seja dividido em N_f por N_f células de mesma área, para N_f é o tamanho da divisão, sendo que todas as células são usadas para compor o vetor de características.
[TAY01]	Extração de características geométricas para cada <i>frame</i> da imagem segmentada. Características geométricas: dimensão e aspectos do <i>bouding box</i> dos pixel nos <i>frames</i> , centro de gravidade, distâncias centro-zona, perfil em 8 direções, números de transições verticais e horizontais na direção diagonal de 45° e -45° .
[FRE01]	Primitivas geométricas. Propõem a rotulação do fundo da imagem (<i>background</i>).
[PED95]	Propõem um particionamento da imagem de entrada em sub-imagens (janelas). Um conjunto fixo de operadores é aplicado em cada janela. O conjunto de características consiste das saídas geradas em cada operador em cada janela. O primeiro operador é um contador simples de bits que calcula a densidade média dos <i>pixels</i> na janela. Os outros operadores tentam estimar a extensão de <i>pixels</i> pretos (significativos) alinhados na janela ao longo de algumas direções.

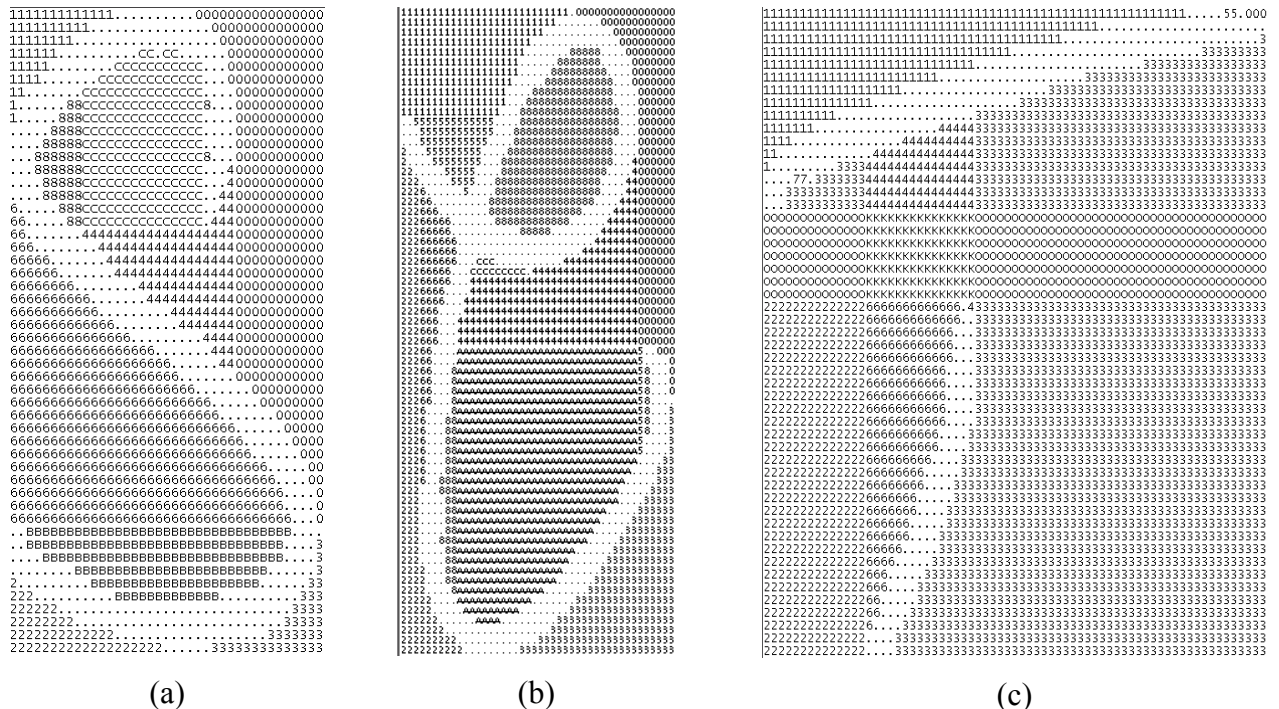


Figura 3.3: Rotulação do fundo da imagem (*background*)

Tabela 3.2: Rótulos do Vetor de Características

Rótulo	Tipo	Busca
0	Aberto à direita e acima	(NSLO)
1	Aberto à esquerda e acima	(NSLO)
2	Aberto à esquerda e abaixo	(NSLO)
3	Aberto à direita e abaixo	(NSLO)
4	Aberto à direita	(NSLO)
5	Aberto acima	(NSLO)
6	Aberto à esquerda	(NSLO)
7	Aberto abaixo	(NSLO)
8	Laço fechado	(NSLO)
A	Escape para direita e acima	Sondagem
B	Escape para esquerda e acima	Sondagem
C	Escape para direita e abaixo	Sondagem
D	Escape para esquerda e abaixo	Sondagem
E	Escape acima e direita	Sondagem
F	Escape acima e esquerda	Sondagem
G	Escapa abaixo e direita	Sondagem
H	Escape abaixo e esquerda	Sondagem
J	Aberto esquerda/direita	(NSLO)
K	Aberto acima/abaixo	(NSLO)
L	Fechado abaixo	(NSLO)
M	Fechado acima	(NSLO)
N	Fechado direita	(NSLO)
O	Fechado esquerda	(NSLO)

3.3 Representação das Primitivas

3.3.1 Formas de Representação

As principais formas de representação de características são [HEU94 *apud* KAP04]:

- **Vetores de características e matrizes:** Normalmente a imagem é dividida em zonas (*zoning*) utilizando-se uma grade fixa ou variável (segmentação implícita). Para cada zona se extraem vetores ou matrizes de dados;
- **Seqüências:** A imagem é representada por uma seqüência de símbolos (*codebooks*). A obtenção da seqüência respeita a ordem de ocorrência dos símbolos na imagem;
- **Estruturas de grafos:** A imagem é representada por um grafo tendo as primitivas como nós e a relação espacial entre estas como as arestas (ligações);
- **Contagem e verificação (*Assertions*):** Contagem de pixels, número de ascendentes, descendentes, laços, entre outros. A verificação da presença ou ausência de barras (letra T), entre outros, é muito utilizada em abordagens globais. Ela é freqüentemente utilizada para descartar objetos não similares.

Neste trabalho o estudo está relacionado com a criação de regiões (*zoning*), as quais compõem um vetor de características. Em cada região da imagem faz-se a contagem dos rótulos encontrados conforme Tabela 3.2 e, ainda, adiciona-se ao vetor de características a contagem de *pixel* preto em cada região. Todos os valores são normalizados pelo tamanho da região. A Figura 3.4 apresenta um exemplo de divisão do caractere T em 4 regiões.

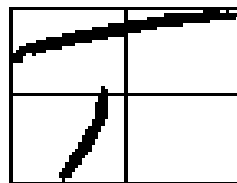


Figura 3.4: Divisão do caractere em 4 regiões

Uma imagem dividida em 4 regiões apresenta um vetor de características de dimensão 24 para cada região; neste caso a dimensão total do vetor é de 96. A Figura 3.5 representa este vetor para a primeira região da imagem. Sendo que cada posição do vetor corresponde a contagem de uma rotulação apresentada na Tabela 3.2 encontrada na imagem.

0	1	2	3	4	5	6	7	8	9	10	11	12	13
R01Q	R11Q	R21Q	R31Q	R41Q	R51Q	R61Q	R71Q	R81Q	RA1Q	RB1Q	RC1Q	RD1Q	RE1Q

14	15	16	17	18	19	20	21	22	23
RF1Q	RG1Q	RH1Q	RJ1Q	RK1Q	RL1Q	RM1Q	RN1Q	RO1Q	CPPRETO

Figura 3.5: Representa parte de um vetor de características – 1º região da imagem

3.3.2 Percepção Humana das Formas

Em Freitas [FRE02] realiza-se uma revisão dos aspectos da percepção visual, envolvendo as abordagens analítica e sintética, do processo de leitura e reconhecimento de palavras manuscritas, relacionando estes aspectos de forma a auxiliar no entendimento do processo humano de reconhecimento de formas manuscritas.

Enfatiza-se a necessidade de incorporar a percepção humana para contribuir no processo de reconhecimento. Freitas [FRE02] relaciona a percepção humana e a leitura com a extração de primitivas para os processos automáticos de reconhecimento.

Observe a Figura 3.6, se você leu "NÃO PARE NA PISTA", é melhor olhar de novo. Existem pessoas que olham várias vezes e não vêem onde está o erro. Esse fenômeno deve-se ao um fato muito comum de ver os padrões globalmente, sem notar os detalhes. Os psicólogos chamam isso de Gestalt [FRE02].

Gestalt é uma Escola de Psicologia Experimental alemã que teve sua origem com Max Wertheimer (1880-1943), Wolfgang Köhler (1887-1967) e Kurt Koffka (1886-1941), por volta de 1910, esta atua efetivamente na área da teoria da forma, com contribuições relevantes aos estudos da percepção, linguagem, memória, inteligência, entre outros. A teoria da Gestalt sugere uma resposta ao porque de certas formas agradarem mais que outras, não baseia-se no subjetivismo do "feio x bonito", mas sim apoiada na fisiologia do sistema nervoso e na psicologia, sempre através de rigorosos experimentos e pesquisas.



Figura 3.6: Placa de sinalização

Fonte: Adaptado de [FRE02] <<http://omnis.if.ufrj.br/~coelho/DI/olho.html>>

O princípio enunciado por Wertheimer sobre a organização perceptiva demonstra que o olho humano tende a agrupar as várias unidades de um campo visual para formar um todo.

Segundo essa teoria, o organismo percebe um conjunto de elementos como uma forma completa em que os componentes estão integrados entre si, de um modo que não é possível decompor-los sem destruir o conjunto. Ou seja, um conjunto passa a ser uma nova entidade, que não é simplesmente a soma dos seus componentes, mesmo porque esses componentes podem ser dispostos de diferentes maneiras, formando diferentes conjuntos [GOM02 *apud* FRE02].

Segundo a Gestalt, a percepção da forma pelo cérebro é sempre uma percepção global dos estímulos, ou seja, o cérebro não enxerga elementos isolados, e sim as relações entre eles. Portanto, enxergamos o todo e não partes dele.

[FRE02] utiliza os conceitos da Gestalt para buscar um relacionamento da percepção humana e o reconhecimento de palavras manuscritas. Apresenta-se a seguir os princípios da Gestalt com os quais pode-se detalhar estas forças de percepção:

- Princípio da Proximidade: descreve a tendência onde elementos individuais são intensamente associados com os elementos mais próximos do que com os que estão distantes. Pode-se observar este fenômeno em dois níveis diferentes na Figura 3.7. Os olhos organizam primeiro os pontos em quatro colunas porque a separação horizontal é maior que a separação vertical. Então, devido a separação entre as duas "colunas" do meio ser maior que as brechas externas, a figura toda é vista como dois grupos de duas colunas cada.

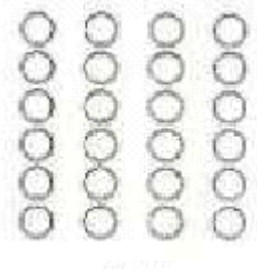


Figura 3.7: Princípio Gestalt da proximidade
Fonte: Adaptado de [FRE02]

- Princípio da Similaridade: observa-se que alguns elementos são associados com mais intensidade quando eles compartilham de características visuais básicas (como é o caso das variáveis visuais de forma, tamanho, cor, textura, valor e orientação) do que quando eles diferem nessas dimensões. A Figura 3.8 apresenta novamente dois grupos de duas colunas cada, apesar do espaçamento interelementos e intercolunas terem sido igualados.

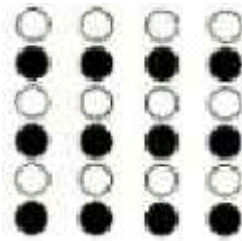


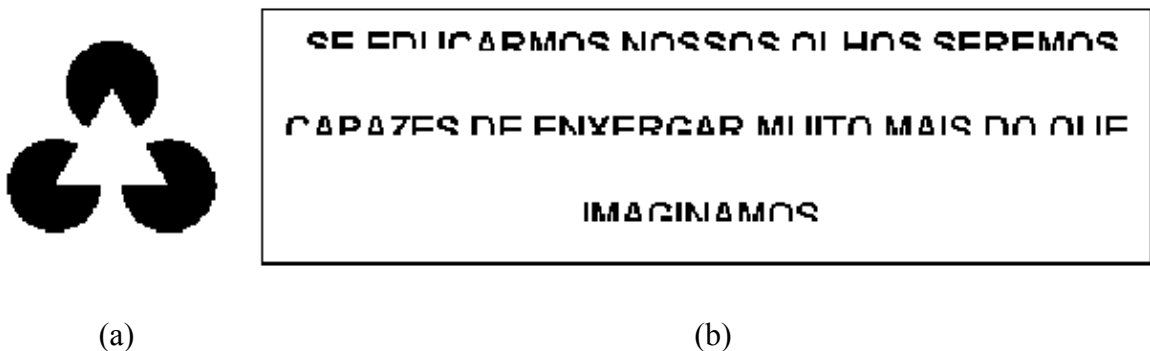
Figura 3.8: Princípio Gestalt da similaridade
Fonte: Adaptado de [FRE02]

- Princípio da Continuidade: descreve a preferência pelos contornos contínuos e sem quebra ao invés de outras combinações mais complexas, mas igualmente aceitáveis de figuras mais irregulares. A forma da Figura 3.9 é então percebida como duas linhas que se cruzam ao invés de quatro linhas que se tocam ou dois (ou mesmo quatro) ângulos opostos.



Figura 3.9: Princípio Gestalt da continuidade.
Fonte: Adaptado de [FRE02]

- Princípio do Fechamento: descreve a tendência humana de interpretar o estímulo visual como completo, como figuras fechadas, até quando algumas das informações de contorno estão ausentes. A Figura 3.10-a é instintivamente vista como um triângulo sobreposto em três círculos completos mesmo que nenhuma destas formas esteja tecnicamente presente. Este exemplo clássico é conhecido como Triângulo de Kanizsa. A Figura 3.10-b exemplifica o princípio do fechamento para a leitura de palavras incompletas [MAC96 *apud* FRE02].



(a) (b)
 Figura 3.10: Princípio do Fechamento: a) triângulo de Kanizsa e b) palavras incompletas.
 Fontes: Adaptado de [FRE02] <<http://www.psicologia.freeservers.com/gestalt/>>

O princípio do fechamento é importante para a formação de unidades, uma vez que a Gestalt parte da teoria que a primeira sensação da forma é global e unificada. Não se vê partes isoladas, mas relações. A Figura 3.11 exemplifica este princípio, pois se percebe facilmente um cavalo, um triângulo e uma mulher de vestido preto. Comprova-se assim, que existe a tendência psicológica de se unir intervalos e estabelecer ligações [FRE02].

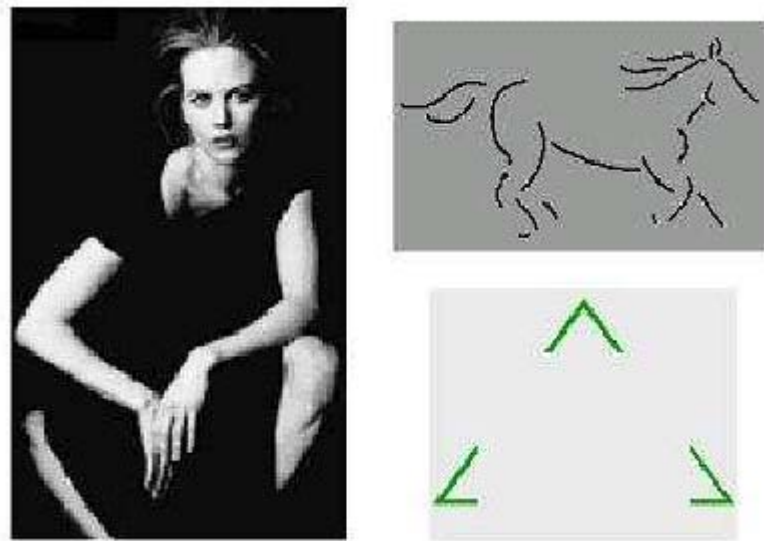


Figura 3.11: Princípio da Gestalt: fechamento

Fonte: Adaptado de [FRE02]

http://www.belasartes.br/aulas_virtuais/joaogomes/gestalt/leitura-visual/

- Princípio de Figura-Fundo: baseia-se no fato de que a atenção visual do observador alterna entre o padrão que emerge como fundo ou figura. Por exemplo, na Figura 3.12 o observador pode ver inicialmente um vaso branco em fundo preto e se alternar seu objeto de interesse poderá ver duas faces pretas uma de frente para a outra. Este exemplo clássico foi apresentado por Rubin em 1921 e é conhecido como *Face-Vase*. Ambos podem ser vistos como figura (objeto de interesse) ou como fundo sobre o qual está apoiada a figura.



Figura 3.12: Princípio da Gestalt: figura-fundo.

Fonte: Adaptado de [FRE02] <http://www.psicologia.freeservers.com/gestalt/>

- Princípio da Área Relata: onde a menor de duas figuras sobrepostas tenderá a ser interpretada como figura, enquanto que a maior será interpretada como

fundo. Na Figura 3.13-a, o quadrado interno é percebido como uma forma distinta na frente de um quadrado maior, em vez de um buraco em uma forma maior. Sabe-se ainda, que o princípio da simetria descreve o agrupamento baseado nas propriedades emergentes da forma, ao invés das características das partes que a constituem, ou seja, a Figura 3.13-b é vista como dois objetos sobrepostos ao invés de três.

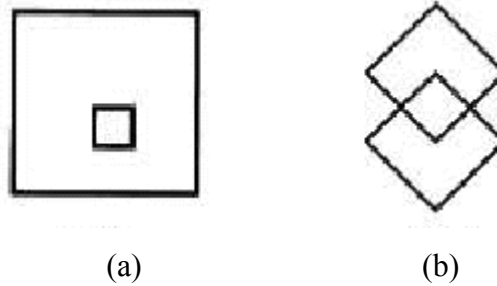


Figura 3.13: Princípio de área relativa

Fonte: Adaptado de [FRE02]

- Princípio da Pregnância: este princípio é geral e abrange os demais princípios. Este princípio afirma que as forças de organização tendem a se dirigir da melhor forma possível, no sentido da clareza, unidade e equilíbrio, tanto quanto permita a imagem dada, ou seja, quanto melhor a forma, mais pregnância ela terá, e melhor será sua relação com o cérebro. Uma imagem de boa Gestalt é enxergada com muito mais clareza pelo cérebro, e conseqüentemente de forma harmoniosa. É importante lembrar que para a Gestalt a noção de unidade é primordial, pois para a percepção as partes são inseparáveis do todo e, ainda, que a imagem percebida é o resultado da interação das forças externas (luz na retina) com as forças internas (a tendência de organizar da melhor forma possível os estímulos externos). Assim, afirmou Wertheimer em 1910: "O todo é mais que a soma das partes". A Figura 3.14 apresenta diferentes formas de uma mesma frase, percebe-se que a primeira e última frase possuem maior pregnância devido ao fato de que pode-se dizer que tem melhor relação com o cérebro.



Figura 3.14: Princípio da pregnância
 Fonte: Adaptado de [FRE02]

Para [AUM93 *apud* FRE02] espera-se do mundo que o mesmo tenha sempre a mesma aparência ou espera-se que se encontre uma certa quantidade de elementos invariáveis. É a percepção desses aspectos invariantes do mundo (tamanho dos objetos, forma, localização, orientação, propriedade das superfícies, etc) que se designa pela noção de constância perceptiva. Ou seja, apesar da variedade de percepções, localizam-se as constantes.

Sabe-se, por experiência própria, que os olhos estão equipados para localizar pequenos detalhes. Alguns traços relevantes podem identificar um objeto ou uma pessoa. Por exemplo, quando um cartunista cria uma semelhança expressiva de uma pessoa utilizando somente algumas linhas bem escolhidas. Ou ainda, quando se reconhece um amigo a grande distância unicamente pelas proporções e movimentos mais elementares.

Durante o processo de percepção de um objeto, os olhos humanos se movimentam sucessivamente e se fixam nas partes mais relevantes da imagem. Para [ARN97 *apud* FRE02] capta-se um rosto humano, exatamente como todo o corpo é captado, ou seja, como um padrão global de componentes essenciais (olhos, nariz, boca) aos quais se pode agregar detalhes.

A Figura 3.15 demonstra um procedimento que utiliza uma câmera para descrever os movimentos oculares realizados pelo ser humano quando da percepção e reconhecimento de

faces. Estas imagens foram geradas pelos pesquisadores do A. B. Kogan Research Institute for Neurocybernetics da Rostov State University - Rússia.

Em [FRE02] apresenta-se que os estudos da percepção visual ocorrem sob duas abordagens: a abordagem analítica e a abordagem sintética, a saber:

- *Abordagem Analítica*: esta abordagem parte de uma análise da estimulação do sistema visual pela luz, buscando fazer com que os componentes assim isolados correspondam a diversos aspectos da experiência perceptiva dela [AUM93 *apud* FRE02]. Esta abordagem vem ao encontro da idéia de que o cérebro possui células especializadas nas funções elementares (percepção das bordas, linhas, dos movimentos direcionais, entre outros). Esta hipótese consiste em supor que, entre uma família de objetos, o observador escolhe uma e só uma opção. Assim, em seguida aplica-se repetidamente esta hipótese por "tentativa e erro". Portanto, se uma opção se revela errônea, o sistema visual revê suas opções de invariância e emite outras opções, de forma a fazer coincidir todas as opções com uma configuração possível (aqui o sistema leva em conta a experiência adquirida e as associações possíveis);
- *Abordagem Sintética*: para esta abordagem a imagem óptica na retina contém toda a informação necessária à percepção dos objetos no espaço. Assim, a abordagem sintética se opõe as teorias que supõem uma aprendizagem da visão [AUM93 *apud* FRE02]. Vale ressaltar que se entende como aprendizagem da visão a capacidade do sistema visual como um todo aprender, ou seja, inclui-se aí o cérebro. Para esta teoria, cada imagem provoca uma percepção global única. Sendo assim, não cabe ao sistema visual decodificar as informações, nem tampouco, construir percepções, mas sim extrair informações. A percepção é, então, uma atividade direta.

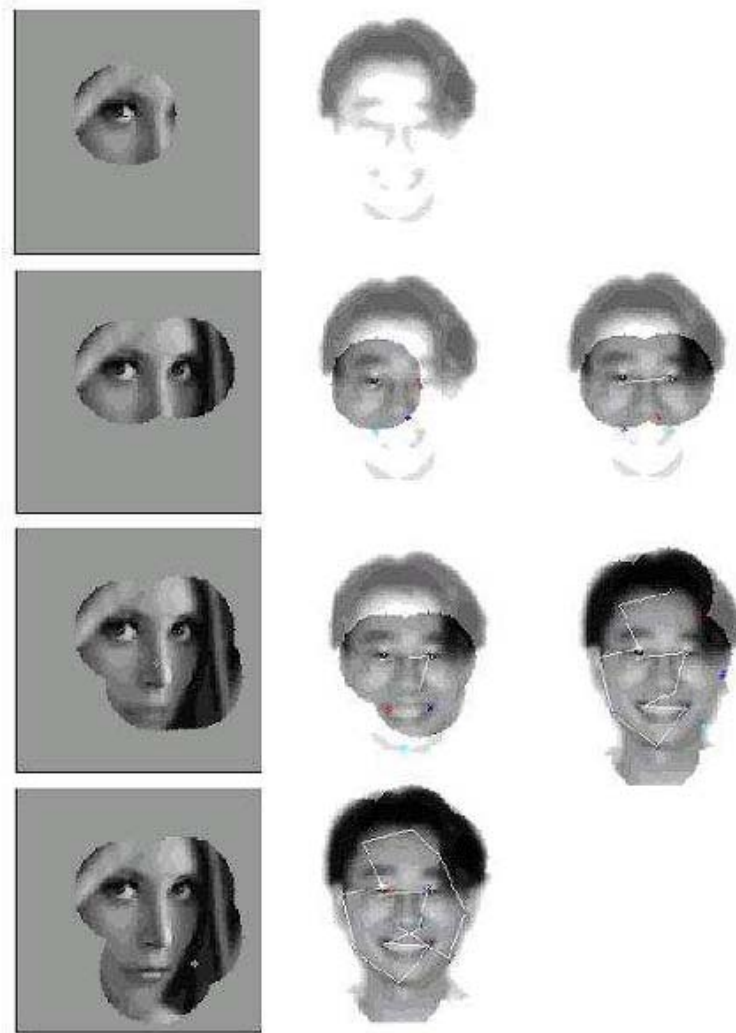


Figura 3.15: Movimentos oculares durante a percepção visual.
 Fonte: Adaptado de [FRE02] (<http://www.rybak-et-al.net/vnc.html>)

Para [FRE02] não existe uma melhor abordagem. Para palavras manuscritas, o reconhecimento consiste em a partir de uma forma (uma palavra ou caractere) desconhecida estabelecer um conjunto de formas conhecidas, verificando entre as formas desse conjunto as quais mais se assemelha a forma desconhecida, e sobre tudo isso tomar uma decisão da melhor hipótese. Essa decisão é realizada geralmente medindo-se a semelhança da forma desconhecida com um conjunto de referência (ou modelos) armazenados na memória e descritos em uma representação análoga. As referências ou modelos são obtidos através da etapa denominada treinamento.

Nos sistemas de reconhecimento de manuscritos uma dada palavra é ligada a uma forma particular. Após a fase de aquisição do sinal, a extração de primitivas permite

transformar a forma em uma representação, mais fácil de manipular que a forma original. O reconhecimento consiste em decodificar essa representação atribuindo-se uma das classes conhecidas ou referências do sistema.

Conclui [FRE02] que a percepção visual e o reconhecimento devem ser considerados como processos comportamentais e, provavelmente, não podem ser completamente entendidos pelos limitados esquemas dos sistemas computacionais sem que se leve em consideração aspectos comportamentais e cognitivos destes processos.

No trabalho apresentado por Correia et al. [COR02] a percepção humana pode ser avaliada através de alguns aspetos:

- os olhos se movimentam e se fixam sucessivamente nos pontos da imagem como maior informação;
- os olhos executam ativamente uma coleção de informação seletiva e orientação de problemas do mundo visível;
- os neurônios no córtex visual executam a seleção orientados pelo sentido da detecção das bordas e de barras locais. A Figura 3.16 descreve este modelo.

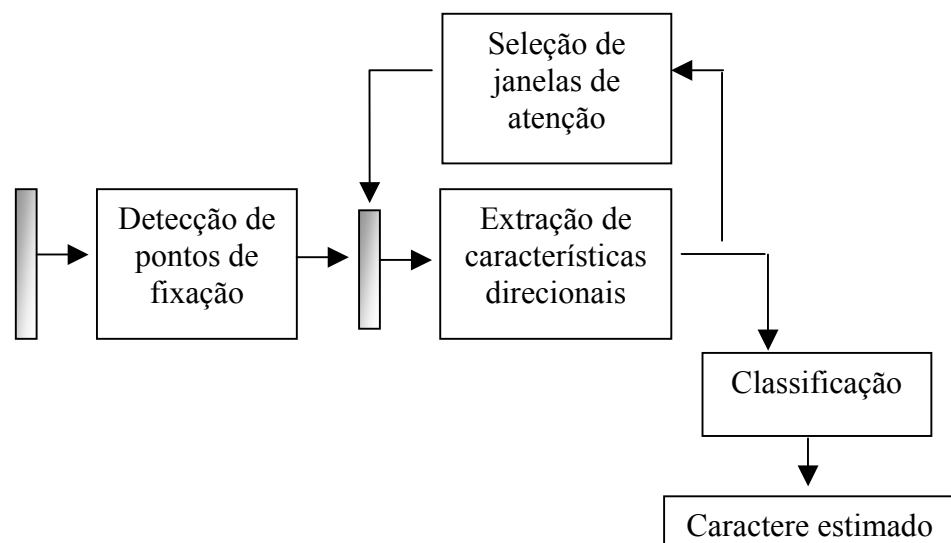


Figura 3.16: Arquitetura da percepção humana no sistema de reconhecimento de caracteres. Fonte: Adaptados de [COR02]

Em seu experimento Correia et al. [COR02] apresenta um zoneamento por janelas de atenção, onde cada ponto de fixação simula o movimento dos olhos e permite localizar direções seletivas para extração de características. Os pontos de fixação determinam “onde”

está a informação nas janelas de atenção, e a extração de transformadas Wavelet-2D identificam a informação. A Figura 3.17 exemplifica os pontos de fixação e as respectivas janelas de atenção. Correia et al. [COR02] apresenta os resultados de seus experimentos com uma taxa de reconhecimento de 98,25% usando a base de dados NIST.

O estudo desenvolvido por Yang [YAN98] realiza a imitação de processos psicológicos humanos em reconhecimento de caracteres chineses. O estudo compara os processo de padrões globais e análise de sub-padrões que são similares as rotinas de reconhecimento humano. No processo desenvolvido cria-se uma hierarquia de estruturas do caractere, sendo o caractere subdividido em: radicais, *strokes*, *stroxels*, conforme descreve a Figura 3.18.

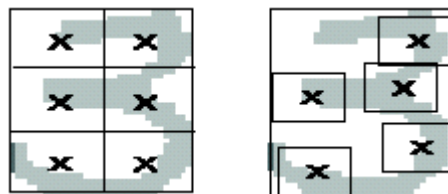


Figura 3.17: Exemplos de pontos de fixação e suas respectivas janela de atenção

Fonte: Adaptado de [COR02]

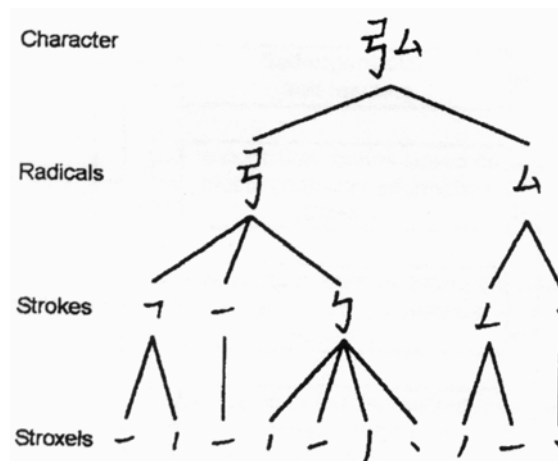


Figura 3.18: Subdivisão do caractere

Fonte: Adaptado de [YAN98]

3.3.3 Mecanismos de Zoneamento

Para Suen [SUE94], é necessário capturar indícios da percepção humana e reconhecimento de padrões para embutir nas máquinas. Em uma contínua busca para

distinguir características dos caracteres, as imagens podem ser divididas em partes: direita, esquerda, acima, meio, e abaixo, como mostrado na Figura 3.19.

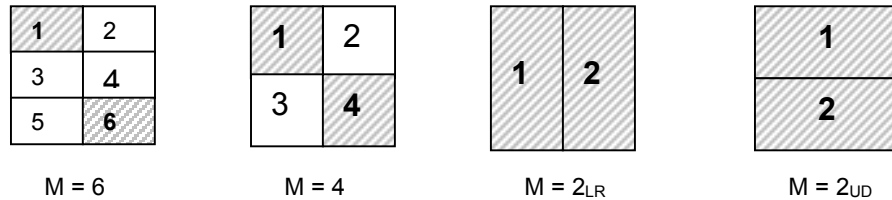


Figura 3.19: Partições onde 1,2,3,4,5 e 6 são o número de posições das partes

Fonte: Adaptado de [SUE94]

Para seus experimentos foram utilizados 89 padrões, entre 10 numerais e 26 letras do alfabeto inglês, estas escolhidas por experimentos computacionais e humanos. Cada padrão pode ser dividindo em até seis partes. A Figura 3.20 apresenta a divisão da letra k conforme os padrões de zoneamento proposto por Suen et al [SUE94].

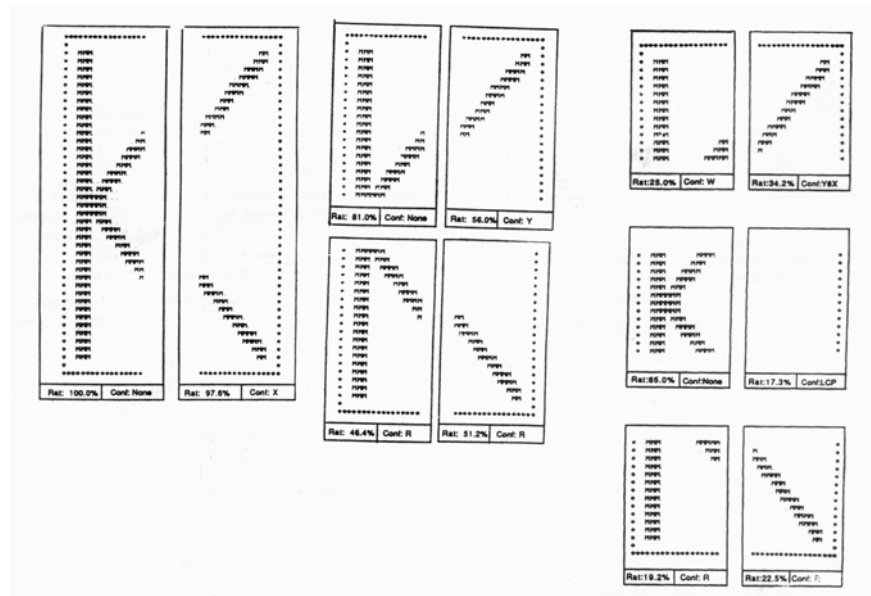


Figura 3.20: Divisão do caractere K conforme padrão de zoneamento

Fonte: Adaptado de [SUE94]

Suen et al. [SUE94] observa que existem partes do padrão chamadas de cruciais, onde o reconhecimento é perfeito, ou seja, 100%. O número de padrões identificados como perfeitos dependem largamente do número de partições M utilizadas. Encontrar partes distintas de várias partições de um padrão é um passo preliminar em direção a uma robusta

análise de padrões e reconhecimento de caracteres. As partes cruciais são partes efetivas no reconhecimento de padrões, porque são absolutamente distintas de outras partes. Depois da partição todas as partes dos caracteres tem uma diferença gradual. Entretanto, a parte crucial apresenta uma característica invariante e distinta das outras características do caractere original. As idéias e definições dadas como partes cruciais, partes de confusão e identificação perfeita de padrões podem ter um papel chave no reconhecimento de padrões e caracteres. Neste trabalho observa-se que a letra “D” sempre fica no topo (100%), letras como “A”, “K” e “G” resultam em uma taxa reconhecimento mais alta (100%) , sendo que “P”, “I” e “T” (54%) e, as taxas de reconhecimento considerando $Z=2LR$ (direita/esquerda), 2UD (inferior/superior), 4 e 6 eram: 86,12%, 85,88%, 61,73% e 42,91%, respectivamente. Os autores comentam sobre o caso 2LR para “Y” e explicam que este zoneamento está perfeito para reconhecimento; mas traz uma dificuldade para a letra “B” porque a esquerda e o meio está confundindo com “E”. Então, nota-se que diferentes partições podem produzir diferenças grandes em relação as taxas de reconhecimento. Mais partições provocarão mais partes de confusão. Por exemplo, em $Z=6$ um caractere pode ser confundido com 6 caracteres diferentes, “B” é confundido com: “C”, “G”, “J”, “O”, “S”, “U”.

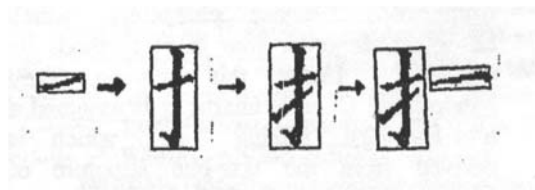


Figura 3.21: Representa a formação das caixas. Fonte: Adaptado de [HUN00]

No estudo apresentado por [HUN00] um esquema para caracteres é assinalado. Uma caixa de código de n ou menos dígitos representa um tipo de movimento do caractere. Este esquema independe do estilo da escrita e da seqüência de movimentos. Os símbolos Chineses são formados por segmentos de linha chamados de “*strokes*”. Os movimentos básicos são agrupados em um pequeno número de categorias ou tipos de movimentos para a classificação. Quando um caractere é impresso, a seqüência de movimentos s_1, s_2, \dots, s_n , representa uma ordem em que os caracteres são impressos. A Figura 3.21 apresenta a formação das caixas de código.

As caixas são reordenadas em seqüências de caixas $b_1, b_2, b_3, \dots, b_m$, de acordo com as regiões, na qual eles residem. As regiões são criadas desenhando linhas verticais e horizontais que divide o caractere impresso. Inicialmente o caractere é dividido em regiões desenhando

exclusivamente linhas verticais, ou linha horizontais que cortam o caractere sem cortar quaisquer das caixas. Preferências são dadas as linhas horizontais ou verticais que podem ser usadas para cortar os caracteres. As regiões futuramente são divididas em sub-regiões e repetidas da mesma maneira em cada sub-região até que se tenha somente uma caixa. A Figura 3.22 apresenta um exemplo desta divisão.

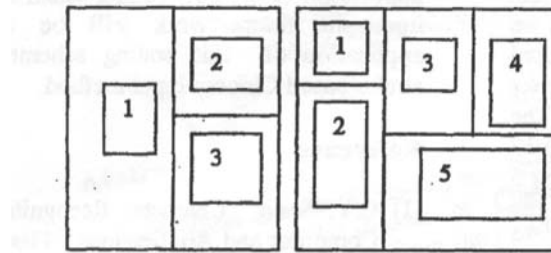


Figura 3.22: Divisão das caixas de código

Fonte: Adaptado de [HUN00]

Os movimentos em cada caixa são classificados em t predeterminado tipos de movimentos, e um código de reordenação inicial é gerado, e subseqüentemente melhorado, com referência para o canto à esquerda superior da caixa. A Tabela 3.3 apresenta as interseções e as distâncias de prioridade. A Formação da caixa de código é realizada onde todos os códigos reordenados são concatenados em uma string de acordo com as seqüências das caixas.

Tabela 3.3: Interseções e prioridades. Fonte: Adaptado de [HUN00]

Stroke Name	Intersection Priority	Distance Priority
Horizontal	2	1
Vertical	4	3
Left-Falling	5	2
Right-Falling	6	5
Left-Turning	1	6
Right-Turning	3	4

No trabalho apresentado em [AIR05a] é proposto um mecanismo de zoneamento não simétrico, utilizando um vetor de características baseadas em convexidade e concavidade extraídas pela rotulação do *background* (conforme apresentado na seção 3.2). Para isto, o caractere é inserido em uma “caixa” e particionado em Z partes, onde $Z = 4, 5H$ (horizontal), $5V$ (vertical) e 7 , conforme Figura 3.23.

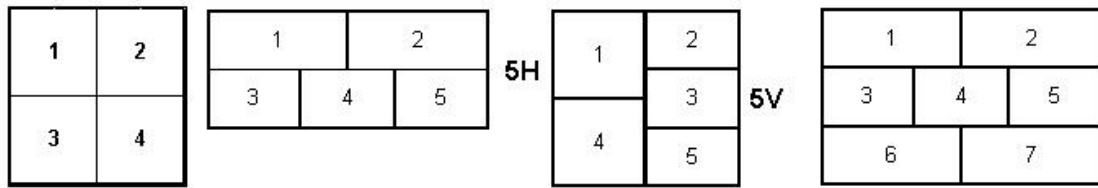


Figure 3.23: Mecanismo de Zoneamento: $Z = 4$, 5H, 5V e 7 regiões

Para o presente estudo o mecanismo de *zoning* tem por base as áreas propostas por Suen [SUE94]. Características da percepção humanas são utilizadas durante os estudos a fim de contribuir com a definição das regiões perceptivas relacionadas com os pontos de atenção apresentados por Correia et al. [COR02].

3.4 Classificador Neural

3.4.1 Redes Neurais

Redes Neurais são utilizadas para diversos propósitos em reconhecimento de padrões e áreas afins, como agrupamentos, classificação, regressão, aproximação de funções, entre outros. Dada a grande diversidade de aplicações, existem diferentes tipos de redes neurais, que diferem entre si pelo tipo de grafo subjacente, algoritmo de treinamento, tipo de processamento executado nos neurônios, etc.. Um dos modelos mais difundido, responsável pela retomada dos pesquisas na área à partir de meados dos anos de 1980, é o modelo *Multi Layer Perceptron* (MLP).

Considerando os métodos de classificação apresentados no item 2 (Revisão Bibliográfica) e tendo em vista que o conjunto de primitivas deste trabalho possui sua representação através de vetores, optou-se por um classificador neural.

As redes neurais artificiais (RNA) são conjuntos de elementos de processamento (PE) interconectados, denominados de neurônios, células ou nós, cada qual realizando cálculos. Elas podem possuir diversas entradas, sendo que um valor (peso) é associado a cada uma, formando as ligações entre os PEs (sinapses). As saídas dos PEs também podem ser entradas de outros PEs. O valor associado a qualquer neurônio é chamado de sua ativação e representa a soma ponderada das entradas. Ou seja, para um neurônio k :

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

Para N sendo o número de entradas do neurônio, x_j as entradas do neurônio e w_{kj} os pesos sinápticos associados a cada entrada. A Figura 3.24 ilustra o modelo do neurônio.

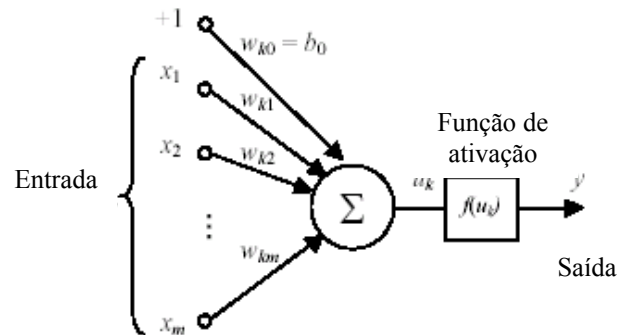


Figura 3.24: Modelo de neurônio. Fonte: Adaptado de [OLI02]

As redes neurais possuem sua força na sua capacidade de gerar uma região de decisão a partir de uma forma qualquer (aproximadores universais), requerida por um algoritmo de classificação, ao preço da integração de camadas de células suplementares nas redes.

Os trabalhos realizados por diversos autores, tais como: [OLI02][PED95][COR02][SIL03], utilizam redes neurais para o reconhecimento de palavras e caracteres manuscritos.

3.4.2 Arquitetura Classe-Modular

Para Oh & Suen [OH_02] as redes neurais 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 dimensional elevado de características, isto devido a complexidade do problema como é o caso do reconhecimento de caracteres manuscritos. As limitações existem também em diversos aspectos dos processos de treinamento e de reconhecimento. Para tanto, introduz-se o conceito da modularidade de classes ao classificador da rede neural na tentativa de superar tais limitações. No conceito classe-modular (*Class-Modular*), o problema original K da classificação decompõem-se em K -subproblemas. Uma arquitetura modular consiste em K sub-redes, cada uma responsável para discriminar uma classe das outras $K-1$ classes, como apresentado na Figura 3.25. Em seu trabalho os autores Oh & Suen [OH_02] apresentam a eficácia das redes neurais 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.

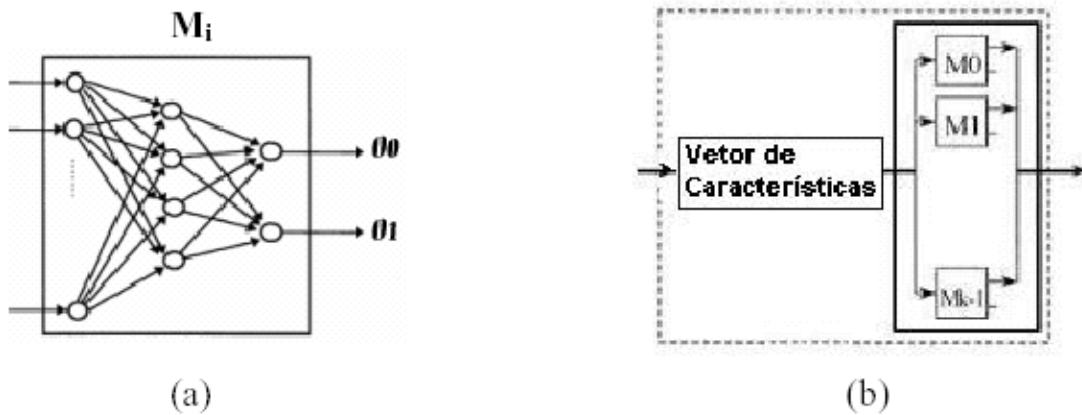


Figura 3.25: Arquitetura Classe Modular. a) Módulo M_i da rede e b) configuração completa com K módulos. Fonte: Adaptado de [KAP04]

Silva e Thomé apresentam [SIL03] a utilização de mais de uma rede neural para a classificação dos caracteres, em um “time” de redes neurais, que podem ser de um 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. Consideram [SIL03] que, os sub-problemas específicos são menores que o problema original, logo, as redes neurais que atuam nos mesmo são menores. Sendo assim, espera-se que estas redes menores alcancem um 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%.

Kapp et al. [KAP03] avaliam o uso da arquitetura convencional de MLP (*Multiple Layer Perceptron*) e classe-modular para o reconhecimento de manuscritos no contexto de cheques bancários. No modelo convencional MLP todas as classes são treinadas juntas e a classe que apresentar um valor máximo de saída é considerada com a classe reconhecida. Já no modelo classe-modular MLP para cada K o classificador é treinado independente das outras classes utilizando o conjunto de treinamento e validação. As experiências mostraram que a arquitetura classe-modular é melhor do que arquitetura convencional. As taxas médias

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

Aires et al [AIR05b] visa contribuir com o estudo de redes neurais que utilizam um conjunto de primitivas extraído com base em regiões perceptivas para o processo de reconhecimento das formas. Trata-se de uma abordagem analítica do problema, tendo em vista um léxico de pequena dimensão: caracteres manuscritos maiúsculos, ou seja, 26 letras do alfabeto. Para a tarefa de reconhecimento são realizados experimentos que utilizam redes neurais artificiais com arquitetura MLP (*Multiple Layer Perceptron*) Classe-Modular, ou seja, um comitê de redes neurais MLP, de forma que cada classe do problema possui uma rede específica, assim como descrito em [Kapp et al. 2003] [Oh e Suen 2002] [Silva e Thomé 2003]. No trabalho determinaram-se redes neurais especializadas em sub-problemas buscando melhorar o desempenho do sistema como um todo. Deste modo, conclui-se que algumas representações são mais robustas e discriminantes entre as classes de caracteres do que outras e, sendo assim, as combinações dos classificadores com base em diferentes representações resultam em taxas de reconhecimento mais elevadas.

Para o presente estudo definiu-se a aplicação das redes neurais artificiais com arquitetura MLP (*Multiple Layer Perceptron*) Classe-Modular, ou seja, um comitê de redes neurais MLP, de forma que cada classe do problema possui uma rede específica, assim como descrito em [OH_02] e [KAP03]. Esta arquitetura permitirá estudar as regiões perceptivas mais significativas para cada classe de caractere a ser reconhecido.

3.4.3 Múltiplos Classificadores

Métodos de extração de características e técnicas de classificação foram muito estudados nas últimas décadas para auxiliar no reconhecimento de manuscritos. Muitos métodos de reconhecimento foram propostos, mas isoladamente nenhum conseguiu uma solução completa para o problema. Entretanto, algumas técnicas de combinação de múltiplos classificadores foram propostas para melhorar o desempenho desses sistemas, apresentando resultados promissores. Esses métodos, em sua maior parte, são propostos utilizando o resultados dos classificadores individuais que são combinados de acordo com diversas estratégias [OLI04].

Observa-se em alguns estudos de classificação [LAM95] [KIT98] [LIU02], que o conjunto de padrões reconhecidos erroneamente pelos diferentes classificadores não são

necessariamente os mesmos. Para [OLI04], isto sugere que diferentes classificadores oferecem informações complementares sobre os padrões, podendo melhorar o desempenho do classificador selecionado. A idéia é utilizar todos os sub-conjuntos para a tomada de decisão, combinando as opiniões individuais de modo a obter uma opinião em consenso.

Basicamente, pode-se dividir os problemas de classificação em dois cenários distintos [OLI04]:

- 1) Todos os classificadores utilizam o mesmo padrão de entrada, por exemplo, um conjunto de classificadores k-NN, utilizando o mesmo vetor de características, mas diferentes parâmetros de classificação; outro exemplo é um conjunto de classificadores neurais de arquitetura fixa, apresentando conjuntos de pesos distintos obtidos por diferentes estratégias de treinamento.
- 2) Cada classificador utiliza sua própria representação dos padrões de entrada. Ou seja, as características extraídas dos padrões são únicas para cada classificador, permitindo desta forma, integrar medidas/características fisicamente diferentes.

Xu et al. [XU_92] apresenta uma definição matemática da combinação de múltiplos classificadores. Dado um espaço de padrões P constituído de M conjuntos mutuamente exclusivos $P = C_1 \cup \dots \cup C_M$ para cada $C_i \forall_i \in \Lambda = \{1, 2, \dots, M\}$ representa um conjunto de padrões específicos, denominado classe. Dada uma amostra x de P , a tarefa do classificador e é atribuir a x um índice $j \in \Lambda \cup \{M + 1\}$ como rótulo para representar que x é observado como sendo da classe C_j se $j \neq M + 1$, sendo que $j = M + 1$ representa que x é rejeitado por e .

[Pham *apud* OLI04], apresenta que diversos métodos de fusão tem sido utilizados para combinar múltiplos classificadores, tais como: formalismo bayesiano ou probabilidade máximo *a posteriori*, integral *fuzzy*, regras *fuzzy*, cadeias de *Markov*, teoria da evidência (regra *Dempster-Shafer*), redes neurais, voto majoritário, entre outros. O método *bayesiano* atribui um objeto desconhecido à classe que possua a máxima probabilidade *a posteriori*. A regra de *Dempster-Shafer* combina evidências de diferentes fontes de informação baseada em uma função chamada de atribuição de probabilidade básica. No voto majoritário, se a maior quantidade de classificadores rotulam uma amostra para uma classe, mais do que para outra, então a amostra é atribuída àquela classe. O método baseado em rede neural tal como rede multicamadas é usado para combinar vários classificadores usando dados de treinamento destes classificadores. Na aplicação da integral *fuzzy*, com uma ferramenta de combinação, foi

apresentado um método baseado no conceito de combinar imagens características de diversas fontes para o reconhecimento automático.

Pode-se classificar os métodos de combinação como os baseados em regras fixas ou estáticas e baseados em treinamento [WEB02 *apud* MAT04]. A seguir serão estudadas algumas regras de combinação [MAT04].

3.4.3.1 Métodos baseados em regra fixa

Estes métodos baseiam-se na regra de decisão *bayesiana*. Sendo Z um objeto que se deseja classificar e, tendo L classificadores com entradas x_1, \dots, x_L . Sendo a regra de *Bayes* para erro mínimo atribui Z à classe w_j se:

$$p(w_j | x_1, \dots, x_L) > p(w_k | x_1, \dots, x_L); \quad k = 1, \dots, C, \quad k \neq j, \quad (3.1)$$

ou, de forma equivalente, atribui Z à classe w_j se

$$p(x_1, \dots, x_L | w_j) > p(x_1, \dots, x_L | w_k); \quad k = 1, \dots, C, \quad k \neq j, \quad (3.2)$$

Isto requer o conhecimento das probabilidades conjuntas $p(x_1, \dots, x_L | w_j)$, $j = 1, \dots, L$ que não são disponíveis.

3.4.3.1.1 Regra do produto

Assumindo que os classificadores são independentes, então a regra de decisão expressa pela Equação 3.2 torna-se:

Atribua Z a classe w_j se,

$$\prod_{i=1}^L p(x_i | w_j) p(w_j) > \prod_{i=1}^L p(x_i | w_k) p(w_k); \quad k = 1, \dots, C, \quad k \neq j, \quad (3.3)$$

ou, em termos de probabilidade *a posteriori* dos classificadores individuais:

Atribua Z à classe w_j se,

$$[p(w_j)]^{-(L-1)} \prod_{i=1}^L p(w_j | x_i) > [p(w_k)]^{-(L-1)} \prod_{i=1}^L p(w_k | x_i); \quad k = 1, \dots, C, \quad k \neq j. \quad (3.4)$$

Esta é a regra do produto, podendo ser simplificada considerando-se $p(w_i)$ equiprovável, de modo que, atribua a Z à classe w_j se,

$$\prod_{i=1}^L p(w_j | x_i) > \prod_{i=1}^L p(w_k | x_i); \quad k = 1, \dots, C, \quad k \neq j. \quad (3.5)$$

Assumir a hipótese de independência condicional pode ser muito rigorosa, mas essa condição tem sido utilizada satisfatoriamente em muitos problemas práticos. A regra requer o conhecimento das probabilidades *a posteriori* dos classificadores individuais, $p(w_j|x)$, $j=1,\dots,C$, que podem ser estimados a partir dos dados de treinamento. O problema principal deste método é sua sensibilidade à erros na estimativa das probabilidades *a posteriori*, o que prejudica o método quando os erros aumentam. Se um dos classificadores determinar que a probabilidade de que uma dada amostra pertença a uma classe em particular é nula, então a regra do produto atribuirá uma probabilidade zero, mesmo que os outros determinem que esta é a classe mais provável.

3.4.3.1.2 Regra da soma

Consiste numa variação da regra do produto, pela introdução da hipótese de que $p(w_k|x)$ é próximo de $p(w_k)$. Ou seja, admite-se que:

$$p(w_k|x_i) = p(w_k)(1 + \delta_{ki}) \quad (3.6)$$

com $\delta_{ki} \ll 1$, isto é, as probabilidades *a posteriori* $p(w_k|x_i)$ usadas na regra do produto, dada pela Equação 3.5, não são substancialmente diferentes das probabilidades *a priori* $p(w_k)$. Substituindo $p(w_k|x_i)$ na Equação 3.5 e desprezando os termos de segunda ordem e de ordens superiores em δ_{ki} e usando a Equação 2.6, temos a regra da soma:

$$(1-L)p(w_j) + \sum_{i=1}^L p(w_j | x_i) \succ (1-L)p(w_k) + \sum_{i=1}^L p(w_k | x_i); \quad k = 1, \dots, C, \quad k \neq j. \quad (3.7)$$

Esta é a regra da soma, que pode ser simplificada considerando-se $p(w_i)$ equiprovável: Atribua Z à classe w_j se,

$$\sum_{i=1}^L p(w_j | x_i) \succ \sum_{i=1}^L p(w_k | x_i); \quad k = 1, \dots, C, \quad k \neq j. \quad (3.8)$$

A hipótese usada para derivar a aproximação da regra da soma a partir da regra do produto não seria realista em muitas aplicações práticas. Contudo, esta regra é relativamente insensível a erros de estimação das densidades de probabilidades conjuntas e estudos comparativos mostram que a regra da soma é mais robusta a erros e reduz os efeitos de um possível superespecialização dos classificadores individuais.

Pode-se modificar a regra da soma introduzindo uma ponderação, de modo que se atribua Z à classe w_j se,

$$\sum \alpha_i \cdot p(w_j | x_i) > \sum \alpha_i \cdot p(w_k | x_i); \quad k = 1, \dots, C, \quad k \neq j, \quad (3.9)$$

em que α_i , $i = 1, \dots, L$ são pesos para os classificadores. Uma questão importante é a escolha dos pesos, que podem ser estimados usando o conjunto de treinamento para minimizar a taxa de erro do classificador combinado. Deste modo a mesma ponderação é aplicada em todo o espaço de características. Uma alternativa é permitir que os pesos variem com a localização da amostra no espaço de características. Um exemplo extremo disto é a *seleção dinâmica de classificadores* em que atribui-se o valor unitário a um dos pesos e anula-se os demais. Para um dado padrão, a seleção dinâmica procura selecionar o melhor classificador. De modo que o espaço de características é particionado em regiões com um classificador diferente para cada região.

3.4.3.1.3 Regra do máximo, mínimo e mediana

A regra de máximo pode ser obtida aproximando-se as probabilidades *a posteriori* mostrada na Equação 3.7 por um limiar superior, $L \max_i p(w_k | x_i)$, resultando na seguinte regra de decisão:

Atribua Z à classe w_j se,

$$(1-L)p(w_j) + L \max_i p(w_j | x_i) > (1-L)p(w_k) + L \max_i p(w_k | x_i); \quad k = 1, \dots, C, \quad k \neq j. \quad (3.10)$$

Este é a regra de máximo, que pode ser simplificada considerando-se $p(w_i)$ equiprovável:

Atribua Z à classe w_j se,

$$\max_i p(w_j | x_i) > \max_i p(w_k | x_i); \quad k = 1, \dots, C, \quad k \neq j. \quad (3.11)$$

Do mesmo modo, aproximando a regra do produto definida pela Equação 3.5 por um limiar superior, $\min_i p(w_k | x_i)$, resultando na seguinte regra de decisão:

$$\left[p(w_j) \right]^{-(L-1)} \min_i p(w_j | x_i) > \left[p(w_k) \right]^{-(L-1)} \min_i p(w_k | x_i); \quad k = 1, \dots, C, \quad k \neq j \quad (3.12)$$

Esta é a regra de mínimo, que pode ser simplificada considerando-se $p(w_i)$ equiprovável: Atribua Z à classe w_j se,

$$\min_j p(w_j | x_j) \succ \min_j p(w_k | x_j); \quad k = 1, \dots, C, \quad k \neq j. \quad (3.13)$$

Finalmente, a regra da mediana pode ser derivada observando que a regra da soma calcula a média das saídas dos classificadores e que uma estimativa robusta da média é a mediana. Então, considerando-se $p(w_i)$ equiprovável, a regra da mediana pode ser definida como: Atribua Z à classe w_j se,

$$\text{med}_j p(w_j | x_j) \succ \text{med}_j p(w_k | x_j); \quad k = 1, \dots, C, \quad k \neq j. \quad (3.14)$$

3.4.3.1.4 Voto majoritário

Aproximam-se as probabilidades *a posteriori* $p(w_k|x_j)$ por funções de valor binário Δ_{ki} de modo que:

$$\Delta_{ki} = \begin{cases} 1 & \text{se } p(w_k|x_j) = \max_i p(w_k|x_j) \\ 0 & \text{caso contrário} \end{cases} \quad (3.15)$$

Assumindo-se que $p(w_i)$ é equiprovável, determinamos a regra do voto majoritário da seguinte forma:

$$\sum_{i=1}^L \Delta_{ji} \succ \sum_{i=1}^L \Delta_{ki}; \quad k = 1, \dots, C, \quad k \neq j. \quad (3.16)$$

Para cada classe w_k , a regra do voto majoritário simplesmente conta os votos recebidos por esta hipótese pelos classificadores individuais. A classe que receber o maior número de votos é então definida como a decisão em consenso. Pode-se admitir que a classe seja considerada desconhecida caso exista empate entre os rótulos no processo de votação.

3.4.3.2 Métodos baseados em treinamento

Estes métodos requerem algum grau de treinamento prévio. Serão descritas algumas das principais técnicas [WEB02 *apud* OLI04]:

3.4.3.2.1 Misturas de especialistas

Este modelo de mistura adaptativa de especialistas locais é um processo de aprendizagem que treina vários classificadores (especialistas) e um combinador (função *gating*) formando um agrupamento paralelo de classificadores baseados no mesmo espaço de características. Cada especialista emite um vetor de saída, O_i ($i=1, \dots, L$), para um dado vetor de entrada x , e a rede *gating* realiza uma combinação linear dos vetores de saída. A função *gating* pode ser analisada como se atribuísse uma probabilidade para cada especialista, baseado na entrada corrente. A ênfase do processo de treinamento é encontrar a função *gating* ótima e, para cada função *gating*, treinar cada especialista para obter o máximo desempenho.

3.4.3.2.2 Regra baseada no formalismo bayesiano

Este método utiliza a regra do produto com estimativas de probabilidade *a posteriori* derivadas das predições de cada classificador individual, junto com um resumo do seu desempenho em um conjunto de treinamento previamente rotulado.

Especificamente, a regra de combinação bayesiana aproxima as probabilidades *a posteriori* por uma estimativa baseada em resultados de um processo de treinamento. A matriz de confusão C de cada classificador em um dado conjunto de treinamento é utilizada como indicação do seu desempenho. Para um problema com M classes, H é uma matriz de tamanho $M \times M$ em que cada elemento H_{ij} denota o número de padrões da classe i que são atribuídos à classe j pelo classificador. A partir da matriz H , obtém-se o número total de amostras pertencentes a classe i como a soma das linhas $\sum_{j=1}^M H_{ij}$, enquanto a soma das colunas $\sum_{i=1}^M H_{ij}$ representa o número total de amostras que são atribuídas à classe j pelo especialista. Quando há K especialistas, se têm K matrizes de confusão $H(k)$, $1 \leq k \leq K$. Conseqüentemente, a probabilidade condicional de um padrão x realmente pertença à classe i , dado que o especialista k indica que ele pertença à classe j .

3.4.3.3 Métodos de amostragem do conjunto de treinamento

Um dos principais problemas envolvendo combinação de classificadores é a existência de dependência entre os mesmo. Mesmos classificadores distintos podem ser correlacionados se forem treinado com os mesmos dados. Uma maneira efetiva de fazê-los discordar é treiná-los com conjuntos de treinamento distintos. Dois métodos principais serão apresentados [MAT04]:

- *Bagging*: contração de *bootstrap aggregating*, é um método de geração de conjuntos de treinamento para um dado número de classificadores previamente definidos. Consiste em gerar conjuntos de treinamento por um processo de amostragem aleatória com reposição. Os classificadores são treinados e aplica-se no combinador a regra do voto majoritário. Como a amostragem é feita com reposição, surgirão alguns elementos replicados nos novos conjuntos de treinamento. É importante observar que, se os classificadores que constituem o agrupamento forem instáveis, por exemplo, árvores de decisão, isto é, se respondem de forma bastante diferenciada quando treinados com dados ligeiramente distintos, então o processo de amostragem resultará em um conjunto de classificadores distintos, sendo portanto um procedimento válido. Portanto se, os classificadores forem estáveis, por exemplo, k-NN, então a combinação resultante teria pouco efeito, já que os classificadores tenderiam a apresentar a mesma predição;
- *Boosting*: este método também é aplicado a classificadores baseados no mesmo espaço de características. Assim com o anterior, também é um método para geração de conjuntos de treinamento e utiliza-se no combinador a regra do voto majoritário. Mas, diferentemente do *bagging*, os conjuntos de treinamento não são gerados simultaneamente. Neste caso, os conjuntos são gerados em série e para cada um deles atribui-se um peso usado no processo de combinação. A regra do voto majoritário pe regida com base nas ponderações determinadas no treinamento. O método foi criado para combinar classificadores que podem tratar entradas com pesos. Considera-se que os classificadores envolvidos sejam capazes de levar em consideração além do par ordenado padrão/rótulo (x_i, y_i) – um peso associado a ele – w_i . De um modo geral, a idéia consiste em, dado um conjunto de treinamento $\zeta = \{(x_i, y_i); i = 1, \dots, n\}$ e um conjunto de classificadores, $h_j(x)$, $j=1, \dots, m$, inicia um processo iterativo em que, a

cada iteração, sejam estabelecidos pesos w_i para os elementos do conjunto de treinamento, seja treinado um classificador $h_j(x)$ e seja estabelecido um peso associado ao classificador, denotado por e_j . Os pesos w_i são estabelecidos de maneira que seu somatório seja sempre igual a 1 em todas as iterações. Quando os classificadores não são capazes de treinar padrões com pesos, podemos eliminá-los fazendo um processo de amostragem com reposição, em que considera-se como peso a proporção ou a probabilidade de se sortear um padrão a partir do conjunto de treinamento original.

Os sistemas de múltiplos classificadores buscam obter resultados combinados de classificadores individuais de modo que a combinação apresente um desempenho global melhor do que àquele obtido pelo classificador individual.

Neste estudo utiliza-se redes neurais, de modo que considera-se apenas combinações de regras fixas, aplicando-se a regra da soma.

As combinações foram efetuadas entre 2 e 3 classificadores individuais. Observa-se nas combinações realizadas melhoras nas taxas reconhecimento.

3.5 Diversidade versus Discordância

3.5.1 Medidas de Diversidade e Discordância

Uma metodologia de projeto de sistemas de reconhecimento de padrões, com base em Sistema de Múltiplos Classificadores (MCS – *Multiple Classifiers System*) ou comitê/conjunto (*committee/ensemble*), dirige-se ao problema prático no projeto de sistemas de classificação, com exatidão e eficiência melhoradas [WIN05].

Tentativas de compreender a eficácia do padrão MCS tem alertado para o desenvolvimento de várias medidas, por exemplo: *Margin*, *Bias* e *Concepts of Variances*. Entretanto, recentemente diversas medidas da diversidade foram estudadas com a intenção de determinar a correlação dos dados em termos de exatidão global [KUN03].

Na realização deste objetivo, o principal questionamento é: como medir a eficiência do MCS? Nossa resposta utiliza similaridades e discordância entre os classificadores. Estas informações podem fornecer um mecanismo para entender como classificadores permitem melhorar a predição dos sistema de múltiplos classificadores ou comitês/conjunto. A medida da diversidade pode ser categorizada em dois tipos [KUN03]:

- *pair-wise*: esta abordagem calcula a média de uma distância particular medida entre todas as possibilidades dos pares de classificadores no conjunto. A distância calculada é usada para determinar as características da diversidade medida;
- *non-pair-wise*: Esta abordagem utiliza a idéia de entropia ou outra medida para calcular a correlação de cada membro do conjunto com a saída média do conjunto.

A dificuldade principal com medidas de diversidade é o dilema denominado de exatidão-diversidade. Como explicado em [HAD05], para alcançar os níveis mais altos de exatidão, a diversidade tem que diminuir de forma que espera-se um *tradeoff* entre diversidade e exatidão. Estes autores mencionaram que nenhuma teoria convincente ou estudo experimental houve para sugerir que qualquer medida pode prever com confiança o erro de generalização de um conjunto. E, sabe-se baseados em outros autores [WIN05] [ZOU04] que é necessário achar um ponto de equilíbrio entre diversidade e exatidão. Em outras palavras, estas são as medidas das discordâncias.

Duin et al. [DUI04] aplica o conceito de discordância para medir a diferença entre dois classificadores C_1 e C_2 treinados em um problema de classificação $P_j (j = 1, \dots, N)$, no qual N é o tamanho do conjunto de problemas. Entretanto, a discordância $d_j(C_1, C_2)$ pode ser formulada pela Equação 1:

$$d_j(C_1, C_2) = Prob(C_1(x) \neq C_2(x) \mid x \in P_j) \quad (3.17)$$

em que $C_i(x)$ retorna a rotulação de um padrão x de acordo com o classificador C_i . M classificadores constituem uma matriz $M \times M$ de discordâncias D para o problema P_j , com elementos $D_j^C(m, n) = d_j(C_m, C_n)$.

Duin et al. [DUI04], apresentam a disparidade entre as discordâncias de diversos classificadores quando projetados em um Espaço Euclidiano 2D, chamado de Espaço de Projeção de Classificadores (CPS), Figura 3.26.

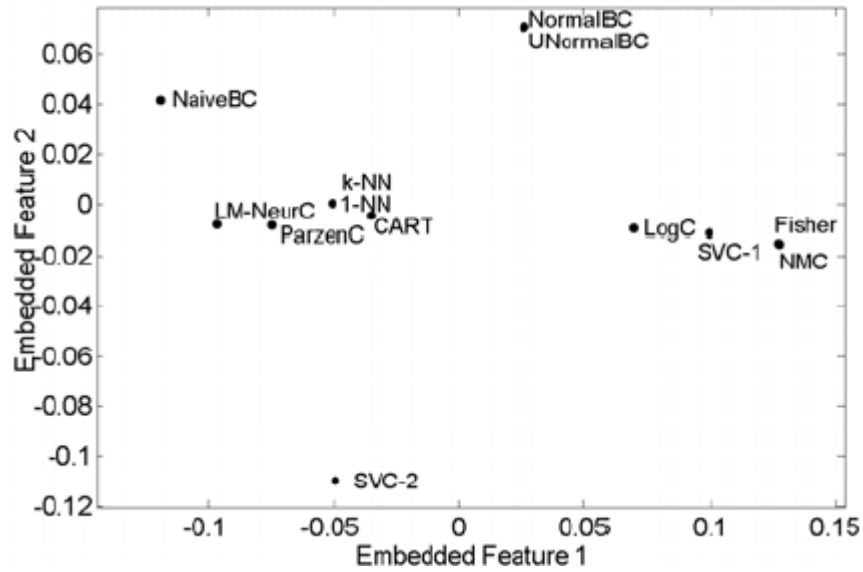


Figura 3.26: Representação de classificadores, sendo que suas distâncias ótimas preservam suas discordâncias. Resultado para o problema Highleyman com 10+10 objetos.

Fonte: Adaptado de [DUI04]

O interesse aqui é diferente de Duin et al. [DUI04], mas também é baseado em discordância. A idéia é utilizar a informação contida nas matrizes de confusão para cada classificador individual e computar as distâncias que representam as discordâncias entre os classificadores. Denomina-se esta abordagem de Discordância baseada no critério da Distância (*DD-based*).

3.5.2 Matrizes de Confusão

Uma análise consistente do comportamento do classificador pode ser fornecida pela matriz semi-global do desempenho, conhecida como Matriz de Confusão. Esta matriz é uma representação quantitativa do desempenho obtido para cada classificador em termos do reconhecimento de cada classe. A matriz de confusão pode ser representada pela Equação 3.18 [ZOU04]:

$$A = \begin{bmatrix} TR_{1,1} & TR_{1,2} & \cdots & TR_{1,N} \\ \vdots & \vdots & & \vdots \\ TR_{i,1} & TR_{i,2} & \cdots & TR_{i,N} \\ \vdots & \vdots & & \vdots \\ TR_{N,1} & TR_{N,2} & \cdots & TR_{N,N} \end{bmatrix} \quad (3.18)$$

onde, $TR_{i,j}$ corresponde ao total de números de entrada na classe C_i na qual a solução correta é colocada na posição j ; a diagonal principal indica o número total de exemplos para cada classe C_i reconhecido corretamente pelo sistema. Baseado na matriz A , é possível computar as medidas de desempenho globais do classificador como indicado pela Equação 3.19:

$$TR = \frac{1}{N} \sum_{i,j=1}^N TR_{i,j} \quad (3.19)$$

A abordagem *DD-based* utiliza a informação contida nas matrizes de confusão, para cada classificador individual é computada as distâncias que representam as discordâncias entre classificadores. A distância pode ser obtida considerando que todas as matrizes de confusão possuem o mesmo tamanho como definido na Equação 3.20:

$$D^{C_A, C_B} = \sum_{i=1}^N \sum_{j=1}^N |TR_{i,j}^A - TR_{i,j}^B| \quad (3.20)$$

Considerando C_A e C_B as matrizes de confusão, nas quais os elementos são utilizados para o cálculo das distâncias entre as matrizes. Cada matriz resultante retorna a distância final entre a combinação. As distâncias foram medidas considerando os classificadores individuais dois a dois, e, três a três, computando-se desta forma 10 medidas de distância, com as combinações: 4-5H, 4-5V, 4-7, 5H-5V, 5H-7, 5V-7, 4-5H-5V, 4-5H-7, 4-5V-7, 5H-5V-7.

3.5.3 Hipótese: *Soft-Correlation*

A hipótese é baseada na seguinte idéia proposta por Hadjitorov et. al. em [HAD05] : “*A seleção do conjunto através da diversidade mediana permite obter um valor melhor que a seleção randômica do conjunto ou seleção do conjunto com a discordância máxima*”. Estes autores observam que esses conjuntos mais diversos são menos exatos do que os conjuntos menos diversos. Então atribuí-se este fenômeno para intuir que mais diversidade sendo associada com muitos grupos não estrutura os grupos e, conseqüentemente, tem a exatidão individual mais baixa. Portanto, denomina-se esta hipótese de regra *soft-correlation*.

Analisando estas considerações, a idéia é computar as distâncias entre as matrizes de confusão e observar a mediana das distâncias com a intenção de definir meta-classes com base no conjunto de validação, para posteriormente aplica-las no conjunto de teste. Assim, define-se uma arquitetura baseada em múltiplos classificadores. O objetivo é maximizar as taxas de reconhecimento utilizando-se para isto as matrizes de confusão, suas discordâncias em termos de distâncias baseada na mediana, sem necessidade de combinar efetivamente os classificadores.

No trabalho apresentado em [FRE05] as informações da matrizes de confusão são utilizadas para computar as distâncias entre estas matrizes que representam a discordâncias entre os classificadores. O estudo utiliza estas informações para prover um mecanismo de conhecimento *a priori* das possíveis combinações de classificadores, sem precisar combiná-los, evitando buscas exaustivas para encontrar a melhor combinação.

3.5.4 Definição de Meta-Classes

Utilizando-se de uma definição de Linguagem de Programação, meta-classe é definida como: “*Uma meta-classe é uma classe de classes. Pode-se julgar conveniente que, em uma linguagem ou ambiente, classes também possam ser manipuladas como objetos. Por exemplo, uma classe pode conter variáveis com informações úteis, como, o número de objetos instanciados pela classe e valor médio de determinada propriedade*” [KAM96].

Para o presente estudo as meta-classes são conjunto de características comuns ou semelhantes a várias classes de objetos. A Figura 3.27 apresenta exemplos de meta-classes em trabalhos de reconhecimento de palavras manuscritas [FRE01] [OLI04]. Observa-se que no contexto de palavras manuscritas em cheques bancários, a meta-classe “*enta*” contém as

palavras que possuem o mesmo sufixo, no caso, “enta”. Assim, esta meta-classe pode ser modelada, ou seja, obter-se um modelo, por exemplo, um HMM capaz de reconhecer a meta-classe, inicialmente, para depois a partir de um modelo HMM para cada palavra realizar-se o reconhecimento das respectivas classes (ou palavras). Outro exemplo, é a meta classe “eiro”, também encontrada no contexto de cheques bancários brasileiros. Neste caso, esta meta-classe contém as palavras: “janeiro” e “fevereiro”. Tendo-se aqui o mesmo objetivo, ou seja, classificar inicialmente a meta-classe “eiro” entre as demais meta-classes, por exemplo, “embro” que contém as palavras: “setembro”, “novembro” e “dezembro”.

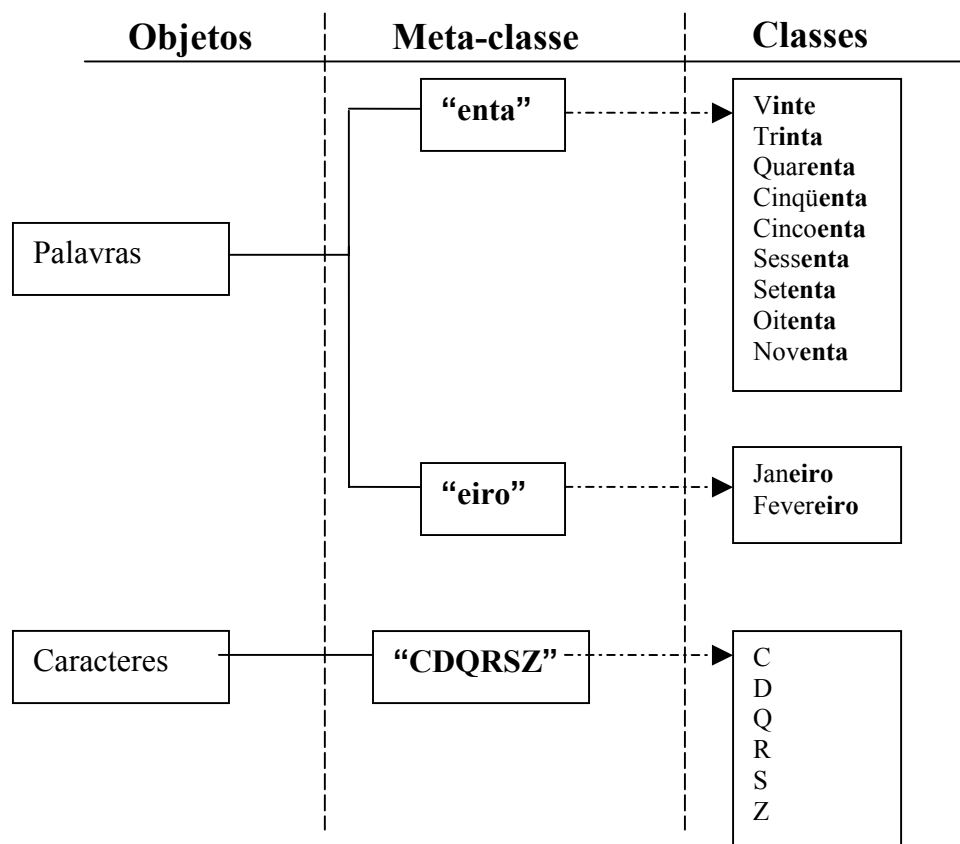


Figura 3.27: Exemplo de meta-classes

Observando-se que nos exemplos de palavras manuscritas a similaridade encontra-se no sufixo das palavras, no contexto do presente trabalho busca-se verificar a similaridade nas regiões perceptivas. O exemplo de meta-classe “CDQRSZ” definida com base no cálculo das distâncias (*DD-based*) para os caracteres apresenta sua similaridade na concavidade, visto que na Figura 3.28 as classes apresentam estas similaridades evidenciadas. A Figura 3.29

apresenta a meta-classe AB, observa-se a similaridade no traço horizontal e na concavidade superior.

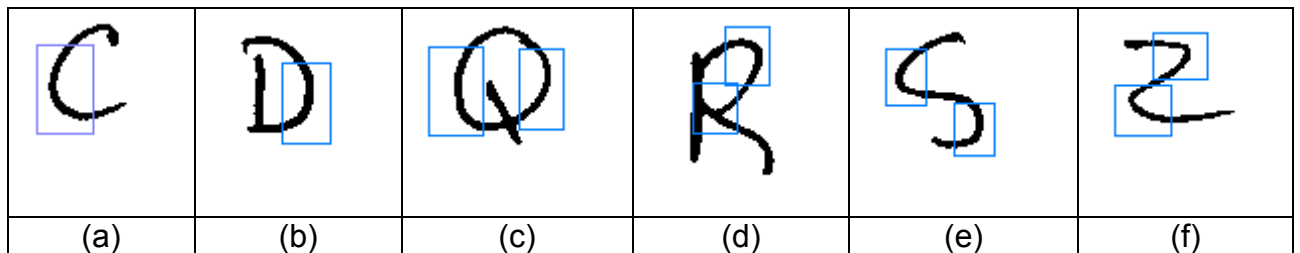


Figura 3.28: Meta-classe CDQRSZ

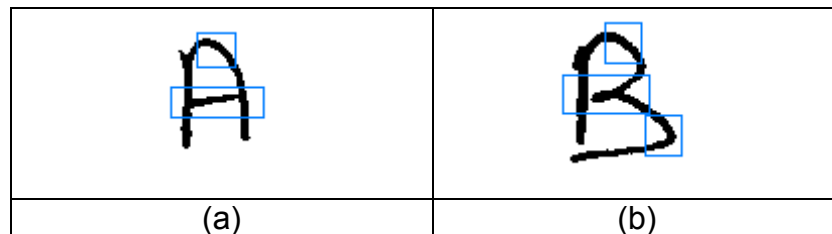


Figura 3.29: Meta-classe AB

A definição das meta-classes passa pela determinação dos conjuntos de classificadores que estão próximos em termos de distância, representadas pelas medianas e, assim, verificar quais classes de caracteres possuem o mesmo comportamento, ou seja, quais classes de caractere possuem suas medianas no mesmo classificador.

Observa-se na Tabela 3.4 as combinações dos classificadores, bem como, as distâncias resultantes em cada combinação. Para cada combinação dois a dois e três a três define-se um valor de mediana. As classes “A” e “B” possuem suas medianas nos classificadores 4-5H e 4-5V, podendo significar um caminho mais rápido para a melhor taxa de reconhecimento, partindo da combinação destes classificadores. Essas verificações foram realizadas para todas as classes, permitindo definir as meta-classes em função dos classificadores apontados pelas medianas, esses valores foram definidos em função do conjunto de validação.

As meta-classes resultantes desta análise foram: AB, CDQRSZ, EJM, FPT, GX, HKLNO, IY, UVW. Estas meta-classes foram validadas com o conjunto de teste, conforme apresentado no Capítulo 5.

3.6 Considerações Finais

Neste capítulo, foi descrito o sistema proposto para o reconhecimento de caracteres utilizando regiões perceptivas. Este é composto por: pré-processamento, extração de primitivas, classificação por múltiplos classificadores que busca uma abordagem alternativa a taxa de reconhecimento, sendo esta abordagem definida como uma Discordância baseada na medida da Distância (*DD-based*), aplica-se uma distância euclidiana computada entre as matrizes de confusão dos classificadores e, uma regra de *soft-correlation* é proposta para indicar o melhor conjunto de classificadores. No próximo capítulo são apresentados os experimentos realizados e os resultados obtidos para validar a metodologia proposta.

Tabela 3.4: Distâncias com base nas matrizes de confusão para as classes:

a)caractere “A”, b)caractere “B”

A	
4-7	0,089552
5v-7	0,119403
4-5h	0,149254
4-5v	0,149254
5h-5v	0,179104
5h-7	0,179104
Mediana	0,149254
4-5h-7	1,059701
4-5h-5v	1,059701
4-5v-7	1,059701
5h-5v-7	1,149254
Mediana	1,059701

(a)

B	
5v-7	0,149254
4-7	0,238806
4-5h	0,298507
4-5v	0,328358
5h-7	0,358209
5h-5v	0,41791
Mediana	0,313433
4-5h-5v	1,059701
4-5h-7	1,059701
4-5v-7	1,089552
5h-5v-7	1,149254
Mediana	1,074627

(b)

Capítulo 4

Experimentos Realizados

Neste capítulo são apresentados os experimentos realizados e os resultados obtidos com o objetivo de investigar a eficiência da metodologia proposta. Como descrito na Seção 1.1, o problema abordado na presente trabalho é o reconhecimento de caracteres baseado em regiões perceptivas. A metodologia proposta para este problema foi apresentada no Capítulo 3.

4.1 Base de Dados – IRONOFF

Para a realização dos experimentos utilizou-se a base de dados IRONOFF, gerada pela IRESTE (*University of Nantes - France*) 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* (Figura 4.1). Para este estudo foram utilizados arquivos da base *off-line*. Esta base de dados está dividida em um léxico de 26 caracteres conforme apresentado na Tabela 4.1. Os conjuntos de treinamento possuem 200 exemplares da classe em questão (por exemplo, da classe A) e 8 exemplares das demais classes (no caso, as 25 classes restantes). Para a validação, um esquema semelhante foi adotado considerando-se 67 exemplares da classe e 3 para cada uma das demais classes. Esta organização é necessária devido ao treinamento ser realizado com base em uma arquitetura Classe-Modular, ou seja, uma rede neural para cada classe de caractere [OH_02].

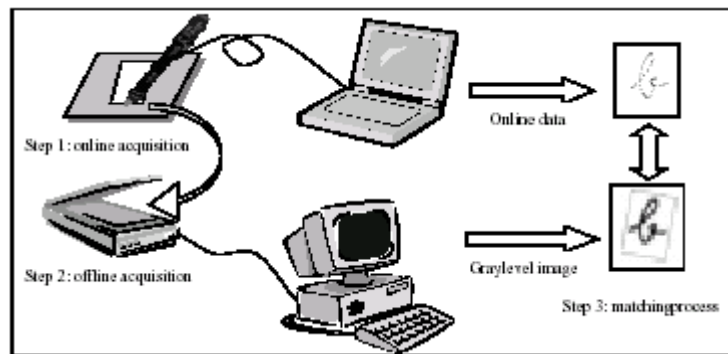


Figura 4.1 Visão geral da construção da base IRONOFF. Fonte: Adaptado de [VIA01]

Tabela 4.1: Apresenta a distribuição de caracteres – Base IRONOFF

Léxico	Quantidade	Treinamento 200 classe + 8 por classe	Validação 67 classe + 3 por classe	Teste 67 classe
A	403	400	142	67
B	405	400	142	67
C	403	400	142	67
D	404	400	142	67
E	406	400	142	67
F	406	400	142	67
G	406	400	142	67
H	405	400	142	67
I	405	400	142	67
J	404	400	142	67
K	404	400	142	67
L	405	400	142	67
M	406	400	142	67
N	405	400	142	67
O	398	400	142	67
P	405	400	142	67
Q	403	400	142	67
R	404	400	142	67
S	403	400	142	67
T	405	400	142	67
U	405	400	142	67
V	404	400	142	67
W	403	400	142	67
X	405	400	142	67
Y	405	400	142	67
Z	403	400	142	67
Total	10.510	10.400	3.692	1.742

Analisando a matriz de confusão para $Z = 4$, Tabela 4.2, as seguintes confusões entre classes são evidentes: “B”, “D” e “O”; “C” e “E”; “D” e “O”; “H” e “M”; “I” e “F”; “G” e “Q”; “J” e “D”; “K” e “M”; “N” e “W”; “R” e “A”; “S” e “D”; “W”, “U” e “V”; “X” e “K”; “Y” e “X”. Assim, foram realizados os experimentos com 5 regiões, para $Z = 5$ -Horizontal (Tabela 4.3) e 5-Vertical (Tabela 4.4). A idéia é constatar uma melhor solução para os problemas de confusão entre as formas não simétricas, tais como: “G” e “Q” (Figura 4.2a); “D” e “O”; “Y” e “X”.

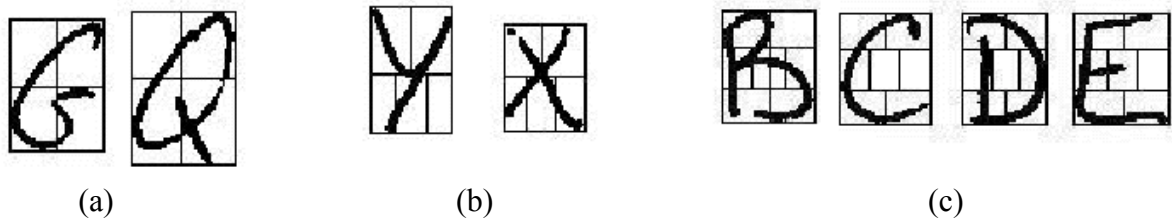


Figura 4.2: Zoneamento baseado em regiões perceptivas

Tabela 4.3: Matriz de Confusão para conjunto de Teste, zoneamento $Z=5H$

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	NI	%Rec.	
A	58	1	0	2	0	0	0	0	0	0	0	0	1	0	0	0	0	3	0	1	0	0	0	0	1	0	0	86,57	
B	0	43	0	10	0	0	1	0	0	0	0	0	1	0	9	0	0	2	1	0	0	0	0	0	0	0	0	64,18	
C	1	0	53	0	7	0	4	0	0	0	0	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	79,10	
D	0	5	0	44	0	0	0	1	0	0	0	0	0	0	14	0	0	0	0	0	1	0	0	0	0	0	2	65,67	
E	0	0	7	0	57	0	1	0	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	85,07	
F	0	0	0	1	2	61	0	0	1	1	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	91,04	
G	1	0	0	2	2	0	58	0	0	0	0	0	0	0	2	0	1	0	1	0	0	0	0	0	0	0	0	86,57	
H	0	0	0	0	0	0	0	57	0	0	0	0	4	3	0	0	0	0	0	0	2	0	1	0	0	0	0	85,07	
I	0	1	0	0	10	0	0	0	48	7	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	71,64	
J	0	0	0	5	1	0	2	0	3	53	0	0	0	0	1	0	0	0	0	0	2	0	0	0	0	0	0	79,10	
K	0	0	0	0	4	0	0	1	1	0	51	0	1	1	0	0	0	2	0	0	3	0	0	3	0	0	0	76,12	
L	0	1	1	1	1	0	1	0	0	0	0	60	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0	89,55	
M	1	1	0	0	0	0	0	6	0	0	0	0	55	0	0	0	1	2	0	0	0	0	1	0	0	0	0	82,09	
N	0	1	0	0	0	0	0	2	0	1	0	0	3	52	0	0	0	0	0	0	1	2	4	1	0	0	0	77,61	
O	0	2	1	0	0	0	0	2	0	0	0	0	0	0	60	0	0	0	0	0	0	1	1	0	0	0	0	89,55	
P	0	0	0	0	4	0	0	0	0	0	0	1	0	0	0	62	0	0	0	0	0	0	0	0	0	0	0	92,54	
Q	1	2	0	1	0	0	2	0	0	0	0	0	0	0	8	9	43	0	1	0	0	0	0	0	0	0	0	64,18	
R	4	0	0	0	0	0	0	0	0	0	0	0	1	0	0	1	0	60	0	1	0	0	0	0	0	0	0	89,55	
S	0	0	0	4	2	0	1	0	1	3	0	0	0	0	0	0	0	0	53	0	0	0	0	0	2	1	0	79,10	
T	0	0	0	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	65	0	0	0	0	0	0	0	97,01	
U	0	0	0	0	0	0	0	1	1	0	0	0	1	2	2	0	0	0	0	0	57	3	0	0	0	0	0	85,07	
V	0	0	0	0	0	0	0	1	0	0	0	0	1	0	0	0	0	0	0	0	2	55	4	0	4	0	0	82,09	
W	0	0	0	0	0	0	0	3	0	0	0	0	0	4	0	0	0	0	0	0	8	2	50	0	0	0	0	74,63	
X	1	0	0	0	2	0	0	2	0	0	7	0	1	0	0	0	0	2	0	0	0	0	1	50	1	0	0	74,63	
Y	0	0	0	1	0	0	0	2	0	2	0	0	0	1	0	0	0	0	0	1	0	0	0	0	60	0	0	89,55	
Z	0	0	0	2	0	0	1	0	1	2	0	0	0	0	0	1	0	0	0	1	0	0	0	0	0	59	0	88,06	
																												Média	81,75

Observa-se que a matriz de confusão para $Z = 5H$ (Tabela 4.3) apresenta melhores resultados para as seguintes confusões: “G”, “O”, e “Y”. Na verdade, este mecanismo de zoneamento contribui para solucionar problemas em letras que não são simétricas horizontalmente (Figura 4.2b).

Observa-se na Tabela 4.4 que o zoneamento considerando 5 regiões verticalmente não contribui para o aumento da taxa de reconhecimento do classificador individual e, conseqüentemente, não auxilia na solução das confusões identificadas.

Tabela 4.4: Matriz de Confusão para conjunto de Teste, zoneamento $Z=5V$

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	NI	%Rec.	
A	60	0	0	2	0	0	1	0	0	0	0	0	2	0	0	0	1	1	0	0	0	0	0	0	0	0	0	89,55	
B	0	50	0	8	0	0	1	0	0	0	0	0	1	0	7	0	0	0	0	0	0	0	0	0	0	0	0	74,63	
C	0	0	46	0	16	0	2	0	0	0	0	1	0	0	0	0	1	0	0	1	0	0	0	0	0	0	0	68,66	
D	0	5	0	46	1	0	0	0	0	0	0	0	0	0	13	1	0	0	0	0	0	0	0	0	0	0	1	68,66	
E	0	1	4	0	60	0	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	89,55	
F	0	1	0	0	2	60	0	0	1	2	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	89,55	
G	0	2	0	1	5	1	54	0	0	0	0	0	1	0	2	0	1	0	0	0	0	0	0	0	0	0	0	80,60	
H	1	0	0	0	0	0	0	47	0	0	1	0	16	1	0	0	0	0	0	0	0	0	0	0	0	0	1	70,15	
I	0	1	0	1	8	1	0	0	51	2	0	0	0	0	0	0	0	0	0	1	0	0	0	1	0	1	0	76,12	
J	0	1	0	5	0	0	1	0	0	53	0	0	1	0	0	0	2	0	0	2	0	0	0	0	1	1	0	79,10	
K	0	0	0	0	5	0	0	2	0	0	52	0	3	1	0	0	0	0	0	0	1	0	0	3	0	0	0	77,61	
L	0	0	3	1	3	0	1	0	0	0	0	58	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	86,57	
M	1	0	0	0	0	1	0	3	1	0	0	0	57	1	0	0	0	2	0	0	0	0	1	0	0	0	0	85,07	
N	2	0	0	0	0	0	0	1	0	0	0	0	4	47	0	0	0	0	0	0	2	2	8	1	0	0	0	70,15	
O	0	2	0	1	2	0	0	0	1	0	0	0	0	0	59	0	0	0	0	0	0	0	1	0	0	1	0	88,06	
P	0	0	0	2	3	0	0	0	0	0	0	0	0	0	0	61	0	0	0	0	0	0	0	0	0	1	0	91,04	
Q	0	2	0	1	1	0	1	0	0	0	0	0	0	0	5	4	51	1	0	0	0	0	1	0	0	0	0	76,12	
R	1	2	0	0	1	0	0	0	0	0	0	0	1	0	0	2	0	59	0	0	0	0	0	1	0	0	0	88,06	
S	0	0	1	4	0	0	1	0	2	3	0	0	0	0	1	0	0	0	53	0	0	0	0	0	1	1	0	79,10	
T	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	65	0	0	0	0	0	1	0	97,01	
U	0	1	0	0	1	0	1	3	0	0	0	0	0	2	1	0	0	0	0	1	55	2	0	0	0	0	0	82,09	
V	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	5	59	1	0	1	0	0	88,06	
W	0	0	0	0	1	0	0	1	0	0	0	0	2	5	0	0	0	0	0	0	0	7	7	44	0	0	0	65,67	
X	0	0	0	0	3	0	0	0	2	0	7	0	2	0	0	0	0	2	0	0	0	0	1	47	3	0	0	70,15	
Y	0	0	0	3	0	0	0	1	0	1	1	0	1	0	0	0	0	0	0	1	0	2	0	0	57	0	0	85,07	
Z	0	1	1	1	1	0	1	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	59	0	88,06	
																												Média	80,94

Deste modo, realizou-se um experimento com $Z = 7$ regiões (Tabela 4.5). A idéia é, além de alcançar um melhor resultado para letras não simétricas, extrair e representar a parte central dos caracteres diferentemente das partes superior e inferior, tal qual nos seguinte caracteres: “D” e “C”, “N” e “W”; “Y” e “X”. Este zoneamento resultou melhor para os seguintes caracteres: “B”, “C”, “D”, “E”, “K”, “N”, “P”, “R”, “U”, “W” e “X” (Figura 4.2c). Assim, a taxa de reconhecimento alcançada com $Z = 7$ regiões foi a mais alta (84,73%).

Tabela 4.5: Matriz de Confusão para conjunto de Teste, zoneamento Z=7

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	NI	%Rec.		
A	61	2	0	0	0	0	0	0	0	0	0	0	2	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	91,04	
B	0	53	0	10	0	0	1	0	0	0	0	0	1	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	79,10	
C	0	0	59	0	3	0	2	0	0	0	0	1	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	88,06	
D	0	4	0	55	0	0	0	1	0	0	0	0	0	0	5	1	0	0	0	0	1	0	0	0	0	0	0	0	82,09	
E	0	0	0	0	64	1	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	95,52	
F	0	0	0	0	3	62	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	92,54	
G	1	1	3	1	4	0	54	0	0	0	0	0	0	0	2	0	0	0	0	1	0	0	0	0	0	0	0	0	80,60	
H	0	0	0	0	0	0	0	51	0	0	0	0	12	2	0	0	0	0	0	0	2	0	0	0	0	0	0	0	76,12	
I	0	0	0	1	4	0	0	0	48	5	0	0	0	1	0	0	0	2	1	0	0	0	0	1	1	3	0	71,64		
J	0	0	0	6	0	0	2	0	0	55	0	0	0	1	0	0	1	0	0	1	0	0	0	0	1	0	0	82,09		
K	0	0	1	0	1	0	0	0	1	0	54	1	3	0	0	0	0	0	0	0	1	0	0	5	0	0	0	80,60		
L	0	0	1	0	2	0	1	0	0	0	1	61	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	91,04	
M	1	0	0	0	0	0	0	3	0	0	0	0	59	1	1	0	1	1	0	0	0	0	0	0	0	0	0	0	88,06	
N	0	0	0	0	0	0	0	1	0	1	0	0	5	58	0	0	0	0	0	0	0	0	1	1	0	0	0	86,57		
O	0	2	0	2	0	0	1	1	0	0	0	0	0	0	56	0	2	1	0	1	0	0	0	0	1	0	0	83,58		
P	0	1	0	0	1	0	0	1	0	0	0	0	0	0	0	63	0	0	0	1	0	0	0	0	0	0	0	0	94,03	
Q	1	2	1	1	0	0	1	0	0	0	0	0	0	0	3	3	54	1	0	0	0	0	0	0	0	0	0	80,60		
R	1	2	0	0	0	0	1	0	0	0	0	0	1	0	0	0	1	61	0	0	0	0	0	0	0	0	0	0	91,04	
S	0	2	1	3	1	0	1	0	1	6	0	0	0	0	0	0	0	0	51	0	0	0	0	0	1	0	0	76,12		
T	0	0	0	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	65	0	0	0	0	0	0	0	97,01	
U	0	0	0	0	0	0	0	1	0	0	0	0	1	1	1	0	0	0	0	0	2	58	2	0	0	1	0	86,57		
V	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	55	6	0	4	0	0	82,09		
W	0	0	0	0	0	0	0	1	0	0	0	0	0	2	0	0	0	0	0	0	0	5	5	53	0	1	0	79,10		
X	1	0	0	0	0	0	0	0	0	0	5	0	2	0	0	0	1	0	0	0	0	0	1	53	3	1	0	79,10		
Y	0	0	0	0	0	0	1	0	0	3	1	0	1	1	0	0	0	0	0	2	0	2	1	0	55	0	0	82,09		
Z	0	1	1	0	0	0	1	2	0	1	0	1	0	0	0	0	0	0	0	1	0	0	0	1	0	58	0	86,57		
																													Média	84,73

A Figura 4.3 apresenta um gráfico explicativo das taxas de reconhecimento alcançadas pelos classificadores individuais para os conjuntos de validação e teste, respectivamente.

As Tabelas 4.6 e 4.7 apresentam os resultados obtidos com os conjuntos de validação e teste, respectivamente, para os classificadores individuais e combinados (dois a dois e três a três). Observa-se na Tabela 4.6 que as taxas de reconhecimento melhoraram com a combinação de dois ou três classificadores, somente as classes H, M, N, Y, conseguiram taxas de reconhecimento em classificadores individuais maiores que as taxas dos classificadores combinados. As classes F, J, O, T, Z, conseguira taxas individuais iguais as melhores taxas dos classificadores combinados. As combinações de classificadores foram eficientes para a melhoria nas taxas de reconhecimento das classes A, B, C, D, E, G, I, K, L, P, Q, R, S, U, V, W, X, correspondentes a 65,3% das classes do problema. Considerando as combinações que conseguiram taxas iguais aos classificadores individuais temos 84,6% das classes com as melhores taxas de reconhecimento.

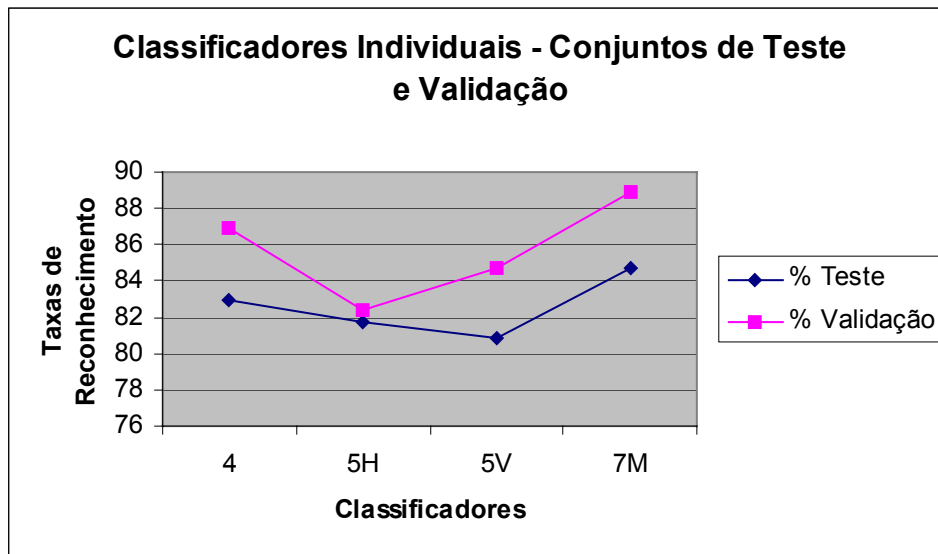


Figura 4.3: Taxas de Reconhecimento dos conjuntos de Validação e Teste para Classificadores Individuais

Observa-se na Tabela 4.6 que o melhor resultado para os classificadores combinados é encontrado com a combinação dos classificadores **4-5H-7**, atingindo 91,0% de taxa de reconhecimento. Este resultado vem de encontro dos resultados obtidos com as Tabelas 4.2 até 4.5, nas quais se observa que o classificador individual **5V** não contribui para a melhoria do reconhecimento das formas, nem mesmo, auxiliando com as confusões entre classes. Para a combinação composta por 2 classificadores observa-se que o melhor resultado encontra-se com a composição baseada em **4-7**, obtendo-se 90,3% de taxa de reconhecimento. Assim, observa-se que o classificador **5H** complementa os classificadores **4** e **7**.

Para a Tabela 4.7, conjunto de teste, os resultados obtidos demonstram que a combinação dos classificadores **5H-5V-7** resulta na melhor taxa de reconhecimento, ou seja, 85,9%. Em termos de combinação de 2 classificadores, observa-se que os classificadores **4-7** geram o melhor resultado, ou seja, 85,8%. Esta taxa é próxima da taxa alcançada com 3 classificadores, não demonstrando um ganho significativo em termos de taxa quando comparado ao esforço computacional para efetivação da combinação. Porém, observa-se que esta combinação também é satisfatória para o conjunto de teste, tal qual, para o conjunto de validação.

Tabela 4.6: Taxas de Reconhecimento para conjunto de Validação - Classificadores individuais e combinados

	4	5H	5V	7	4-5H	4-5V	4-7	5H-5V	5H-7	5V-7	4-5H-5V	4-5H-7	4-5V-7	5H-5V-7
A	86,6	85,1	89,6	89,6	91,0	91,0	86,6	92,5	89,6	91,0	92,5	89,6	91,0	92,5
B	76,1	68,7	86,6	83,6	82,1	83,6	82,1	85,1	79,1	83,6	83,6	80,6	83,6	83,6
C	89,6	88,1	67,2	97,0	92,5	89,6	95,5	91,0	95,5	95,5	91,0	95,5	95,5	97,0
D	92,5	82,1	89,6	94,0	89,6	91,0	94,0	88,1	91,0	92,5	91,0	94,0	95,5	94,0
E	85,1	77,6	85,1	85,1	85,1	82,1	82,1	82,1	86,6	82,1	83,6	83,6	82,1	82,1
F	91,0	82,1	89,6	91,0	92,5	91,0	92,5	92,5	91,0	91,0	92,5	92,5	91,0	91,0
G	92,5	83,6	80,6	91,2	91,0	89,6	94,0	85,1	92,5	88,1	91,0	95,5	92,5	91,0
H	76,1	82,1	69,7	73,1	82,1	76,1	76,1	85,1	79,1	79,1	85,1	82,1	79,1	82,1
I	77,6	64,2	77,6	77,6	77,6	74,6	77,6	77,6	76,1	79,1	76,1	79,1	79,1	79,1
J	80,6	74,6	79,1	83,6	85,1	82,1	89,6	86,6	88,1	89,6	86,6	88,1	89,6	88,1
K	86,6	80,6	83,6	89,6	89,6	91,0	92,5	86,6	91,0	92,5	89,6	94,0	95,5	92,5
L	97,0	94,0	92,5	97,0	97,0	95,5	97,0	97,0	97,0	97,0	97,0	97,0	97,0	97,0
M	82,1	86,6	91,0	89,6	86,6	88,1	85,1	83,6	85,1	86,6	85,1	85,1	86,6	83,6
N	76,1	83,6	76,1	76,1	80,6	74,6	79,1	79,1	82,1	74,6	77,6	79,1	76,1	79,1
O	97,0	97,0	98,5	98,5	98,5	98,5	100,0	97,0	100,0	100,0	98,5	100,0	100,0	100,0
P	92,5	97,0	91,0	94,0	94,0	92,5	92,5	95,5	95,5	92,5	94,0	94,0	92,5	95,5
Q	83,6	61,2	70,1	88,1	82,1	82,1	88,1	68,7	85,1	82,1	82,1	86,6	86,6	80,6
R	82,1	85,1	92,5	98,5	85,1	85,1	95,5	91,0	98,5	97,0	89,6	97,0	95,5	95,5
S	92,5	91,0	85,1	85,1	92,5	92,5	91,0	91,0	89,6	91,0	94,0	91,0	94,0	89,6
T	98,5	97,0	98,5	98,5	100,0	100,0	100,0	100,0	100,0	100,0	100,0	100,0	100,0	100,0
U	85,1	89,6	85,1	94,0	89,6	83,6	86,6	91,0	92,5	89,6	89,6	89,6	88,1	91,0
V	97,0	91,0	98,5	94,0	100,0	98,5	100,0	95,5	98,5	95,5	98,5	100,0	98,5	97,0
W	83,6	68,7	67,2	80,6	88,1	91,0	95,5	73,1	82,1	79,1	88,1	94,0	92,5	80,6
X	83,6	79,1	83,6	82,1	79,1	83,6	86,6	85,1	88,1	85,1	80,6	91,0	86,6	88,1
Y	85,1	85,1	95,5	91,0	97,0	98,5	97,0	94,0	94,0	91,0	100,0	97,0	98,5	95,5
Z	89,6	68,7	77,6	88,1	89,6	94,0	91,0	92,5	91,0	91,0	94,0	91,0	95,5	95,5
	86,9	82,4	84,7	88,9	89,2	88,5	90,3	87,9	90,0	89,1	89,7	91,0	90,9	90,1

Outra análise é referente ao resultado obtido com os classificadores **4-5H-7**, no conjunto de validação, a fim de entender o que ocorre no conjunto de teste. Observa-se na Tabela 4.7 que esta combinação atinge uma taxa de 85,5% de reconhecimento em comparação com o maior valor alcançado de 85,9% (**5H-5V-7**).

Deste modo, observa-se que a taxa de reconhecimento é um parâmetro insuficiente para determinar os melhores conjuntos de classificadores a combinar. Principalmente devido ao fato de que para se conhecer a taxa de reconhecimento é necessário combinar efetivamente os classificadores (através das probabilidades *a posteriori*), o que dependendo do número de classificadores a combinar se torna uma tarefa inviável. Nestes casos, os métodos de *Bagging* e *Boosting*, ou ainda, de programação dinâmica; devem ser aplicados (ver Seção 3.4.3). Assim, a metodologia apresentada no presente trabalho possibilita determinar a priori os conjuntos de classificadores a serem analisados.

Tabela 4.7: Matriz Geral que apresenta as Taxas de Reconhecimento para conjunto de Teste
Classificadores individuais e combinados

	4	5H	5V	7	4-5H	4-5V	4-7	5H-5V	5H-7	5V-7	4-5H-5V	4-5H-7	4-5V-7	5H-5V-7
A	92,5	86,6	89,6	91,0	88,1	94,0	94,0	89,6	89,6	91,0	91,0	92,5	95,5	92,5
B	65,7	64,2	74,6	79,1	65,7	79,1	76,1	73,1	74,6	82,1	76,1	76,1	79,1	82,1
C	82,1	79,1	68,7	88,1	82,1	83,6	83,6	83,6	86,6	86,6	83,6	83,6	83,6	85,1
D	73,1	65,7	68,7	82,1	70,1	73,1	77,6	68,7	74,6	74,6	68,7	74,6	73,1	73,1
E	83,6	85,1	89,6	95,5	83,6	85,1	89,6	86,6	95,5	89,6	83,6	89,6	86,6	89,6
F	92,5	91,0	89,6	92,5	91,0	91,0	92,5	89,6	92,5	92,5	91,0	92,5	92,5	92,5
G	82,1	86,6	80,6	80,6	88,1	83,6	89,6	85,1	86,6	85,1	85,1	88,1	86,6	86,6
H	88,1	85,1	70,1	76,1	86,6	83,6	85,1	85,1	85,1	77,6	85,1	85,1	82,1	85,1
I	76,1	71,6	76,1	71,6	74,6	77,6	74,6	77,6	68,7	77,6	77,6	71,6	77,6	76,1
J	83,6	79,1	79,1	82,1	83,6	80,6	83,6	79,1	82,1	80,6	80,6	80,6	83,6	82,1
K	77,6	76,1	77,6	80,6	80,6	80,6	82,1	77,6	79,1	82,1	83,6	80,6	83,6	80,6
L	92,5	89,6	86,6	91,0	91,0	89,6	91,0	92,5	92,5	94,0	91,0	92,5	92,5	94,0
M	92,5	82,1	85,1	88,1	85,1	85,1	89,6	83,6	86,6	88,1	86,6	86,6	86,6	88,1
N	68,7	77,6	70,1	86,6	76,1	73,1	82,1	79,1	82,1	83,6	77,6	83,6	80,6	83,6
O	86,6	89,6	88,1	83,6	89,6	88,1	83,6	88,1	86,6	85,1	88,1	86,6	85,1	86,6
P	86,6	92,5	91,0	94,0	92,5	88,1	94,0	91,0	95,5	94,0	94,0	94,0	94,0	95,5
Q	82,1	64,2	76,1	80,6	80,6	85,1	85,1	79,1	80,6	82,1	83,6	85,1	86,6	82,1
R	86,6	89,6	88,1	91,0	89,6	89,6	91,0	94,0	95,5	92,5	94,0	94,0	94,0	95,5
S	79,1	79,1	79,1	76,1	77,6	79,1	76,1	82,1	79,1	80,6	79,1	77,6	79,1	82,1
T	95,5	97,0	97,0	97,0	95,5	95,5	95,5	97,0	97,0	97,0	97,0	97,0	97,0	97,0
U	80,6	85,1	82,1	86,6	82,1	83,6	85,1	85,1	86,6	86,6	83,6	85,1	85,1	88,1
V	95,5	82,1	88,1	82,1	94,0	92,5	97,0	89,6	82,1	89,6	92,5	95,5	92,5	89,6
W	70,1	74,6	65,7	79,1	70,1	67,2	79,1	74,6	82,1	77,6	70,1	77,6	77,6	80,6
X	76,1	74,6	70,1	79,1	80,6	76,1	83,6	73,1	77,6	73,1	76,1	80,6	77,6	73,1
Y	77,6	89,6	85,1	82,1	82,1	82,1	82,1	88,1	85,1	85,1	83,6	83,6	83,6	85,1
Z	89,6	88,1	88,1	86,6	88,1	89,6	88,1	89,6	89,6	88,1	89,6	88,1	89,6	88,1
	83,0	81,7	80,9	84,7	83,4	83,7	85,8	83,9	85,1	85,2	84,3	85,5	85,6	85,9

Para tal, as Tabelas 4.8 e 4.9 apresentam os resultados obtidos com o cálculo das distâncias entre as matrizes de confusão dos classificadores individuais para as combinações analisadas considerando os conjuntos de validação e teste, respectivamente.

As Figuras 4.4 a 4.11, apresentam as distâncias dos classificadores projetado em um Espaço Euclidiano 2D, chamado por [DUI04], como Espaço de Projeção de Classificadores (CPS), apresentado na Seção 3.5.1. As Figuras apresentam gráficos que foram projetados conforme as distâncias e seus classificadores em cada meta-classe.

Observa-se nos gráficos que as distâncias entre os classificadores dois a dois e três a três são bem representadas e distintas, verificando que quanto maior a quantidade de classificadores combinados, maior será a distância dos conjuntos. Para uma melhor visualização foram gerados gráficos por meta-classe, porém, quando as distâncias são

projetadas em um mesmo espaço pode-se observar as similaridades presentes entre as classes constituintes de cada meta-classe. Isto significa que as classes possuem um mesmo comportamento, inclusive no que se refere à combinação dos zoneamentos buscando uma complementaridade no reconhecimento do caractere. Para algumas combinações de classificadores, as distâncias entre as classes de uma mesma meta-classe são mínimas, sendo que a representação muitas vezes apresenta sobreposição de valores.

Tabela 4.8: Matriz Geral que apresenta as distâncias entre as matrizes de confusão para conjunto de Validação - Classificadores individuais e combinados

	4-5H	4-5V	4-7	5H-5V	5H-7	5V-7	4-5H-5V	4-5H-7	4-5V-7	5H-5V-7
A	0,149254	0,149254	0,089552	0,179104	0,179104	0,119403	1,059701	1,059701	1,059701	1,149254
B	0,298507	0,328358	0,238806	0,41791	0,358209	0,149254	1,059701	1,059701	1,089552	1,149254
C	0,029851	0,447761	0,149254	0,447761	0,179104	0,597015	1	1	1	1,029851
D	0,298507	0,179104	0,119403	0,179104	0,238806	0,089552	1,089552	1,089552	1,119403	1,149254
E	0,179104	0,089552	0,179104	0,268657	0,358209	0,119403	1	1	1,029851	1,208955
F	0,298507	0,089552	0,119403	0,298507	0,298507	0,119403	1,029851	1,059701	1,029851	1,268657
G	0,238806	0,298507	0,088235	0,208955	0,18086	0,240562	1	1,029851	1,000439	1,119842
H	0,179104	0,218001	0,119403	0,336952	0,238806	0,247852	1,059701	1,029851	1,029851	1,029398
I	0,447761	0,179104	0,268657	0,507463	0,447761	0,238806	1,059701	1,149254	1,059701	1,358209
J	0,268657	0,179104	0,208955	0,328358	0,38806	0,119403	1,029851	1,059701	1,059701	1,268657
K	0,328358	0,238806	0,179104	0,208955	0,328358	0,268657	1,179104	1,149254	1,119403	1,208955
L	0,089552	0,119403	0,029851	0,149254	0,089552	0,119403	1	1,029851	1	1,089552
M	0,179104	0,208955	0,179104	0,208955	0,059701	0,149254	1,059701	1,119403	1,059701	1,059701
N	0,298507	0,119403	0,119403	0,268657	0,208955	0,119403	1,089552	1,089552	1	1,059701
O	0,059701	0,059701	0,059701	0,059701	0,059701	0,029851	1,059701	1,059701	1,059701	1,059701
P	0,119403	0,059701	0,089552	0,149254	0,119403	0,119403	1,029851	1,059701	1	1,029851
Q	0,477612	0,328358	0,208955	0,328358	0,597015	0,38806	1,029851	1,029851	1,059701	1,268657
R	0,268657	0,208955	0,358209	0,208955	0,268657	0,149254	1,149254	1,268657	1,208955	1,179104
S	0,149254	0,179104	0,179104	0,179104	0,179104	0,119403	1,029851	1,029851	1	1,029851
T	0,059701	0,029851	0,029851	0,059701	0,059701	0	1,029851	1,029851	1,029851	1,059701
U	0,238806	0,119403	0,208955	0,208955	0,119403	0,179104	1,119403	1,179104	1,119403	1,119403
V	0,149254	0,029851	0,089552	0,149254	0,059701	0,089552	1	1	1	1,059701
W	0,328358	0,447761	0,179104	0,119403	0,268657	0,268657	1	1	1,059701	1,029851
X	0,208955	0,089552	0,149254	0,208955	0,268657	0,179104	1,029851	1	1	1,089552
Y	0,208955	0,298507	0,238806	0,268657	0,238806	0,149254	1,208955	1,179104	1,238806	1,208955
Z	0,567164	0,328358	0,149254	0,447761	0,447761	0,268657	1,089552	1,119403	1,089552	1,358209

Tabela 4.9: Matriz Geral que apresenta as distâncias entre as matrizes de confusão para conjunto de Teste - Classificadores individuais e combinados

	4-5H	4-5V	4-7	5H-5V	5H-7	5V-7	4-5H-5V	4-5H-7	4-5V-7	5H-5V-7
A	0,17910	0,08955	0,05970	0,14925	0,17910	0,08955	1,00000	1,00000	1,00000	1,08955
B	0,20896	0,23881	0,26866	0,20896	0,29851	0,14925	1,08955	1,08955	1,08955	1,08955
C	0,14925	0,41791	0,11940	0,32836	0,20896	0,44776	1,08955	1,02985	1,05970	1,02985
D	0,20896	0,20896	0,20896	0,11940	0,35821	0,32836	1,00000	1,02985	1,05970	1,02985
E	0,08955	0,17910	0,29851	0,11940	0,23881	0,14925	1,05970	1,05970	1,14925	1,11940
F	0,05970	0,11940	0,05970	0,08955	0,08955	0,11940	1,02985	1,02985	1,05970	1,05970
G	0,17910	0,14925	0,23881	0,20896	0,20896	0,14925	1,05970	1,14925	1,08955	1,02985
H	0,17910	0,35821	0,23881	0,44776	0,23881	0,20896	1,00000	1,00000	1,00000	1,02985
I	0,17910	0,14925	0,29851	0,23881	0,26866	0,29851	1,05970	1,05970	1,05970	1,00000
J	0,14925	0,26866	0,20896	0,17910	0,17910	0,14925	1,05970	1,05970	1,17910	1,14925
K	0,29851	0,26866	0,23881	0,14925	0,26866	0,20896	1,17910	1,11940	1,11940	1,08955
L	0,11940	0,14925	0,11940	0,14925	0,11940	0,14925	1,02985	1,02985	1,02985	1,08955
M	0,20896	0,17910	0,11940	0,14925	0,17910	0,11940	1,00000	1,00000	1,00000	1,02985
N	0,29851	0,17910	0,38806	0,23881	0,23881	0,38806	1,00000	1,17910	1,05970	1,02985
O	0,20896	0,20896	0,26866	0,14925	0,23881	0,23881	1,14925	1,14925	1,17910	1,08955
P	0,20896	0,11940	0,20896	0,08955	0,11940	0,14925	1,11940	1,17910	1,08955	1,02985
Q	0,44776	0,20896	0,11940	0,32836	0,38806	0,14925	1,05970	1,02985	1,02985	1,08955
R	0,08955	0,17910	0,14925	0,14925	0,14925	0,11940	1,05970	1,00000	1,08955	1,08955
S	0,14925	0,14925	0,20896	0,08955	0,17910	0,17910	1,08955	1,05970	1,05970	1,02985
T	0,02985	0,05970	0,02985	0,02985	0,00000	0,02985	1,02985	1,02985	1,02985	1,00000
U	0,14925	0,11940	0,23881	0,17910	0,11940	0,17910	1,00000	1,08955	1,05970	1,02985
V	0,26866	0,17910	0,32836	0,23881	0,08955	0,23881	1,00000	1,00000	1,02985	1,05970
W	0,17910	0,11940	0,20896	0,26866	0,20896	0,29851	1,00000	1,05970	1,00000	1,02985
X	0,17910	0,32836	0,20896	0,17910	0,23881	0,26866	1,11940	1,08955	1,11940	1,05970
Y	0,29851	0,23881	0,26866	0,17910	0,23881	0,17910	1,17910	1,20896	1,14925	1,02985
Z	0,14925	0,14925	0,17910	0,11940	0,17910	0,14925	1,05970	1,05970	1,08955	1,08955

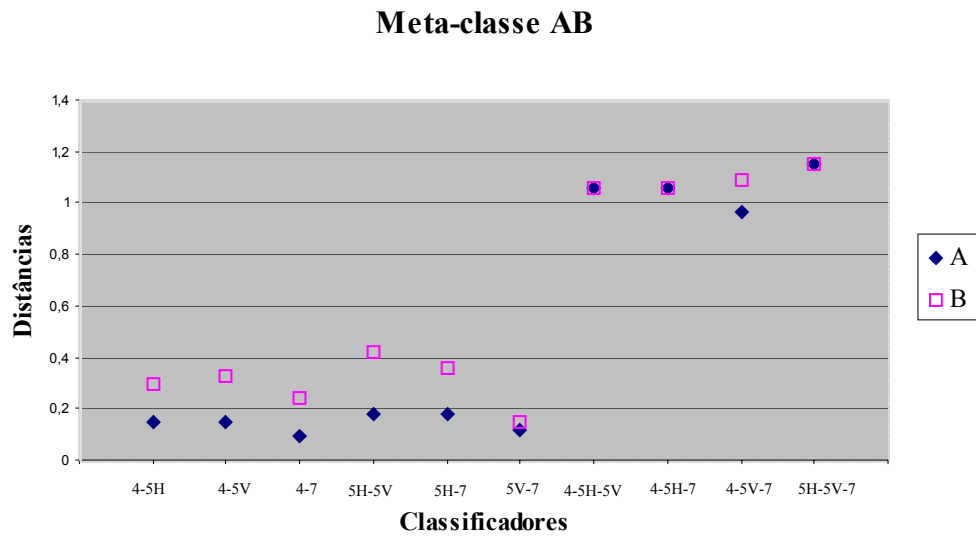


Figura 4.4: Gráfico que representa a meta-classe AB projetada no espaço, conforme a distância entre os classificadores

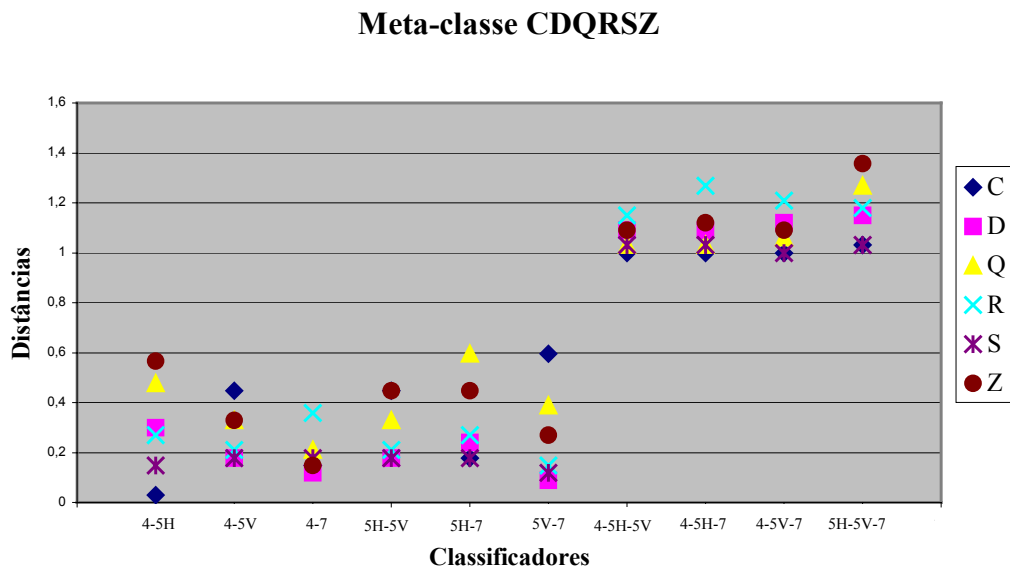


Figura 4.5: Gráfico que representa a meta-classe CDQRSZ projetada no espaço, conforme a distância entre os classificadores

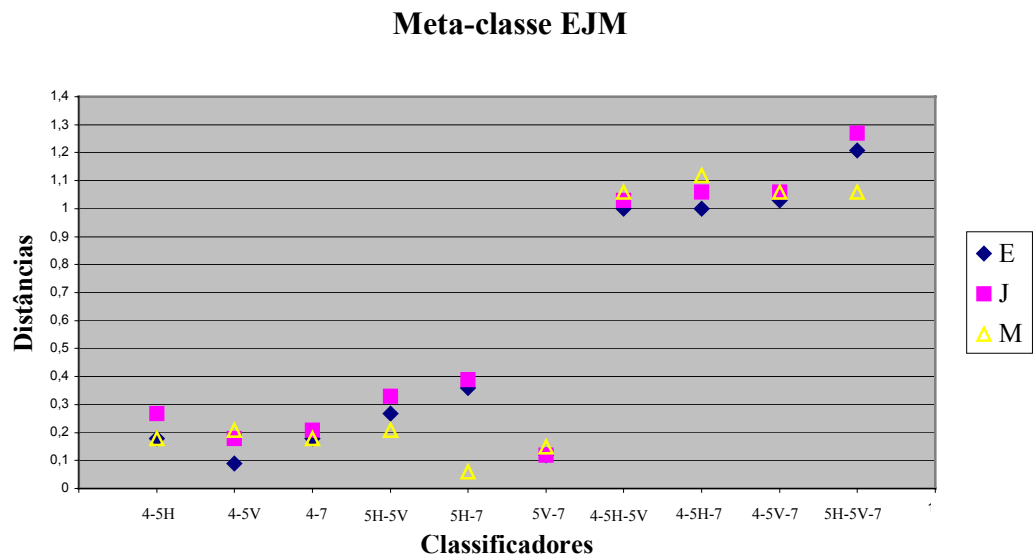


Figura 4.6: Gráfico que representa a meta-classe EJM projetada no espaço, conforme a distância entre os classificadores

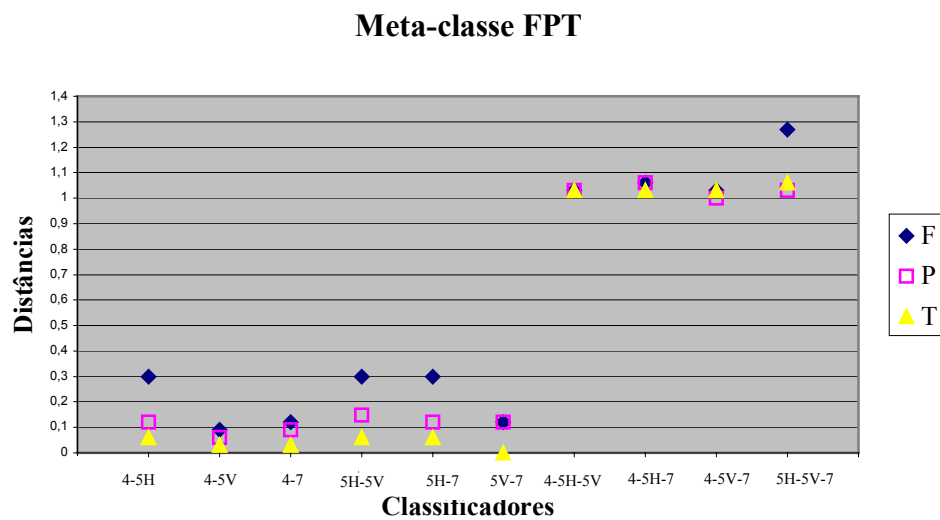


Figura 4.7: Gráfico que representa a meta-classe FPT projetada no espaço, conforme a distância entre os classificadores

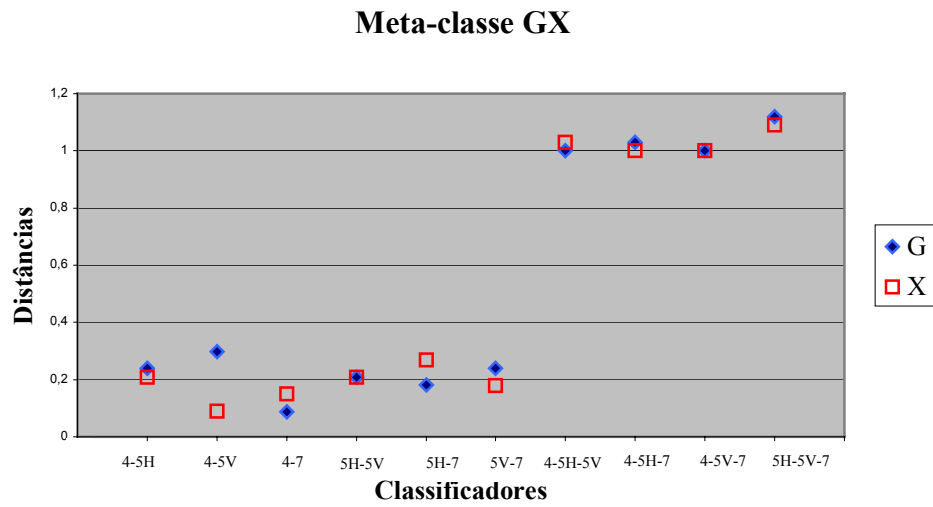


Figura 4.8: Gráfico que representa a meta-classe GX projetada no espaço, conforme a distância entre os classificadores

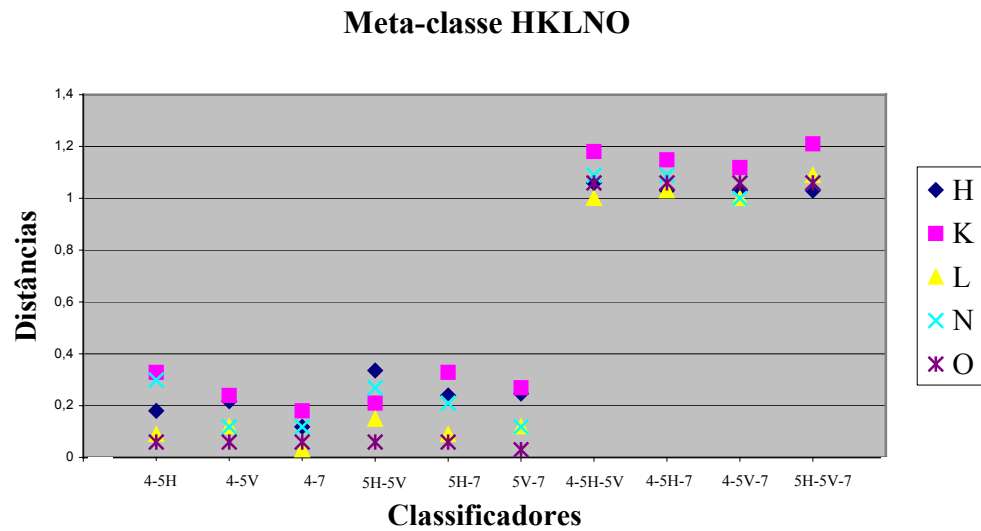


Figura 4.9: Gráfico que representa a meta-classe HKLNO projetada no espaço, conforme a distância entre os classificadores

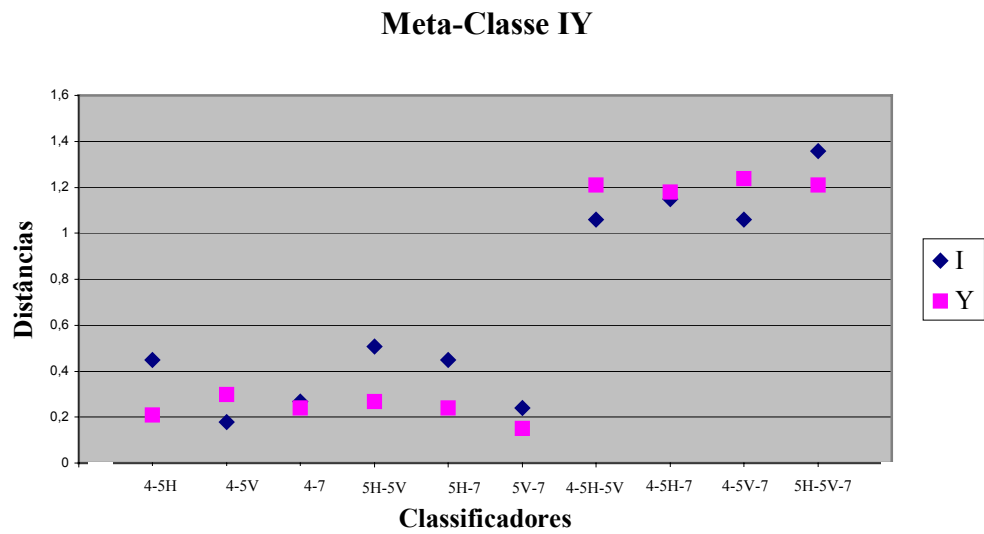


Figura 4.10: Gráfico que representa a meta-classe IY projetada no espaço, conforme a distância entre os classificadores

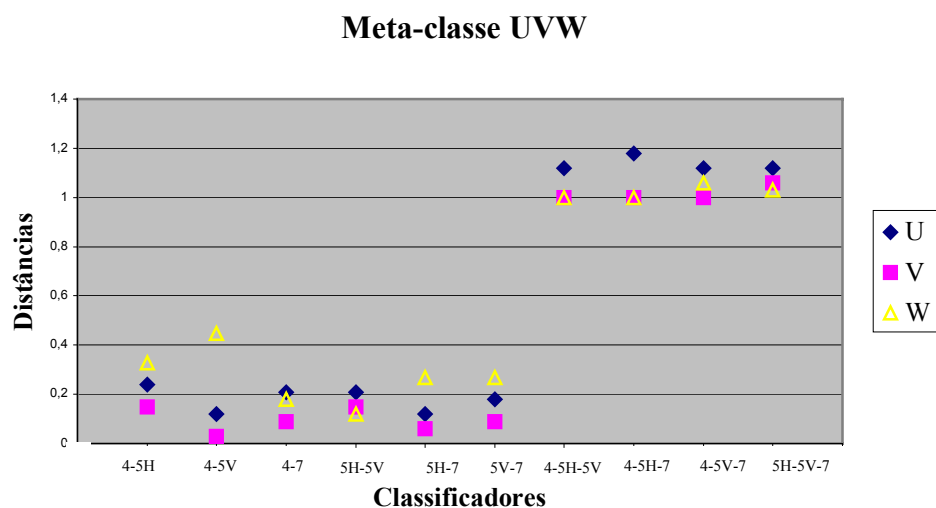


Figura 4.11: Gráfico que representa a meta-classe UVW projetada no espaço, conforme a distância entre os classificadores

As Tabelas 4.10 e 4.11 mostram os resultados obtidos para a metodologia baseada no cálculo das distâncias entre matrizes de confusão aplicando-se a regra da mediana ou *soft-correlation (DD-based)*. Assim, observa-se que *a priori*, com base na validação, os conjuntos a serem analisados são:

- 4-5V e 4-5H para combinação de 2 classificadores e

- 4-5V-7 e 4-5H-7 para combinação de 3 classificadores.

Observa-se que a indicação da combinação **4-5H-7** pela mediana pode ser confirmada através da taxa de reconhecimento (Tabelas 4.6 e 4.10). Em relação a indicação dos classificadores 4-5V e 4-5H (Tabela 4.10) observa-se que estes classificadores não correspondem com as melhores taxas de reconhecimento alcançadas (Tabela 4.6 e 4.10).

Tabela 4.10: *DD-based* e taxas de reconhecimento para os múltiplos classificadores - Conjunto de Validação

Classificador	DD	%Rec.
4-7	4,03	90,3
5V-7	4,64	89,1
4-5V	5,02	88,5
Mediana	5,57	-----
4-5H	6,12	89,2
5H-7	6,24	90,0
5H-5V	6,40	87,9
4-5H-5V	27,49	89,7
4-5V-7	27,52	90,9
Mediana	27,70	-----
4-5H-7	27,88	91,0
5H-5V-7	29,64	90,1

Para o conjunto de teste, Tabela 4.11, considerando-se o cálculo das distâncias entre matrizes de confusão e aplicando-se a regra da mediana ou *soft-correlation (DD-based)*, observa-se que *a priori*, os conjuntos a serem analisados são:

- 4-5V e 5V-7 para combinação de 2 classificadores e
- 4-5H-5V e 4-5H-7 para combinação de 3 classificadores.

Observa-se que apesar das distâncias apontarem as mesmas combinações 4-5V e 4-5H-7, as taxas de reconhecimento com o conjunto de teste diferenciam-se das alcançadas pelo conjunto de validação, demonstrando a necessidade de novos experimentos com mais classificadores ou em contextos diversos (palavras ou dígitos).

Tabela 4.11: *DD-based* e taxas de reconhecimento para os múltiplos classificadores - Conjunto de Teste

Classificador	DD	%Rec.
5H-5V	4,78	83,9
4-5H	4,87	83,4
4-5V	5,01	83,7
Mediana	5,07	-----
5V-7	5,13	85,2
5H-7	5,22	85,1
4-7	5,28	85,8
5H-5V-7	27,49	85,9
4-5H-5V	27,52	84,3
Mediana	27,65	-----
4-5H-7	27,79	85,5
4-5V-7	27,88	85,6

A Figura 4.12 mostra a arquitetura final do sistema composto com as meta-classes definidas e os conjuntos de classificadores determinados pelas distâncias entre matrizes de confusão e, ainda, pela aplicação da regra da mediana ou *soft correlation (DD-based)*.

Observa-se que o presente trabalho comparado aos apresentados na literatura [POI02] para os classificadores individuais estão no mesmo nível, visto que os resultados aqui apresentados podem ser ainda melhorados. A taxa de reconhecimento de [POI02] era de 87,1% para a arquitetura MLP, visto que neste estudo para o conjunto de teste a taxa de reconhecimento foi de 84,7% para $Z=7$, e no conjunto de validação é de 88,9% para $Z=7$.

A Tabela 4.12 apresenta um resumo dos resultados obtidos com os testes considerando-se a arquitetura de um time ou comitê de redes neurais [SIL03] para o contexto em questão. Observa-se que a média geral da taxa de reconhecimento alcançada é de 84,15%, ficando próximo da maior taxa alcançada com o classificador individual $Z = 7$ e, ainda, entre a mediana considerada para a combinação de 2 classificadores (4-5V para 83,7% e 5V-7 para 85,2%) e de 3 classificadores (4-5H-5V para 84,3% e 4-5H-7 para 85,5%).

Considera-se na Tabela 4.12 como rejeição os exemplares analisados que resultam como não pertencentes a meta-classe durante o reconhecimento. Por exemplo, o sistema recebe um caractere W para analisar na meta-classe AB, atribuindo ao mesmo uma probabilidade *a posteriori* muito baixa, sendo classificado como não identificado para a meta-classe AB. Como erro, considera-se os exemplares analisados que pertencem a meta-classe e quando são reconhecidos por um classificador individual resultam incorretos. Por exemplo, o sistema recebe um caractere A para ser analisado na meta-classe AB, o qual resulta uma

probabilidade *a posteriori* como A ou B, porém no nível seguinte, o classificador individual determina o caractere como não identificado.

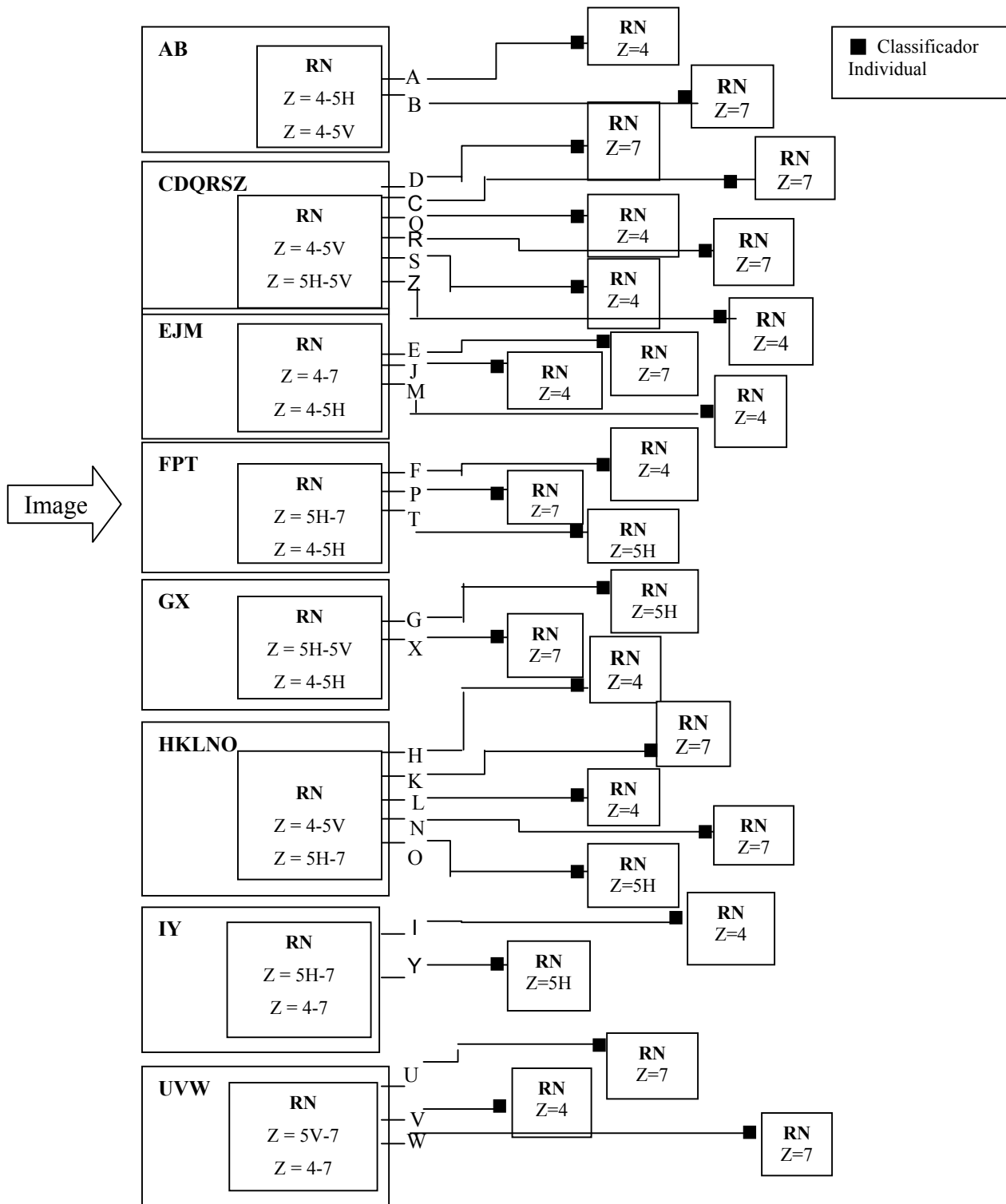


Figura 4.12: Meta-classes geradas pelas das medianas das distâncias do conj. de Validação.

Tabela 4.12: Matriz que apresenta os acertos, erros e rejeições das meta-classes definidas, validadas no conjunto de Teste.

	Classificadores	Acerto	Rejeição	Erro	% Acerto	%Rejeição	%Erro
A	4-5V	60	4	3	89,55	5,97	4,47
B	4-5V	49	14	4	73,13	20,89	5,97
C	4-5V	55	11	1	82,08	16,41	1,49
D	4-5V	48	17	2	71,64	25,37	2,98
Q	4-5V	55	9	3	82,08	13,43	4,47
R	4-5V	60	6	1	89,55	8,95	1,49
S	4-5V	50	6	11	74,62	8,95	16,41
Z	4-5V	60	6	1	89,55	8,95	1,49
E	4-7	60	7	0	89,55	10,44	0,00
J	4-5H	55	10	2	82,08	14,92	2,98
M	4-7	59	7	1	88,05	10,44	1,49
F	4-5H	60	5	2	89,55	7,46	2,98
P	5H-7	63	3	1	94,02	4,47	1,49
T	5H-7	65	2	0	97,01	2,98	0,00
G	4-5H	57	8	2	85,07	11,94	2,98
X	4-5H	50	13	4	74,62	19,40	5,97
H	5H-7	55	7	5	82,08	10,44	7,46
K	5H-7	54	11	2	80,59	16,41	2,98
L	4-5V	60	7	0	89,55	10,44	0,00
N	5H-7	55	10	2	82,08	14,92	2,98
O	4-5V	59	8	0	88,05	11,94	0,00
I	4-7	48	17	2	71,64	25,37	2,98
Y	5H-7	56	10	1	83,58	14,92	1,49
U	5V-7	56	8	3	83,58	11,94	4,47
V	4-7	64	0	3	95,52	0,00	4,47
W	5V-7	53	2	12	79,10	2,98	17,91
Média Geral					84,15	11,95	3,90

Capítulo 5

Conclusão

O foco principal desta dissertação foi o desenvolvimento de um método automático para reconhecimento de caracteres manuscritos baseado em regiões perceptivas. Trata-se de uma abordagem analítica do problema, tendo em vista um léxico de pequena dimensão: caracteres manuscritos maiúsculos, ou seja, 26 letras do alfabeto. Pode-se concluir analisando a Tabela 4.7 que algumas representações são mais robustas e discriminantes entre as classes do caractere do que outras, sendo assim as combinações dos classificadores resultam em taxas de reconhecimento mais elevadas.

Uma metodologia para múltiplos classificadores foi proposta buscando definir uma abordagem alternativa ao processo de combinação de classificadores, ao invés da utilização do critério da taxa de reconhecimento. Esta abordagem define uma Discordância baseada na medida da distância (*Disagreement based on Distance - DD-based*), aplicando uma distância euclidiana computada entre as matrizes de confusão dos classificadores e uma regra de *soft-correlation* para indicar o melhor conjunto de classificadores. O interesse é prover uma enfoque alternativo que permita avaliar a priori os conjuntos de classificadores para determinar a melhor combinação entre eles, sem combina-los diretamente. A razão principal para isto é a dificuldade de executar uma busca exaustiva dentro do espaço de combinação dos classificadores quando se tem um grande número de classificadores a verificar. A abordagem usa informação contida nas matrizes de confusão para cada classificador individual., normalmente, os sistemas de reconhecimento utilizam a matriz de confusão para avaliar confusões locais, essas matrizes são utilizadas para prover mais informações sobre os classificadores e os problemas. A abordagem utiliza uma metodologia simples para avaliar a combinação dos classificadores, sem necessitar de algoritmos complexos. A matriz de

confusão é utilizada para computar as distâncias entre os classificadores, sendo que estas representam a discordância entre os classificadores a serem combinados, como apresentado nas Tabelas 4.8 e 4.9. A medida de diversidade não só avalia e compara os pares de classificadores, mas também pode ser aplicada a grupos de classificadores de tamanho distinto, ao contrário das medidas *pair-wise*.

A hipótese *soft-correlation* pode ser testada e validada nos experimentos apresentados na Seção 4. Observa-se que apesar das distâncias apontarem as mesmas combinações as taxas de reconhecimento com o conjunto de teste diferenciam-se das alcançadas pelo conjunto de validação. Demonstrando, a necessidade de novos experimentos com mais classificadores ou em contextos diversos (palavras ou dígitos).

A regiões perceptivas foram verificadas nas meta-classes geradas, observando-se que as similaridades são evidenciadas entre as classes, como apresentado nas Figuras 3.28 e 3.29. Observe a Figura 5.1, é possível identificá-la facilmente? Pode-se ler este caractere como um “H”, “M” ou “U”, pois existem confusões realizadas também pelo ser humano. O princípio da Gestalt nos auxilia a entender a percepção humana das formas para que possam utilizá-lo como auxílio ao desenvolvimento de sistemas computacionais. A busca dos mecanismos de zoneamento apresentados na Seção 3 auxiliam na resolução das confusões encontradas pelo sistema, mas muitas destas confusões (por exemplo, U e V) também ocorrem com o ser humano.



Figura 5.1: Exemplo de caractere da base IRONOFF

As taxas médias de reconhecimento obtidas, para os zoneamentos avaliados, foram as seguintes, considerando-se Z: **4** = 82,89%, **5H** = 81,75%, **5V** = 80,94% e **7** = 84,73%. As combinações realizadas entre os classificadores individuais apresentam uma melhoria na taxa de reconhecimento, sendo o melhor resultado de 85.9% para a rede **5H-5V-7**. O resultado global considerando uma arquitetura composta por 2 níveis de classificação (meta-classe e classe) atinge uma taxa média de reconhecimento de 84,15%, com rejeição de 11,95% e erro de 3,90%.

A Figura 5.2 apresenta exemplos de imagens que tiveram melhor taxa de reconhecimento com classificadores individuais. Observa-se em algumas classes a possibilidade de uma melhoria na fase de pré-processamento, como por exemplo, a correção da inclinação do caractere. Observando a Figura 5.2 acredita-se que esta correção possa evidenciar as características da imagem, melhorando a extração de primitivas que irão compor o vetor de entrada para a rede neural.

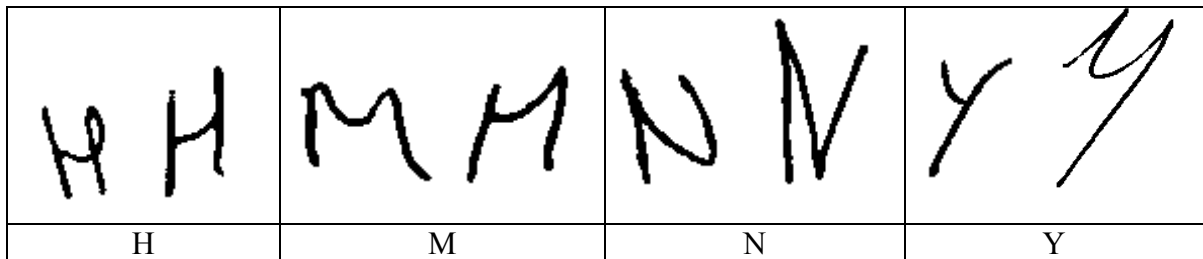



Figura 5.2: Imagens da base IRONOFF

A melhoria na extração de primitivas que complementem o vetor de características é, também, necessária visando aumentar o poder discriminante entre as classes, buscando-se um acréscimo nas taxas de reconhecimento. As extrações de primitivas direcionais poderiam ser utilizadas para complementar os conjuntos já estudados [OLI02b]. Não se descarta a aplicação de métodos baseados no contorno das formas, devido ao fato que a extração de primitivas locais necessita de uma representação da forma e de suas relações.

Referências Bibliográficas

- [AIR05a] AIRES, S.B. K., FREITAS, C. O. A., BORTOLOZZI, F. , SABOURIN, R. *Perceptual Zoning for Handwritten Character Recognition*. 12th Conference of the International Graphonomics Society - IGS, 2005 (publicado em junho de 2005, aguardando o recebimento dos Proceedings).
- [AIR05b] AIRES, S.B. K., FREITAS, C. O. A., BORTOLOZZI, F., NIEVOLA, J. *Redes Neurais Baseadas em Regiões Perceptivas para o Reconhecimento de Caracteres Manuscritos*. V Encontro Nacional de Inteligência artificial, V ENIA, 2005 (publicado em julho de 2005, aguardando o recebimento dos Anais).
- [COR02] CORREIA, S.E.N., CARVALHO, J.M., SABOURIN, R. *Human-Perception Handwritten Character Recognition using Wavelets*. In Proceedings of XVI Brazilian Symposium on Computer Graphics and Image Processing, IEEE Computer Society, 2002, p.404.
- [DUI04] DUIN, R. P.W., PEKALSKA, E., TAX, D. M. J. . *The characterization of classification problems by classifier disagreements*. In *ICPR'2004*, pages 140–143, Cambridge - UK, 2004.
- [FRE01] FREITAS, C.O.A. *Uso de modelos escondidos de Markov para reconhecimento de palavras manuscritas*. Tese de Doutorado, PUCPR/PPGIA, Curitiba-PR, 2001, 188p.
- [FRE02] FREITAS, C.O.A. *Percepção visual e reconhecimento de palavras manuscritas*. Monografia, Concurso de Promoção da Carreira Docente à Classe de professor Titular, PUCPR/PPGIA, Curitiba-PR, 2002, 95p.
- [FRE05] FREITAS, C.O.A, AIRES, S.B.K., BORTOLOZZI, F., OLIVEIRA JR., J.J., CARVALHO, J.M., SABOURIN, R. *DD-based Multiple Classifiers System Applied to Handwritten Character Recognition*, 10th Iberoamerican Congress on Pattern Recognition - CIARP, 2005 (submetido em 19/julho/2005).

- [HAD05] HADJITODOROV, S. T., KUNCHEVA, L. I., TODOROVA, L. P. . Moderate diversity for better cluster ensembles. Disponível *on-line* at http://www.informatics.bangor.ac.uk/~kuncheva/-recent_publications.htm, 2005.
- [HEU94] HEUTTE, L. *Reconnaissance de caractères manuscrits: application à la lecture automatique des chèques et des enveloppes postales*. Docteur Thèse de L'Université de Rouen. Rouen, France, déc., 1994. 239p;
- [HUN00] HUNG, K.W., LEUNG, W.N., LAI, Y-C. *Boxing Code for Stroke-Order Free Handprinted Chinese Character Recognition*. IEEE, 2,p.2721-2724.
- [KAM96] KAMIENSKI, C. A. *Introdução ao paradigma de orientação a objetos*. Faculdade Paraibana de Processamento de Dados, 1996.
- [KAP03] KAPP, M. N, FREITAS, C.O.A., NIEVOLA, J., SABOURIN, R. *Evaluating the conventional and class-modular architectures feedforward neural network for handwritten word recognition*. In Proceedings of XVI Brazilian Symposium on Computer Graphics and Image Processing, IEEE Computer Society, 2003, p.315–319.
- [KAP04] KAPP, M. N. *Reconhecimento de palavras manuscritas utilizando redes neurais artificiais*. Dissertação de mestrado, PUCPR, PPGIA, Curitiba-PR, 2004. 98p.
- [KIT98] KITTLER, J. HATEF, M., DUIN, R.P.W. e MATAS, J. *On Combining Classifiers*. IEEE Trans. On Pattern Analysis and Machine Intelligence, 20(3):226-239,1998.
- [KUN03] KUNCHEVA, L. I., WHITAKER, C. J. . *Measures of diversity in classifier ensembles and their relationship with the ensemble accuracy*. *Machine Learning*, 51(2):181–207, 2003.
- [LAL00] LALLICAN, P.M., VIARD-GAUDIN, C., KNERR, S. *From Off-line to On-line Handwriting Recognition*. Seventh International Workshop on Frontiers in Handwriting Recognition, pp.303-312, 2000.
- [LAM95] LAM, L, SUEN, C.Y. *Optimal Combinations of Pattern Classifiers*. *Pattern Recognition Letters*, 16(3):945-954, 1995.

- [LI_95] LI, Z.C., SUEN, C.Y., GUO,J. *A Regional Decomposition Method for Recognizing Handprinted Characters*. IEEE, Transactions on Systems, Man, and Cybernetics, Vol.25, junho 1995, p. 998-1010.
- [LIU02] LIU, C.L., SAKO, H., FUJISAWA, H. *Performance Evaluation of Pattern Classifiers for Handwritten Character Recognition*. International Journal on Document Analysis and Recognition, 4:191-204,2002.
- [MAT04] MATOS, L.N. *Utilização de Redes Bayesianas Como Agrupador de Classificadores Locais e Globais. Tese de doutorado*. UFCG, Campina Grande – Paraíba, 2004.
- [OH_02] OH, I-S, SUEN, C. Y. *A class-modular feedforward neural network for handwriting recognition*. Pattern Recognition, 35:229–244, 2002.
- [OLI02] OLIVEIRA JR, J. J.; CARVALHO, J.M. de C.; FREITAS, C. O.A.; SABOURIN R. *Evaluating NN and HMM classifiers for handwritten word recognition*. 15th Brazilian Symposium on Computer Graphics and Image Processing, 2002. p. 210-217.
- [OLI02b] OLIVEIRA, L.S., SABOURIN, R., BORTOLOZZI, F., and SUEN, C.Y. *Automatic Recognition of Handwritten Numerical Strings: A Recognition and Verification Strategy*. IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 24, N. 11, Pages 1438-1454, IEEE Computer Society Press, 2002.
- [OLI04] OLIVEIRA JR, J. J., KAPP, M.N., FREITAS, C., DE CARVALHO, J.M. and SABOURIN, R., *Handwritten Month Word Recognition Using Multiple Classifiers*, XVII Brazilian Symposium on Computer Graphics and Image Processing" (SIBGRAPI), October 17-20, 2004, Curitiba, Brazil.
- [OLI04b] OLIVEIRA JR, J.J. *Combinação de Classificadores no Reconhecimento de palavras manuscritas*. Projeto de qualificação de doutorado. UFCG, 2004.
- [OLS05] OLIVEIRA, L. S., MORITA, M. , SABOURIN, R. , BORTOLOZZI, F., *Multi-Objective Genetic Algorithms to Create Ensemble of Classifiers*, in the Proceedings of the Third

International Conference on Evolutionary Multi-Criterion Optimization (EMO 2005), Guanajuato, Mexico, LNCS 3410, March 9-11, 2005, ISBN 3-540-24983-4, pp 592-606.

- [OTS79] OTSU, N., *A threshold selection method from gray-level histograms*, IEEE Transactions Systems, Man. and Cybernetics, SMC 9, Vol.1, 1979, pp.63-66.
- [PAR02] PARK, Jaehwa. *An Adaptative Approach to Offline Handwritten Word Recognition*. IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol.24, No.7, jul.2002, p.920-931
- [PLA00] PLAMONDON, Réjean; SRIHARI, Sargur N. *On-Line and Off-Line Handwriting Recognition: A Comprehensive Survey*. IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol.22, NO.1, jan.2000, p.63-84.
- [PED95] PEDRAZZI, M, COLLA, A.M. *Simple Feature Extraction For Handwritten Character Recognition*, Proc. IEEE 1995 Int. Conference on Image Processing ICIP-95, vol. III, pp. 320-323, Washington, October 22-25 1995.
- [POI02] POISSON, E., VIARD GAUDIM, C, LALLICAN, P. M. *Multi-Modular Architecture Based On Convolutional Neural Networks For Online Handwritten Character Recognition*. International Conference on Neural Information Processing, Vol. 5, pp. 2444-2448, 2002.
- [SIL03] SILVA, E., THOMÉ, A.C.G. *Reconhecimento de caracteres Manuscritos Utilizando Time de Redes Neurais*, IV Encontro Nacional de Inteligência Artificial, 2003. p.13-16.
- [SUE91] LI, Z.C., SUEN, C.Y. e GUO, J. *Computer algorithms for recognizing the distinct parts of handprinted characters*, “. Proc. Conf. IEEE Sys Man, Cybern at Charlottesville, USA, 1991, pp. 197-201.
- [SUE92] SUEN, C.Y., GUO, J., LI, C. Z. “ *Computer and human recognition of handprinted character by parts*”. Proc. 2nd Int. Wkshp. On Frontiers in Handwriting Recognition. Amsterdam: North-Holand, 1992, pp. 224-236.

- [SUE94] SUEN, C.Y., GUO, J., LI, Z.C. *Analysis and Recognition of Alphanumeric Handprints by parts*. IEEE, Transactions on Systems, Man, and Cybernetics, Vol.24, abril 1994, p. 614-631.
- [TAY01] TAY, Y.H, LALLICAN, P.M, KHALID, M., GAUDIN, C.V, KNERR, S. *An offline Cursive Handwritten Word Recognition System*. IEEE Region 10 Conference, TENCON 2001, Singapore, August 2001,p.19-22.
- [TAY03] TAY, Y. H., KHALID, M, YUSOF, R, VIARD-GAUDIN, C. *Offline Cursive Handwriting Recognition System based on Hybrid Markov Model and Neural Networks*. IEEE International Symposium on Computational Intelligence in Robotics and Automation, pp. 1190-1195, 2003.
- [VIA99] VIARD-GAUDIN, C. . *The Ironoff User Manual*. IRESTE, University of Nantes, France, 1999.
- [VIA01] VIARD-GAUDIN, C., LALLICAN, P.M., KNERR, S. , BINTER, P. . The IRESTE On/Off (IRONOFF) Dual Handwriting Database, 2001.
- [WEB02] WEBB, A. *Statistical Pattern Recognition*. Jonh Wiley & Sons, 2002.
- [WIN05] WINDEATT, T.. *Diversity measures for multiple classifier system analysis and design*. *Information Fusion*, 6(1):21–36, 2005.
- [XU_92] XU, L., KRZYZAK, A., SUEN, C.Y. *Methods of Combining Multiple Classifiers and Their Applications to Handwriting Recognition*. IEEE Trans. On Systems, Man and Cybernetics, 22(3):418-435, 1992.
- [YAN98] YANG, Y-Y. *Adaptive Recognition of Chinese Characters: Imitation of Psychological Process in Machine Recognition*. IEEE Transactions on Systems, Man and Cybernetics – Part A: Systems and Humans, Vol. 28, No. 3, 1998. p.253-265.

[ZOU04] ZOUARI, H. K. . *Contribution à L'évaluation des Méthodes de Combinaison Parallèle de Classifieurs par Simulation*. PhD thesis, Université de Rouen, 2004.