Avaliação de Conjuntos de Características no Reconhecimento de Palavras Manuscritas

José Josemar de Oliveira Júnior

Dissertação de Mestrado submetida à Coordenação dos Cursos de Pós-Graduação em Engenharia Elétrica da Universidade Federal de Campina Grande como parte dos requisitos necessários para obtenção do grau de Mestre em Ciências no Domínio da Engenharia Elétrica .

Área de Concentração: Processamento da Informação

João Marques de Carvalho, Ph.D.
Orientador

Robert Sabourin, Ph.D.
Co-orientador

Cinthia Obladen de Almendra Freitas, D. Sc.
Co-orientadora

Campina Grande, Paraíba, Brasil

Avaliação de Conjuntos de Características no Reconhecimento de Palavras Manuscritas

José Josemar de Oliveira Júnior

Dissertação de Mestrado apresentada em Abril de 2002

João Marques de Carvalho, Ph.D.
Orientador

Robert Sabourin, Ph.D.
Co-orientador

Cinthia Obladen de Almendra Freitas, D. Sc.
Co-orientadora

Francisco Marcos de Assis, D. Sc.
Componente da Banca

Ronei Marcos de Moraes, D. Sc.
Componente da Banca

Campina Grande, Paraíba, Brasil, Abril de 2002

Dedicatória

Dedico este trabalho a memória da minha mãe, Maria da Salete e a minha avó, Gentila, por sempre ter lutado pelos meus sonhos como se fossem seus.

Agradecimentos

Agradeço, primeiramente a Deus, pelo dom da vida e por ter conseguido concluir este trabalho.

Ao Prof. João Marques, que me despertou o interesse pela pesquisa, sempre me incentivando e acreditando no meu potencial.

Ao Prof. Robert Sabourin, pelas sugestões e pelo interesse neste trabalho.

À Profa. Cinthia Freitas, por sua receptividade durante minha estada na PUC-PR, pelo incentivo nos momentos difíceis e pelas valiosas discussões.

À Luciana Veloso, pela amizade de tantos anos, pelas horas de estudo compartilhadas e pela análise e cessão de seus programas.

À Vânia e Suzete, pelas conversas do cotidiano, que tornavam os dias mais alegres.

Aos voluntários, que cederam suas escritas, bem como a todas as pessoas que me auxiliaram na coleta das amostras: Álvaro, Ana Luísa, Ana Karina, a Profa. Maria José Ribeiro, entre outros.

Aos amigos do mestrado, Madhavan, Towar, Felipe, Sérgio, Hallyson, Netto, Flávio e Christian, pela amizade e pelo companheirismo.

À todos do LAPS/UFCG, em especial à Rinaldo pela paciência e a todos que fazem o LARDOC/PUC-PR, principalmente ao Prof. Jacques Facon.

À COPELE na pessoa do Prof. Antônio Marcos e seus funcionários, Ângela, Marcos, Pedro e Eleonora, pela disponibilidade constante.

Ao CNPq e à CAPES, que deram o suporte financeiro para o desenvolvimento do trabalho.

Enfim, a todos que de algum modo contribuíram para a realização deste trabalho.

"É melhor tentar e falhar, que preocupar-se e ver a vida passar;
É melhor tentar, ainda que em vão, que sentar-se fazendo nada até o final;
Eu prefiro na chuva caminhar, que em dias tristes em casa me esconder;
Prefiro ser feliz, embora louco que em conformidade viver"

Martin Luther King

Resumo

Este trabalho apresenta uma avaliação comparativa de conjuntos de características utilizados no reconhecimento de palavras manuscritas. O principal objetivo é determinar um conjunto ótimo de características que representem as palavras referentes aos nomes dos meses do ano e extender as conclusões obtidas para outras aplicações. Neste intuito foi desenvolvido um sistema classificador neural de referência, que é usado na determinação do desempenho das características avaliadas. Três tipos de características são analisadas: características perceptivas, direcionais e topológicas. A avaliação mostra que considerando os conjuntos de forma isolada, o conjunto de características perceptivas produz os melhores resultados para o dicionário em questão. Estes resultados são melhorados quando os conjuntos de características e o sistema de referência são combinados com outro classificador, numa abordagem híbrida, obtendo uma taxa de reconhecimento média de 90,4%.

Abstract

This work presents a comparative evaluation of different feature sets used for handwritten word recognition. The main goal is to determine an optimum feature set to represent the handwritten names for the months of the year in Brazilian Portuguese language and to extend the conclusions obtained to other applications. For that purpose a baseline neural classifier was developed and used to determine the performance of the analysed feature sets. Three kinds of features are evaluated: perceptual, directional and topological. The evaluation shows that taken isolatedly, the perceptual feature set produces the best results for the lexicon used. These results can be improved combining the feature sets and the baseline system with other classifier, in a hybrid approach, that obtained an average recognition rate of 90.4%.

Conteúdo

Lista de Figuras

Lista de Tabelas

Capítulo 1

Introdução

No contexto atual, com os avanços na comunicação eletrônica ocorre a necessidade de disponibilizar a informação de uma forma cada vez mais rápida. Neste enfoque, documentos em papel parecem relíquia de um período distante, principalmente quando se fala em documentos manuscritos. Porém este pré-julgamento é falho, uma vez que o papel como meio de informação continua tendo diversas vantagens sobre outros meios:

- Papel é um meio padronizado, que não têm problemas de interface entre o escritor e o leitor;
- Papel é altamente portátil e seu transporte é bem estabelecido, embora seja mais lento que a transferência eletrônica de documentos;
- A escrita de um recado, de um endereço ou o preenchimento de um formulário à mão não necessita de pré-condições especiais, a menos da habilidade do escritor, da necessidade do papel e de algum instrumento de escrita.

Por outro lado, na era da informação tecnológica, as vantagens dos computadores e a sua superioridade no armazenamento, transferência e processamento de textos, dados e informações não pode ser desperdiçada [7]. Para resolver isso surgem os sistemas de leitura automática cuja tarefa principal é servir como ponte entre *o mundo* do papel e da escrita convencional e *o mundo* dos computadores e do processamento eletrônico.

Hoje em dia, as principais aplicações dos sistemas de leitura manuscrita podem ser encontradas em grandes organizações, onde um grande número de documentos similares tem de ser processados de maneira eficiente. Exemplos bem conhecidos dessas aplicações são a leitura de endereços postais, de cheques bancários e de formulários. Em muitas dessas aplicações os pesquisadores iniciaram explorando a informação numérica, para em seguida adicionar informações em relação aos caracteres do alfabeto, com o intuito inicial de melhorar os resultados do reconhecimento numérico, e depois para extrair informações alfabéticas adicionais.

Como um subconjunto destes sistemas, o reconhecimento de palavras manuscritas têm por objetivo investigar o problema da leitura automática de palavras cursivas. Para isso, o texto manuscrito precisa ser localizado, extraído e separado em palavras isoladas. Uma vez segmentado o texto em palavras, se estabelece o problema de qual seria a melhor forma de representar estas palavras considerando a grande variação existente entre elas quando provenientes de escritores diferentes.

1.1 Motivação

Uma forma correta de representar os dados é o ponto de partida de qualquer sistema de reconhecimento de padrões. Apesar dos esforços já realizados, no problema do reconhecimento de documentos, mais especificamente no problema do reconhecimento de palavras manuscritas, não existe um conjunto de características ou um modelo matemático consolidado.

Na literatura, diversos sistemas apresentados descrevem diferentes tipos de características para representar os dados [8, 9]. Contudo, comparações de conjuntos usando um sistema de referência, como ferramenta de avaliação, é necessário para responder a uma questão fundamental: **Qual o melhor tipo de característica para representar palavras manuscritas numa dada aplicação?**

Alguns autores [1, 10, 8] tem tentado incorporar o conhecimento existente sobre o processo de leitura humano em seus sistemas, justificando que a exploração de uma possível dualidade homem-computador tem sido aplicada em outras áreas com sucesso,

por exemplo, no reconhecimento da fala. Mas outra questão surge: **A introdução do conhecimento relativo à leitura humana no modelamento de sistemas de reconhecimento de palavras manuscritas é realmente eficiente e necessário?**

A solução dessas questões é o ponto de partida deste trabalho, porém é necessário definir a aplicação que dará suporte à essa investigação.

1.1.1 Definição do dicionário

Como as palavras manuscritas são padrões bastante complexos devido à grande variedade de estilos de escrita, a investigação desse problema só é tratável quando se provê um dicionário de palavras válidas. O dicionário é determinado pelo domínio da aplicação.

A aplicação escolhida para este trabalho foi o reconhecimento das palavras que representam os nomes dos meses do ano. Este é um problema importante pois constitui um sub-problema do reconhecimento de datas em cheques bancários. Embora esta aplicação possua um dicionário limitado de 12 classes, há palavras muito semelhantes e/ou com mesma terminação, o que pode afetar o desempenho global do sistema de reconhecimento: Jan***eiro***, Fever***eiro***, ***Ma***rç***o***, Abril, ***Ma***i***o***, ***Ju***n***ho***, ***Ju***l*ho*, Agosto, Set***embro***, Outu***bro***, Nov***embro*** e Dez***embro***.

1.2 Objetivos da dissertação

O objetivo principal deste trabalho é determinar um conjunto de características que representem adequadamente as palavras do dicionário em questão e apresentar o sistema de referên-desenvolvido no decorrer das atividades de pesquisa, que é usado para avaliar diferentes conjuntos de características. As técnicas empregadas em cada etapa do sistema, que vão desde a aquisição até o reconhecimento propriamente dito, passando pela definição dos diferentes conjuntos de características, também serão descritas. Por fim, os resultados obtidos são apresentados e avaliados procurando deste modo tirar conclusões que possam ajudar a responder às questões anteriormente formuladas

e sugerir um conjunto ótimo de características adaptadas à aplicação em questão e que possam ser estendidas para outras aplicações no domínio do reconhecimento de palavras manuscritas.

1.3 Organização do texto

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

O capítulo 2 apresenta um estudo sobre as técnicas de extração de características sugeridas na literatura, bem como uma revisão bibliográfica dos esquemas de representação de características utilizados em diversos sistemas.

O capítulo 3 contém uma descrição de cada etapa que compõe o sistema de referência desenvolvido neste trabalho. São apresentados a base de dados utilizada, os algoritmos usados no pré-processamento e a definição dos conjuntos de características a serem avaliados. Também é apresentada uma introdução às redes neurais e a caracterização do classificador utilizado.

O capítulo 4 apresenta os resultados experimentais obtidos considerando os conjuntos de características de maneira isolada e em conjunto. Também são apresentados os resultados obtidos considerando uma abordagem híbrida de classificação. O capítulo é concluído com uma comparação do sistema implementado com outros desenvolvidos para o mesmo dicionário.

O capítulo 5 contém a conclusão do trabalho e suas principais contribuições. Propostas de trabalhos futuros também são sugeridas no final do capítulo.

Capítulo 2

Técnicas de Extração de Características

O desempenho de qualquer algoritmo de classificação e/ou reconhecimento depende, em grande parte, da representação escolhida, ou seja, das características ou primitivas que são extraídas da entrada [11, 12]. O objetivo da etapa de extração de características é reduzir a variabilidade intraclasses e aumentar o poder discriminante entre as classes consideradas. Estas características devem, tanto quanto possível, resumir as informações que são pertinentes e úteis para a classificação e ao mesmo tempo eliminar as informações irrelevantes e desnecessárias.

Deste modo, na definição do conjunto de características é importante considerar alguns critérios básicos [13]:

- As características devem ser preferencialmente insensíveis à rotação, translação e variações de tamanho;
- As características devem ser de baixo custo computacional;
- As características devem ser independentes umas das outras, garantindo a utilização eficiente da informação contida no vetor de características.

Em relação ao reconhecimento de palavras manuscritas, as características são definidas geralmente em função da estratégia de reconhecimento adotada, que pode

ser analítica ou holística [8]. Quando a abordagem é holística, a palavra é reconhecida como uma unidade única, indivisível, sendo o conjunto de características extraído da palavra como um todo. Por outro lado, quando a abordagem é analítica, a identidade da palavra é determinada através da identificação dos caracteres independentemente, de modo que as características são obtidas a partir dos segmentos que compõem a palavra em questão.

As estratégias holísticas geralmente justificam as características utilizadas por meio de estudos psicológicos sobre o processo de leitura humano. Por sua vez, as estratégias analíticas utilizam características adaptadas dos sistemas de reconhecimento de caracteres numéricos isolados. A seguir, é apresentado um resumo dessas abordagens e das principais características que elas utilizam.

2.1 Abordagem holística

O processo de leitura humano tem sido objeto de diversos estudos que buscam o seu modelamento a fim de incorporá-lo nos sistemas de reconhecimento de palavras. Alguns trabalhos [14, 10] apresentam conclusões muito interessantes sobre este processo, descritas a seguir:

- Em um primeiro nível, as pessoas utilizam os ascendentes (d,k,l,h,t,b) e descendentes (q,y,j,g,p), sendo a letra f um caso especial, pois possui ambas as características;

- As consoantes possuem uma maior importância no processo de leitura do que as vogais, sendo possível ler ou reconhecer uma palavra sem a presença dessas letras ($handwriting = hndwrtng$);

- O processo de leitura das vogais (a,e,i,o) não apresenta confusões entre as mesmas, porém a letra u requer mais informações para ser diferenciada das letras w ou m;

- A primeira e a última letras de uma palavra são muito importantes no processo de reconhecimento;
- Palavras curtas para serem lidas requerem mais informações no final das mesmas;
- O final das palavras, a barra de corte da letra t e o ponto da letra i deterioram o processo de reconhecimento quando são mal interpretados;
- Uma letra é confundida geralmente com outra que tenha mais primitivas do que com aquelas que possuem menos. Por exemplo, l é mais confundido com t do que o inverso;
- As palavras são reconhecidas por seu comprimento, contorno exterior e letras no inicio e no fim da palavra.

Estudos psicológicos também sugerem que a leitura é feita usando codificações das formas das palavras a partir de um conhecimento prévio do leitor [8]. De modo que palavras escritas em minúsculo, por serem mais irregulares, são mais fáceis de ler do que palavras em caixa alta. Também é previsto que o desempenho do reconhecimento é degradado quando a forma da palavra está corrompida.

Esses resultados indicam que as características que melhor se adequam a uma representação holística das palavras são as características estruturais de alto nível, como junções e pontos finais, bem como as características perceptivas, baseadas na percepção do olho humano, como pontos isolados, laços, ascendentes, descendentes, junções T e estimativas do comprimento da palavra.

A partir disso, Madhvanath [8] classifica as características em três níveis, de acordo com sua compatibilidade em relação à representação holística:

- Nível baixo - características estruturais altamente localizadas como a distribuição da direção dos traços;
- Nível intermediário - características que permitem um maior nível de abstração da imagem, incluindo junções, pontos finais, concavidades e traços horizontais e diagonais;

- Nível alto - características perceptivas tais como ascendentes, descendentes, laços e comprimento da palavra.

Apesar de ser uma formulação bastante interessante, a utilização da abordagem holística só se justifica em aplicações com dicionários pequenos devido ao menor número de confusão entre as classes, pela dificuldade de obtenção de modelos individuais para cada palavra, além da necessidade de uma base de dados de treinamento de grande dimensão.

2.2 Abordagem analítica

Na abordagem analítica, as palavras são identificadas a partir de uniões de segmentos, que representam caracteres isolados ou partes de caracteres. Nesta abordagem, são utilizadas características adaptadas do reconhecimento de caracteres numéricos. A classificação clássica feita por diversos autores [1, 9], dividem estas características em três tipos distintos: primitivas globais, primitivas baseadas na distribuição dos pixels e primitivas geométricas e topológicas. Essas categorias são analisadas a seguir.

2.2.1 Primitivas globais

Essa categoria inclui características extraídas de todos os pontos pertencentes a um retângulo, o qual circunscreve o segmento de palavra em questão. Elas representam a imagem como um todo, e não refletem propriedades locais, geométricas ou topológicas de uma região específica.

Geralmente, essas características são obtidas por meio de transformadas globais e expansões em séries, que decompõem a imagem em uma combinação linear de funções de base, buscando extrair características invariantes a operações como rotação e translação. Os métodos mais explorados tem sido a transformada e série de Fourier, a transformada de Walsh, a transformada de Hadamard, entre outros.

Transformações e expansões em série possuem grande facilidade de implementação e alta sensibilidade às distorções e variações de estilo, o que prejudica o poder discri-

minante das características obtidas.

2.2.2 Primitivas baseadas na distribuição dos pixels

As características desse grupo são extraídas a partir da distribuição estatística dos pontos que formam a imagem do segmento, produzindo um conjunto de dimensão reduzida. As características mais empregadas são citadas a seguir:

- **Características de zoneamento:** O retângulo que contém o segmento da palavra é dividido em várias regiões, sobrepostas e não sobrepostas, denominadas zonas. As características usadas para reconhecer o caractere refletem as densidades de pontos nessas regiões.

- **Momentos estatísticos:** Os momentos estatísticos dos pixels pretos em relação a um ponto de referência escolhido no segmento, tal como o centro de gravidade ou uma outra coordenada do sistema, são usados como características.

- **Características *loci*:** Para cada pixel branco do fundo da imagem, um conjunto de vetores verticais e horizontais são gerados e o conjunto de características é dado pelo número de interseções que esses vetores fazem com os contornos que formam o segmento.

- **Distâncias e cruzamentos:** A característica de cruzamento é obtida do número de vezes que o caractere é cortado por segmentos de linha traçados em direções específicas. As distâncias entre os pontos que formam o caractere e esses pontos específicos na imagem (por exemplo, os pontos que determinam o limite do retângulo que contém o caractere), formam um outro conjunto de características.

Uma considerável tolerância às distorções e às pequenas variações de estilos é observada nas três últimas características descritas. Para estes grupos existe uma certa dificuldade de implementação, mas por outro lado, essas técnicas provêm alta velocidade de processamento.

2.2.3 Primitivas geométricas e topológicas

Essa categoria é constituída por características que descrevem aspectos importantes da geometria e da topologia do desenho do caractere, podendo representar assim propriedades globais ou locais do caractere, tais como:

- **Segmentos de reta e de linhas curvas:** Neste caso, são extraídos traços verticais, horizontais e diagonais, bem como convexidades e concavidades apresentados pela geometria do caractere.
- **Pontos finais, interseções de linhas e cavidades:** Estas características são representativas da topologia do caractere, e incluem a representação de pontos finais, interseções de traços e a determinação do número de buracos contidos no segmento em análise.

Estas características apresentam uma alto grau de complexidade, o que as tornam difíceis de serem extraídas. Entretanto uma vez implementadas, permitem ao sistema uma grande velocidade de processamento. Elas possuem alta tolerância em relação à possíveis distorções e variações de estilos presentes na imagem.

As classificações mostradas nesta seção ajudam a compreender as diferentes interpretações das imagens que podem ser feitas pelo extrator de características dependendo da abordagem de reconhecimento utilizada.

2.3 Revisão bibliográfica

Nas seções anteriores foi feita uma tentativa de classificação das diferentes características utilizadas no reconhecimento de palavras manuscritas. Porém a maior parte dos autores formam conjuntos unindo tipos distintos de características, procurando assim uma melhor representação da palavra. Deste modo, esta seção tem por objetivo descrever alguns dos diferentes esquemas de extração de características encontrados em uma seleção dos principais sistemas de reconhecimento de palavras disponíveis na literatura.

Gader *et alli* [2] utiliza dois conjuntos de características. O primeiro é formado pelas características obtidas pela análise de cavidades e o segundo engloba aquelas, denominadas de características de valor de direção, o qual é obtido a partir de padrões que provêm informação direcional sobre o traçado da palavra.

Cavidades são definidas como regiões do fundo da imagem delimitadas pelos traços de um caractere no mínimo em três lados. Existem seis tipos de cavidades: leste, oeste, norte, sul, central e laços, rotuladas de acordo com a direção de suas aberturas. Um laço é uma região fechada completamente delimitada pelos traços que compõem o caractere enquanto uma cavidade central (falso laço) é uma região aberta que está cercada por traços em todos os lados.

Nesse trabalho, as cavidades são determinadas utilizando morfologia matemática. O resultado das operações morfológicas na imagem pré-processada é a criação de seis imagens binárias, uma para cada tipo de cavidade, chamadas de imagens de características, conforme mostra a Figura 2.1. As imagens de características juntamente com a imagem de entrada pré-processada são utilizadas para a atribuição de valores numéricos às características usando divisão em zonas. Cada imagem é dividida em 15 zonas, cujos cantos esquerdos superiores pertencem ao conjunto $\{(linha, coluna) | linha = 4 \times i, coluna = 4 \times j, i = 0, 1, ..., 4, j = 0, 1, 2\}$. É importante ressaltar que na etapa de pré-processamento as imagens são normalizadas em tamanho 24×16. Desta forma, para cada imagem de características e imagem de entrada pré-processada obtém-se 15 valores, contando o número de *pixels* ativos em cada zona. Estes valores são linearmente escalonados entre 0 e 1 e armazenados num vetor de características com dimensão 105.

Os valores de direção fornecem informações sobre as orientações dos traços nas zonas utilizando o contorno e o esqueleto da imagem. Os valores numéricos destas características são obtidos através da contagem do número de ocorrência dos padrões ilustrados na Figura 2.2, em cada uma das quinze zonas, utilizadas no cálculo das características de cavidades. O resultado é um conjunto de características de direção composto por 60 valores.

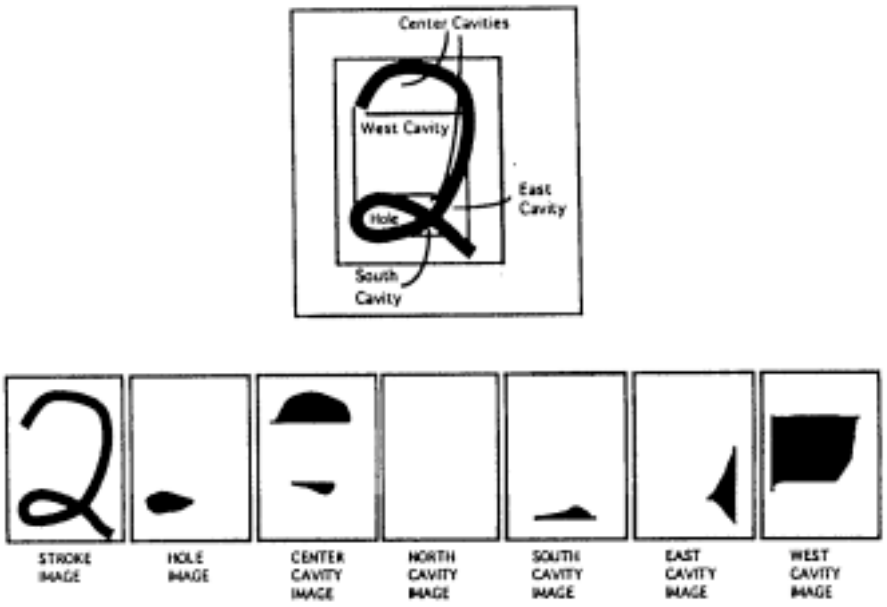


Figura 2.1: Definição de cavidades e exemplo da geração de imagens de características (extraída de Gader *et alli* [2]).

0	0	0
0	1	1
0	0	0

0	0	1
0	1	0
0	0	0

0	1	0
0	1	0
0	0	0

1	0	0
0	1	0
0	0	0

Figura 2.2: Padrões utilizados no conjunto de características direcionais (extraída de Gader *et alli* [2]).

Em outro trabalho [3], Gader *et alli* utilizam um vetor de características com 120 elementos que representam características de barra codificando as informações direcionais. Oito imagens de características são geradas, correspondendo às direções: leste, nordeste, norte e noroeste para regiões do fundo da imagem (*background*) e para o contorno dos caracteres (*foreground*). Na imagem de características, para cada ponto da imagem em análise é associado um valor inteiro que representa o comprimento do traço do caractere que passa pelo ponto em uma determinada direção, conforme pode ser observado na Figura 2.3. O vetor de características é calculado a partir das imagens de características usando a sobreposição de zonas. As zonas utilizadas têm um tamanho aproximado de $h/3 \times w/2$, com h e w representando a largura e a altura da imagem, respectivamente. Os cantos esquerdos superiores estão localizados nas posições aproximadas $\{(r,c)|r = 0, h/6, 2h/6, 3h/6, 4h/6$ e $c = 0, w/4, 2w/4\}$. Os valores em cada zona da imagem de características são somados e normalizados para um intervalo entre 0 e 1. Desta forma, o vetor de características possui uma dimensão de $15 \times 8 = 120$.

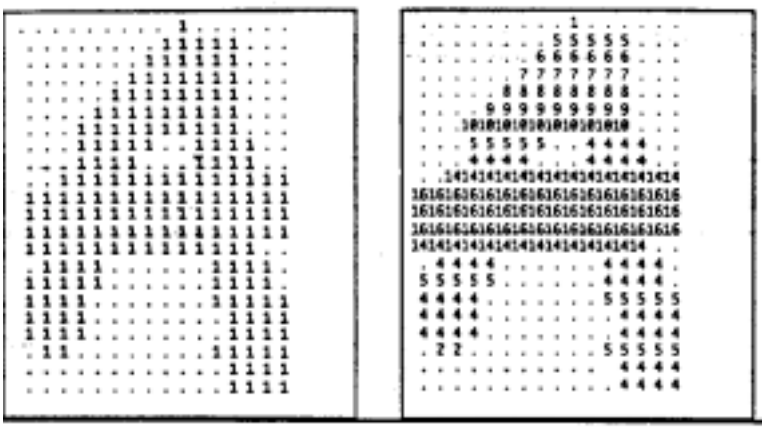


Figura 2.3: Imagem do caractere e a imagem de característica obtida para a direção horizontal (extraída de Gader *et alli* [3]).

No sistema desenvolvido por Gillies [15], o primeiro passo na detecção das características é a localização da região central da palavra. Esta região corresponde a uma faixa horizontal, onde são encontradas as letras minúsculas, tais como e, c e a. As letras ascendentes, descendentes e maiúsculas, geralmente, possuem partes dos seus traços em

regiões acima e abaixo da região central. Uma série de operações morfológicas é utilizada para rotular cada *pixel* na imagem de acordo com seu valor de pertinência nos traços, buracos e concavidades, localizados acima, abaixo e dentro da região central.

Yacoubi *et alli* [4] utiliza dois conjuntos de características. O primeiro conjunto é formado por características globais, como laços, traços ascendentes e descendentes. Os traços ascendentes e descendentes são representados por seus tamanhos relativos à altura da zona superior e inferior da palavra, respectivamente. Os laços são representados de várias maneiras, de acordo com sua localização em cada uma das três zonas (superior, central e inferior), e seus tamanhos relativos ao tamanho de cada zona. A localização dos laços centrais em relação aos traços ascendentes e descendentes dentro do segmento é considerada para permitir uma melhor discriminação entre algumas letras, tais como b e d ou entre p e q.

O segundo conjunto de características é obtido por uma análise bidimensional do histograma de transição do contorno, para cada segmento nas direções horizontal e vertical. Após uma fase de filtragem, as frequências presentes nos histogramas podem ser 2, 4 ou 6. Em cada histograma, o objetivo é sua parte central, representando a parte estável do segmento. Nesta parte central é determinado o número da transição dominante (2, 4 ou 6). Cada par de números da transição dominante é representado por um símbolo diferente. Após a criação de algumas subclasses por análise dos segmentos, esta representação conduz a um conjunto de 14 símbolos. A Figura 2.4 mostra três exemplos dessa codificação: as letras B, C e O, cujos pares de números de transições dominantes são (6,2), (4,2) e (4,4), são codificadas por símbolos chamados de B, C e O, respectivamente.

Nos sistemas propostos por Chen *et alli* [16, 17], são utilizados um conjunto de 35 características (características globais e locais) para identificar os segmentos. As características globais são extraídas dos segmentos da imagem, enquanto as características locais representam características dos traços que compõem os caracteres. As características de momentos (três primeiros momentos) são utilizadas para capturar a informação da forma global da palavra. Características geométricas e topológicas são

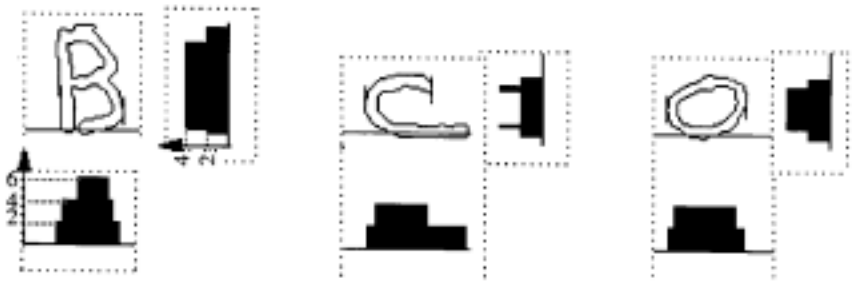


Figura 2.4: Exemplos dos histogramas de transição do contorno nas direções horizontal e vertical (extraída de Yacoubi [4]).

utilizadas para capturar as informações da forma da palavra em âmbito global e local. Neste sentido, os buracos, junções X e o número de interseções do traçado da letra com linhas imaginárias que passam pelo centro de gravidade do caractere nas direções horizontal e vertical são as características utilizadas. Algumas dessas características são úteis quando estão associadas com informações de zonas. Para capturar as características de zonas, o segmento é enquadrado num retângulo, e a seguir é dividido em três zonas verticais de acordo com a densidade dos segmentos. As características de zonas são utilizadas principalmente para capturar a localização vertical das informações topológicas: número de junções T, número de pontos finais e o número de segmentos de traços em cada zona.

Além das características citadas, outras são obtidas da distribuição dos *pixels* pretos, formando dois conjuntos. No primeiro conjunto são contados os números de *pixels* pretos em nove zonas. As características são obtidas por uma combinação linear destes valores. No segundo conjunto estão as características que informam a distribuição de *pixels* em toda a vizinhança das zonas. Por fim, têm-se as características de linhas de referência, que são utilizadas principalmente para capturar as relações entre os segmentos. Estas últimas características são de extrema importância, desde que não é garantido que os segmentos obtidos sejam caracteres completos, o que pode provocar a existência de segmentos de caracteres que são similares a outros segmentos de caracteres distintos, quando considerados isoladamente. Sendo assim, antes do algoritmo

de segmentação ser aplicado, a imagem da palavra é dividida em três zonas verticais, de acordo com o perfil de projeção vertical. A seguir, as características de linhas de referência são calculadas para cada segmento em cada zona como:

$$F_\alpha = \frac{y_{max}(\alpha) - y_{min}(\alpha) + 1}{rows(\alpha)}; \tag{2.1}$$

$y_{max}(\alpha)$ é o número máximo de *pixels* ativos nas linhas da zona α, $y_{min}(\alpha)$ é o número mínimo de *pixels* ativos nas linhas da zona α, $rows(\alpha)$ é o número total de linhas na zona. Calcula-se F_u, F_m e F_l representado as características de linha de referência calculadas nas regiões superior, central e inferior, respectivamente.

Kundu *et alli* [18] em um de seus trabalhos, gera um vetor de características com 14 elementos. As características extraídas são características da forma (números de junções X e T, número de laços, dentre outras) e características de distribuição dos *pixels*. Num outro trabalho, Kundu *et alli* [19] utilizou outras quatorze características. As primeiras três características são baseadas em momentos centrais, sendo portanto independentes de translação, rotação, orientação e do tamanho dos caracteres. A quarta, quinta e sexta características são, respectivamente, o número de laços, o número de junções T e o número de junções X na imagem. A sétima característica é baseada na razão entre a altura e a largura das letras. Algumas letras, por exemplo i e j, possuem pontos isolados que podem ser utilizados como pistas durante a etapa de reconhecimento. Por isso, a oitava característica é o número de pontos isolados. A nona e a décima características são o número de intercessões do traçado da letra com linhas imaginárias que passam pelo seu centro de gravidade nas direções horizontal e vertical. As últimas características são o número de semicírculos presentes na imagem do caractere nas direções norte, sul, leste e oeste. Todas as características foram normalizadas para um intervalo de 0 a 1, garantindo assim que nenhuma característica tenha um peso maior do que outra.

Bunke *et alli* [20] utiliza características baseadas nos nós e nas bordas do esqueleto da imagem a ser reconhecida (bordas são traços formados por *pixels* com dois vizinhos e nós são *pixel*s com um, três ou quatro vizinhos). Portanto, após o esqueleto da imagem

ser obtido, são extraídas as bordas do grafo do esqueleto da imagem. A seguir, cada borda é transformada num vetor de características de comprimento fixo.

Um total de dez características são utilizadas para descrever uma borda. As primeiras quatro características descrevem a localização espacial da borda. O grafo da imagem é dividido em quatro zonas horizontais. As características f_1 à f_4 são definidas como a percentagem de *pixels* da borda encontrados nas quatro zonas. A quinta característica (f_5) é uma característica binária que indica se uma borda é incidente a nodo de grau um ou não. A medida de curvatura é a sexta característica f_6, definida como a razão entre a distância euclidiana entre dois pontos finais da borda da imagem e seu comprimento. As características f_7 à f_{10} contêm mais detalhes sobre a curvatura das bordas.

Côté *et alli* [5] em seu sistema utiliza três tipos de características: primárias, secundárias e de vales. As características primárias são utilizadas para detectar letras chaves no corpo da palavra. As letras chaves são os componentes conectados que possuem traços nas regiões ascendentes e descendentes. Os componentes conectados que possuem laços em seu corpo são também considerados como sendo letras chaves. Características secundárias (*b-loops, d-loops*, ou as barras T) são condicionais, porque são apenas detectadas na presença de características primárias. As características de vale com cavidade para cima e/ou para baixo são extraídas do fundo da imagem. Os vales de cavidade para cima e de cavidade para baixo são componentes conectados do fundo da imagem extraídos entre os contornos superior e inferior da palavra. A Figura 2.5 ilustra as características utilizadas.

Wang *et alli* [21] apresentaram um sistema de reconhecimento de palavras manuscritas que utiliza uma técnica de extração de características tolerante ao erro de detecção da linha de base. Neste sistema, o método de extração de características realiza a codificação em zonas, em que a imagem é dividida em zonas horizontais, correspondendo às regiões ascendentes, descendentes e ao corpo principal da palavra. A região do corpo principal da palavra é dividida em duas partes. Desta forma, a palavra é divida em 4 regiões horizontais. É produzido um vetor de características para cada

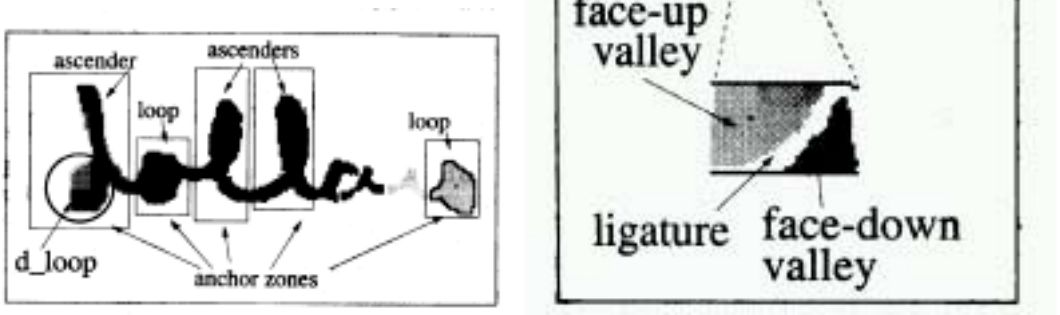


Figura 2.5: Exemplos da determinação de letras chaves e de características de vale (extraída de Côté [5]).

posição i da janela deslizante, dado por:

$$\overrightarrow{f_i} = (f_1, f_2, f_3, f_4)^T$$

Com $f_j = f(z_j)$, z_j sendo a j-ésima zona na direção do topo para a base e $f(z_j)$ representando alguma função de codificação, como por exemplo a percentagem de *pixels* pretos na zona.

Entretanto, a divisão da palavra em zonas é realizada baseada na detecção da linha de base da palavra. Devido à grande variação de estilos de escrita, a detecção precisa desta linha de base é extremamente difícil. Os métodos de detecção da linha de base são todos baseados em regras, que possuem sempre exceções devido às variações nas letras. Em muitos casos, a linha de base é adotada como sendo uma linha horizontal. Entretanto, tal suposição em alguns casos não é aceitável. Embora alguns algoritmos possam encontrar uma linha de base global com bastante precisão, eles ainda não podem evitar erros locais na codificação por zonas. Devido a esta imprecisão na detecção da linha de base, foi proposta uma modificação no método de codificação de zona por janela deslizante, visando diminuir a influência do erro na detecção da linha de base sobre o vetor de características. O novo vetor de características gerado é dado por:

$$\overrightarrow{f_e} = (f_1^-,\ f_1,\ f_1^+, f_2^-,\ f_2,\ f_2^+, ..., f_4^-,\ f_4,\ f_4^+).$$

$f_j = f(z_j)$, $f_j^- = f(z_j^-)$ e $f_j^+ = f(z_j^+)$. A zona z_j^- foi encontrada realizando um deslocamento na linha de base de duas posições para cima e aplicando o mesmo método de divisão por zona. A zona z_j^+ foi encontrada realizando procedimento similar, com deslocamento da linha de base de duas posições para baixo. Obviamente o vetor de característica $\overrightarrow{f_e}$ contém mais informações sobre a forma dos caracteres e possui um maior poder de discriminação do que o vetor de característica original $\overrightarrow{f_i}$. Desta forma, a descrição da letra por $\overrightarrow{f_e}$ pode ser mais tolerante a erros de detecção da linha de base.

Brakensiek *et alli* [22, 23] utiliza um vetor de características composto por 11 características, sendo 8 características obtidas dos coeficientes da DCT (*Discrete Cosine Transform*) e 3 características adicionais (altura, largura e o número de transições de preto para branco na linha média da palavra).

Guillevic e Suen [24] utilizam características extraídas do contorno da palavra (representado pelo código de *Freeman* [25]), por uma janela deslizante que percorre a imagem da esquerda para direita. Para uma dada posição da janela deslizante um vetor de características é extraído. O elemento chave a ser determinado é o tamanho ou a largura da janela deslizante, bem como a sobreposição das janelas. Nesse trabalho, a largura da janela deslizante foi fixada como sendo uma fração da altura do corpo principal da palavra (distância entre as linhas de base superior e inferior da palavra). A sobreposição entre sucessivas janelas deslizantes foi fixada em 50%. Cada janela deslizante foi dividida em regiões horizontais, correspondendo às regiões da palavra onde se encontram os traços ascendentes e descendentes e a região central da palavra. A região central, onde estão localizados os caracteres que não possuem traços ascendentes e descendentes, foi também dividida em três regiões horizontais (parte superior do corpo, área central do corpo, parte inferior do corpo). Para cada janela deslizante i, e para cada sub-janela j, calcula-se o número de pontos do contorno com valor de inclinação k com a horizontal, $Count(i, j, k)$. Os pontos do contorno podem assumir quatro valores de inclinação (k): 0, 1, 2 e 3 correspondendo à inclinação de 0, 45, 90 e 135 graus com relação à horizontal do segmento formado pelo *pixel* i e seu anterior, *pixel* $i - 1$, em que ambos pertencem ao contorno da imagem. Cada janela é representada por 20 características, pois para cada uma das 5 sub-janelas são extraídas 4 características de ângulos. Em complemento às características de ângulo são extraídas características adicionais para as sub-janelas ascendentes e descendentes, as quais codificam as posições verticais médias dos *pixels* em relação à posição da linha superior e inferior, respectivamente. Estas características ajudam a diferenciar as letras minúsculas que possuem traços na sub-janela ascendente (descendente) da verdadeira letra ascendente (descendente).

Freitas [1] em seu trabalho utiliza inicialmente um conjunto de características perceptivas para representar os segmentos da palavra. São utilizados ascendentes, descendentes e laços fechados que são representados pelo seu tamanho em relação ao corpo da palavra e pela sua informação posicional em relação ao eixo horizontal e às zonas

da palavra. Como características complementares para representar segmentos que não tenham essas características e para uma melhor representação do corpo da palavra são extraídas concavidades e convexidades, usando morfologia matemática. Os pontos convexos são determinados com o auxílio de uma família de 5 elementos estruturantes (matrizes 7x3) e os pontos côncavos com uma família de 10 elementos estruturantes (matrizes 9x4). Ambos os procedimentos são aplicados sobre o esqueleto da imagem original pelo processo de *template matching*. Em seguida, é aplicado um procedimento de rotulação para os *pixels* do fundo da área correspondente à parte interna da concavidade e/ou convexidade extraída da imagem. Após a rotulação de todos os *pixels* da área em questão, os rótulos são contados e classificados.

A descrição dos diversos esquemas de extração apresentados nesta seção mostram a utilização de diferentes tipos de características combinadas para a formação de conjuntos representativos. A Tabela 2.1 apresenta um quadro resumo dos principais tipos de características utilizados nos sistemas revisados. A sua análise mostra que cada autor define características diversas, geralmente direcionados para a sua aplicação. Isto dificulta uma conclusão bem fundamentada sobre qual é o conjunto mais representativo, o que corrobora para a motivação deste trabalho apresentada anteriormente.

2.4 Conclusão

A definição de características representativas no reconhecimento de palavras manuscritas é uma tarefa difícil. Diversos pesquisadores tem tentado incorporar modelamentos em relação ao processo de leitura humano, aliados à características já bem estabelecidas, como as utilizadas no reconhecimento de caracteres numéricos.

A revisão bibliográfica apresentada mostrou que os autores definem conjuntos de características diversos, geralmente direcionados para sua aplicação, o que dificulta uma conclusão mais profunda em relação ao conjunto de características mais representativo para o problema do reconhecimento de palavras manuscritas.

Sendo assim, a partir dos estudos mostrados e afim de tirar conclusões em relação ao potencial das diversas características, foram definidos três diferentes conjuntos tomando

Tabela 2.1: Quadro resumo das principais características utilizadas nos sistemas revisados. RF - Rotulação do Fundo da imagem; DT - Direção dos Traços; CP - Características Perceptivas; CE - Características Estruturais e DP - Distribuição dos *Pixels*.

	Referências											
Característica	[2]	[3]	[15]	[4]	[16, 17]	[18]	[19]	[20]	[5]	[21]	[24]	[1]
RF	X	X							X			X
DT	X	X			X		X				X	
CP			X	X					X	X	X	X
CE			X	X	X	X	X	X	X			
DP					X	X	X	X		X	X	

por base a classificação de Madhvanath [8]. Portanto, serão avaliadas desde características simples como o zoneamento até outras mais elaboradas como as perceptivas.

Na literatura não foi encontrada nenhuma ferramenta de avaliação de desempenho para este tipo de problema. A solução encontrada foi desenvolver um sistema de referência e incorporando diferentes conjuntos previamente definidos, avaliar seus comportamentos pela taxa de reconhecimento e pela análise de erros. Este sistema será descrito no próximo capítulo.

Capítulo 3

Descrição do Sistema de Referência

Este capítulo traz uma descrição do sistema de referência desenvolvido para a avaliação de primitivas no reconhecimento de palavras manuscritas, que é o objetivo deste trabalho. Este sistema tem como aplicação o reconhecimento das palavras dos meses do ano.

A Figura 3.1 mostra uma representação do sistema, em diagrama de blocos, cujas partes constituintes são:

- Aquisição - Amostras de palavras obtidas de formulários específicos, que foram digitalizadas usando dispositivo de *scanner*. Uma base de dados com 6000 imagens foi construída para ser usada neste trabalho;
- Pré-processamento - Conjunto de algoritmos aplicados para eliminação do ruído e normalização das imagens;
- Extração de características - Estratégia de segmentação implícita seguida por três diferentes conjuntos de características que extraem informações globais da palavra;
- Classificador neural - Redes neurais utilizadas para atribuição de um valor de confiança à imagem em análise em relação às 12 classes constituintes do problema.

A seguir, é feita uma descrição detalhada de cada um desses blocos.

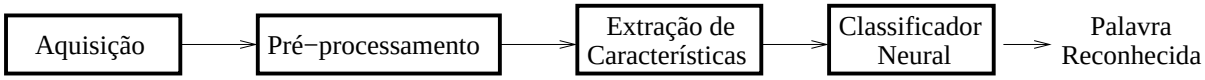


Figura 3.1: Representação em diagrama de blocos do sistema de referência desenvolvido.

3.1 Aquisição

O desenvolvimento de qualquer sistema de reconhecimento de padrões necessita de uma base de dados representativa das classes, para que o sistema possa ser construído e avaliado.

Em relação à lingua portuguesa, existe uma base de dados criada no LARDOC/PUC-PR que possui 2000 imagens de palavras dos meses do ano obtidas utilizando cheques de laboratório [26]. Esta base é dividida em três conjuntos: treinamento, com 1188 imagens, validação com 408 imagens e teste com 402 imagens.

Para o desenvolvimento do sistema proposto neste trabalho especificamente, foi construída uma nova base de dados que pudesse representar tão bem quanto possível os diferentes estilos de escrita presentes na região, no caso a cidade de Campina Grande - PB. Isto foi feito coletando-se amostras do nome de cada mês, de um total de 500 escritores na maioria estudantes de ensino médio e superior de instituições públicas e privadas.

Para isso, foi aplicado um formulário específico, em papel sulfite branco, onde cada voluntário devia escrever uma única vez o nome de cada mês, sem que fosse imposta qualquer restrição quanto ao estilo de escrita. Também não foi proposto nenhum modelo prévio para o voluntário seguir, orientando-o apenas para que escrevesse da forma mais natural possível. Isto resultou em uma base de dados bastante heterogênea, como ilustra a Figura 3.2 que contém algumas amostras retiradas desta base de dados. Após a coleta dos formulários, os mesmos passaram pela etapa de digitalização, em que foi utilizado um scanner HP Scanjet 5200 C [27], disponível no LAPS/DEE/UFPB ajustado para uma resolução de 200 dpi (*dots per inch*) com dois níveis de cinza. As

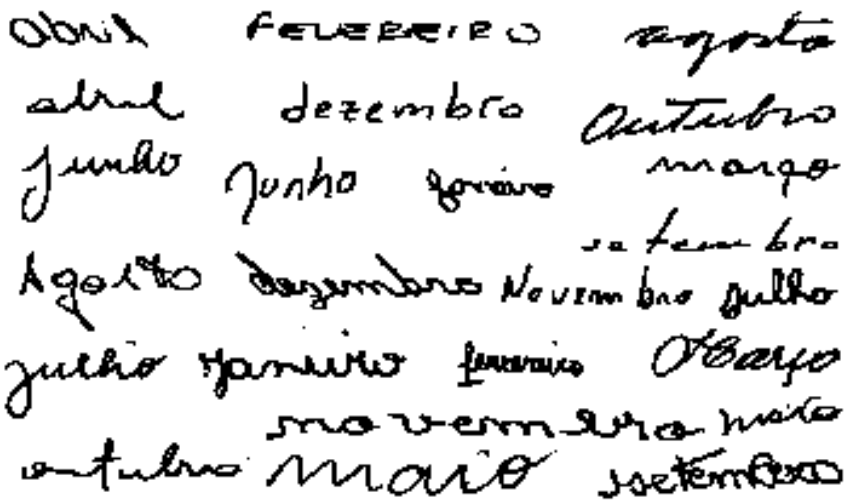


Figura 3.2: Amostras da base de dados de meses do ano do LAPS/UFCG.

imagens foram armazenadas em formato PCX e divididas em 12 conjuntos de arquivos, um para cada mês. Em seguida, a base foi dividida de forma aleatória em três conjuntos: treinamento, validação e teste que possuem 3600, 1200 e 1200 imagens cada, respectivamente.

A seguir é feita uma caracterização da base de dados em relação à distribuição dos diferentes estilos de escrita nos três conjuntos.

3.1.1 Caracterização da base de dados

Para caracterizar a base de dados foi feita uma análise com relação aos estilos de escrita encontrados. Segundo Tappert [6] podemos classificar a escrita cursiva em cinco categorias principais, conforme Figura 3.3:

1. Palavras em caracteres disjuntos contidos em retângulos pré-impressos (caixa alta);

2. Palavras em caracteres disjuntos com espaçamento regular;

3. Palavras em caracteres disjuntos com a presença de vínculos eventuais entre ca-

racteres;

4. Palavras em escrita cursiva pura, ou seja, todos os caracteres de uma palavra são conectados;

5. Palavras em escrita mista, ou seja, misturando os demais tipos de escrita.

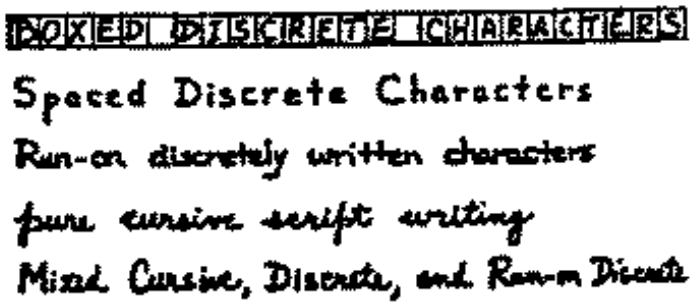


Figura 3.3: Tipos de escrita segundo a classificação de Tappert (extraída de Tappert *et al.* [6]).

Freitas [1] considera que a categoria 3 insere-se na categoria 5, classificando as imagens em quatro grupos: Cursiva pura, caixa alta, caracteres disjuntos e mista. Seguindo esta classificação, a Tabela 3.1 apresenta a distribuição dos tipos de escrita presentes nas bases de treinamento, validação e teste utilizadas neste trabalho.

Outro levantamento realizado foi a porcentagem de palavras com a primeira letra maiúscula, sendo determinado um percentual de 32%, 29% e 33% para os conjuntos de treinamento, validação e teste, respectivamente.

Estes levantamentos mostram que as distribuições dos estilos de escrita é praticamente uniforme nos três conjuntos e que ocorre uma maior predominância da escrita cursiva pura, porém a parcela de palavras em escrita mista é bem representativa, o que comprova a diversidade de estilos presentes na base de dados. O percentual de palavras com inicial maiúscula também é significativo, sendo este fator importante pois aponta que mesmo palavras de uma mesma classe produzem um nível de confusão elevado.

Tabela 3.1: Distribuição dos tipos de escrita nos subconjuntos da base de dados utilizada.

	Treinamento	**Validação**	**Teste**
Cursiva pura	57 %	61 %	61 %
Caixa alta	5 %	3 %	2 %
Caracteres disjuntos	8 %	7 %	11 %
Mista	30 %	28 %	26 %

3.2 Pré-processamento

O pré-processamento é uma parte fundamental de qualquer sistema de reconhecimento de palavras. Seu objetivo principal é reduzir a grande variação observada em diferentes amostras da mesma palavra, escrita pela mesma pessoa em instantes distintos ou por diferentes escritores.

Neste trabalho foram empregadas as técnicas de pré-processamento desenvolvidas por Veloso [28] que consistem de três etapas:

- Normalização da inclinação média dos caracteres da palavra;
- Normalização do declive da palavra;
- Suavização.

As etapas de normalização são necessárias pois os formulários não forneciam linhas de referência para o escritor, ocasionando a presença de palavras com diferentes inclinações em relação aos eixos horizontal e vertical. A etapa de suavização tem como objetivo retirar da imagem original os pontos isolados (ruído) e reduzir os picos e buracos existentes no contorno da imagem, resultantes de problemas ocorridos durante a digitalização das imagens ou ocasionados pelas operações de normalização.

A seguir, é feita uma descrição das técnicas empregadas e uma análise dos seus resultados.

3.2.1 Normalização da inclinação média dos caracteres da palavra

Para obter a normalização da inclinação média dos caracteres da palavra é realizada inicialmente uma operação morfológica de abertura, no intuito de prevenir que traços relativamente horizontais interfiram na determinação da inclinação das letras.

Em seguida, é calculado o perfil de projeção inclinado (PPI) da imagem em diferentes ângulos de inclinação, que variam de -60 a 60 graus em relação à vertical, com passo de 1 grau. O perfil de projeção inclinado indica a quantidade de *pixels* pretos existentes em colunas inclinadas. O algoritmo utilizado para o cálculo do perfil de projeção inclinado, para uma imagem $M \times N$ é descrito a seguir:

Para cada pixel (i, j) *da imagem;* $i = 1, 2, ..., M$ *e* $j = 1, 2, ..., N$.

Para cada ângulo de inclinação (θ).

1. *Determine o novo valor* v *da coordenada* j *na imagem, como sendo:*

$$v = \lfloor j - (M - i).tan(\theta) \rfloor, \tag{3.1}$$

   em que o operador $\lfloor . \rfloor$ *indica o inteiro mais próximo.*

2. *Se o valor do pixel em* (i, v) *for igual a* 1*, incremente a* v*-ésima coluna do* θ*-ésimo perfil de projeção inclinado (PPI).*

$$PPI_\theta(v) = PPI_\theta(v) + 1; \tag{3.2}$$

Uma vez obtidos os perfis é calculada a entropia associada a cada perfil de projeção, segundo a Equação 3.3:

$$H_\theta = -\sum_{v=1}^{L} P_v(\theta) \log P_v(\theta); \tag{3.3}$$

sendo L o número de linhas do perfil de projeção inclinado e P_v a probabilidade de um *pixel* preto ser encontrado na coluna inclinada v.

O ângulo (α) que proporciona a menor entropia é considerado o ângulo de inclinação média dos caracteres da palavra. Em seguida é realizada a normalização propriamente

dita, através de uma transformação, que rotaciona a imagem pelo ângulo de inclinação determinado. Esta transformação é descrita a seguir.

Para cada pixel na imagem original com coordenadas (i, j) *são calculadas as suas novas coordenadas* $(i\prime, j\prime)$ *na imagem normalizada, utilizando a equação 3.4.*

$$\begin{aligned} j' &= \lfloor j - (M - i).tan\alpha \rfloor, \\ i' &= i. \end{aligned} \tag{3.4}$$

Nesta equação, α é o ângulo pelo qual se deseja rotacionar os *pixels* da imagem com relação à normal, (i, j) são as coordenadas do *pixel* na imagem de entrada e (i', j') são as novas coordenadas do *pixel* na imagem de saída.

3.2.2 Normalização do declive da palavra

Para obter a normalização do declive da palavra é realizada inicialmente a extração do contorno inferior da palavra, com a finalidade de evitar que os pontos que não pertençam à linha de base da palavra interfiram no cálculo do declive.

Em seguida, é calculado o perfil de projeção horizontal inclinado (PPHI) em diferentes ângulos de inclinação, que variam de -60 a 60 graus de inclinação com relação à linha de referência horizontal, com passo de 1 grau. O perfil de projeção inclinado informa a quantidade de *pixels* pretos existentes em linhas inclinadas.

O algoritmo utilizado para o cálculo do perfil de projeção horizontal inclinado, para uma imagem $M \times N$, é descrito a seguir:

Para cada pixel (i, j) *da imagem;* $i = 1, 2, ..., M$ *e* $j = 1, 2, ..., N$.

Para cada ângulo de inclinação (θ).

1. *Determine o novo valor* l *da coordenada* i *na imagem, como sendo*:

$$l = i + j \tan(\theta); \tag{3.5}$$

2. *Se o valor do pixel* (l, j) *for igual a* 1, *incremente a* l*-ésima coluna do* θ*-ésimo perfil de projeção horizontal inclinado (PPHI).*

$$PPHI_\theta(l) = PPHI_\theta(l) + 1; \tag{3.6}$$

A seguir, do mesmo modo que na normalização da inclinação média dos caracteres da palavra é determinada a entropia associada a cada perfil. O ângulo (α) que proporciona a menor entropia será o ângulo de declive da palavra. Finalmente rotaciona-se os *pixels* da imagem original utilizando a transformação, descrita a seguir:

$$\begin{aligned} i' &= i + j \tan\alpha, \\ j' &= j. \end{aligned} \tag{3.7}$$

Nesta equação, α é o ângulo pelo qual deseja-se rotacionar os *pixels* da imagem. A Figura 3.4, mostra a representação gráfica do cálculo da nova coordenada (i', j'). O ponto A é *transportado* para o ponto B. Note que os pontos A e B pertencem à mesma coluna da imagem, ou seja, possuem o mesmo valor de coordenada j.

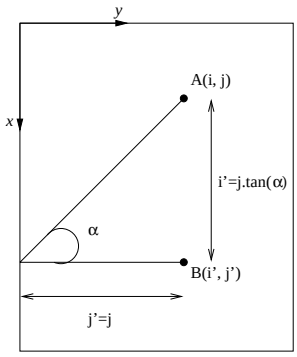


Figura 3.4: Representação gráfica do cálculo da nova coordenada (i', j') da imagem rotacionada.

3.2.3 Suavização

O algoritmo de suavização utilizado baseia-se no deslocamento de máscaras sobre a imagem. Estas máscaras, definidas por Veloso [28], são divididas em duas categorias: as que tratam com *pixels* isolados e as que tratam com mais de um *pixel*.

As máscaras utilizadas na primeira categoria são mostradas na Figura 3.5. Além dessas, são usadas outras 14 máscaras obtidas pelo espelhamento e rotacionamento das máscaras x_c e x_r de 90°, 180° e 270°. As máscaras utilizadas na segunda categoria são ilustradas na Figura 3.6. Em ambos os casos, o x pode representar tanto *pixels* pretos como *pixels* brancos, o número 1 representa os *pixels* pretos e o número 0 os *pixels* brancos. Quando ocorre o casamento entre qualquer uma dessas máscaras e uma janela da imagem, o elemento central tem seu valor modificado (0 para 1 ou 1 para 0).

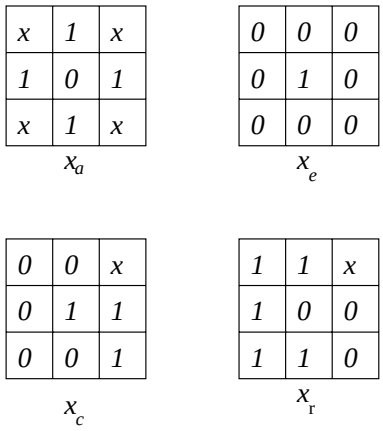


Figura 3.5: Máscaras utilizadas no processo de suavização - primeiro procedimento.

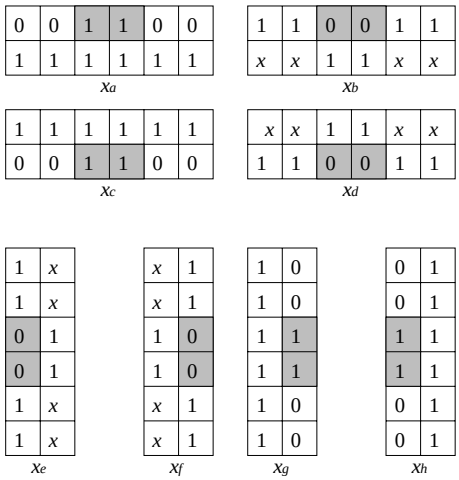
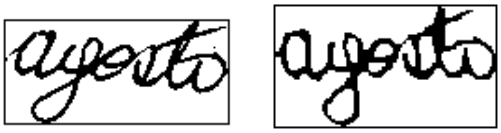


Figura 3.6: Máscaras utilizadas no processo de suavização - segundo procedimento.

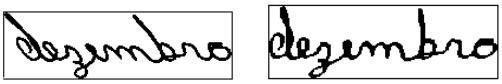
3.2.4 Análise dos resultados

A partir de uma análise visual subjetiva, constatou-se que 99% das imagens pré-processadas apresentaram bons resultados. Porém, como já previsto por Veloso [28], alguns problemas ocorrem no algoritmo de normalização do declive da palavra. Estes problemas são ocasionados por erros na detecção do contorno inferior da palavra, pois em alguns casos este contorno não contém apenas os *pixels* pertencentes à linha de base da palavra ou os *pixels* localizados próximos à esta linha. O método foi desenvolvido baseado na hipótese de que a linha de base da palavra era uma linha reta, porém em alguns casos essa suposição falha, ocasionando problemas. Exemplos do resultado do pré-processamento são apresentados nas Figuras 3.7 e 3.8, que mostram imagens em que o pré-processamento obteve bons resultados e na Figura 3.9, um caso onde ocorreu o problema acima descrito.



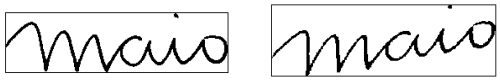
(a) (b)

Figura 3.7: Resultado do pré-processamento aplicado à palavra **agosto**. (a) imagem original e (b) imagem pré-processada.



(a) (b)

Figura 3.8: Resultado do pré-processamento aplicado à palavra **dezembro**. (a) imagem original e (b) imagem pré-processada.



(a) (b)

Figura 3.9: Resultado do pré-processamento aplicado à palavra **maio**. (a) imagem original e (b) imagem pré-processada.

3.3 Extração de características

A etapa de extração de características é vital em qualquer sistema de reconhecimento de padrões, pois quanto melhor a representação dos dados em análise, melhor será seu mapeamento pelo classificador. Portanto, a relação características-classificador é muito importante para o bom desempenho de qualquer sistema. Em função disso, iniciamos esta seção com algumas discussões sobre as limitações dos classificadores neurais em relação à representação das características.

Uma limitação que ocorre com os classificadores neurais é a necessidade de um vetor de entrada de tamanho fixo. Para resolver este problema, um dos caminhos mais fáceis seria uma normalização em escala, redimensionando as imagens para um mesmo tamanho, porém trabalhos anteriores mostraram que esse tipo de solução pode ocasionar grandes deformações nos padrões, o que comprometeria o sistema como um todo [29, 30].

Para resolver este problema, se optou por fazer uma segmentação implícita dividindo cada imagem em 8 sub-regiões de mesmo tamanho. Este número de sub-regiões corresponde ao número médio de letras presentes nas palavras que formam o léxico em análise. Para cada sub-região, são definidos 10 padrões denominados $x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8, x_9, x_{10}$. Deste modo é formado um vetor de características contendo 80 padrões, para cada imagem. Exemplo desse procedimento é mostrado na Figura 3.10 e a definição dos padrões é detalhada nas próximas seções.

Outra limitação do classificador neural é a necessidade de padrões de entrada normalizados [31]. Por este motivo, todos os componentes do vetor de características foram

Figura 3.10: Exemplo do processo de segmentação implícita utilizado.

normalizados no intervalo [0,1], em função da própria definição dos padrões como será mostrado nas próximas seções.

Resolvido os problemas devido às limitações do classificador, foram definidos três diferentes conjuntos de características denominados de características perceptivas, direcionais e topológicas, descritos a seguir.

3.3.1 Características perceptivas (P)

As características perceptivas são consideradas características de alto nível de acordo com a classificação de Madhavanath [8] e sua utilização é justificada pelo processo de leitura humano, que usa características como ascendentes, descendentes e estimação do comprimento da palavra para ler sentenças manuscritas.

O primeiro passo na extração de perceptivas é determinar as zonas da palavra, que são definidas por Freitas [1] do seguinte modo:

- Zona ascendente: compreendida entre o limite superior máximo (LSM) da palavra e o limite superior (LS) do corpo da palavra;
- Zona corpo da palavra: compreendida entre o limite superior (LS) e inferior (LI) do corpo da palavra;
- Zona descendente: compreendida entre o limite inferior (LI) do corpo da palavra e o limite inferior mínimo (LIM) da palavra.

Para determinar estas zonas, inicialmente é determinado o histograma de projeção horizontal das transições branco-preto da palavra (HT). A linha com valor de histograma máximo é denominada linha média (LM). Em seguida, um procedimento de

suavização é aplicado para eliminar as descontinuidades do histograma, de acordo com a Equação 3.8:

$$P'_x = \frac{P_{x-2} + P_{x-1} + P_x + P_{x+1} + P_{x+2}}{5} \tag{3.8}$$

em que P_y é o valor do histograma na posição y e P'_y é o novo valor do histograma, obtido após a suavização, na posição y.

A partir do histograma suavizado, as linhas superior (LS) e inferior (LI) são aquelas acima e abaixo da linha média (LM), respectivamente, com valor igual a 70% do valor máximo do histograma. Este percentual foi obtido heuristicamente por Freitas [1] baseado no estudo da diferença entre os picos do histograma de transição branco-preto e do histograma de densidade de *pixels* para um conjunto de imagens da sua base de treinamento. Exemplos desse processo são apresentados na Figura 3.11.

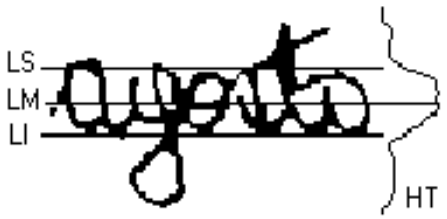


Figura 3.11: Exemplo do processo de detecção das zonas da palavra.

Uma vez determinada as zonas da palavra e após o processo de segmentação implícita, são extraídos os 10 padrões de cada uma das 8 sub-regiões, que são definidos a seguir:

- Posição do maior ascendente: Inicialmente é feita a rotulação dos ascendentes a partir de uma adaptação do algoritmo para rotulação de ilhas descrito por Veloso [28], que o utiliza para segmentação de palavras. Em resumo, este algoritmo faz uma busca dos *pixels* pretos da imagem analisando a sua vizinhança. Se os *pixels* da vizinhança não tiverem rótulos, o *pixel* em análise é rotulado, caso contrário ele recebe o mesmo rótulo da vizinhança. Em seguida é determinada a posição do *pixel* central do maior ascendente, que é normalizada pela largura

da sub-região. Esta posição é determinada calculando a posição média entre os pontos extremos do ascendente em relação à horizontal;

- Tamanho do maior ascendente: É determinada a altura do maior ascendente, obtida pela diferença entre as coordenadas do *pixel* inferior mínimo e do *pixel* superior máximo, que é normalizada pela altura do corpo da palavra;

- Posição e tamanho do maior descendente: Mesmas definições usadas para os ascendentes, considerando a zona descendente da palavra;

- Tamanho do laço: Faz-se a contagem do número de *pixels* interiores ao laço, normalizando pela área da sub-região. Um laço é definido como a região em que a partir de um *pixel* interno independente da direção de busca sempre se encontra um *pixel* preto;

- Localização do laço: É dada pelas coordenadas do centro de massa do laço, que são definidas de acordo com a Equação 3.9:

$$(X_{cm}, Y_{cm}) = (\frac{\sum_{i=1}^{M}\sum_{j=1}^{N} i.f_{ij}}{M}, \frac{\sum_{i=1}^{M}\sum_{j=1}^{N} j.f_{ij}}{N}), \tag{3.9}$$

  em que X_{cm} e Y_{cm} são as coordenadas do centro de massa, M e N as dimensões da imagem e f_{ij} assume o valor 0 quando o *pixel* na posição (i, j) é branco e o valor 1 no caso contrário.

  As coordenadas X_{cm} e Y_{cm} são normalizadas pela largura e altura da sub-região, respectivamente;

- Concavidades: Inicialmente, são extraídos os pontos extremos do contorno externo da forma em análise. Em seguida, os ângulos definidos por dois segmentos de reta, traçados entre o ponto inferior e os pontos mais à direita e mais à esquerda do contorno, em relação à horizontal são medidos. Estes ângulos são normalizados por 90°.

- Comprimento estimado da palavra: Determina-se o número de transições (branco-preto) presentes na linha média da sub-região em análise. Este valor

é então normalizado pelo número total de transições presentes na linha média da palavra. Uma transição é definida como qualquer mudança branco-preto ou preto-branco desde que fora de laços.

Quando um padrão não ocorre em uma sub-região é necessário atribuir um valor que represente a sua ausência, o mais simples seria atribuir $0,0$, porém uma grande quantidade de padrões nulos na entrada da rede afetam o seu desempenho, deste modo preferiu-se atribuir o valor unitário $1,0$.

3.3.2 Características direcionais (D)

As características direcionais podem ser consideradas características de nível intermediário, contendo informações relevantes sobre a região do fundo da imagem. Neste trabalho, as características direcionais definidas foram inspiradas num procedimento de rotulação proposto por Parker [32]. Neste método, para cada *pixel* do fundo da imagem é verificado para cada uma das quatro direções principais (Norte, Sul, Leste e Oeste) se um *pixel* preto pode ser encontrado, como mostra a Figura 3.12.

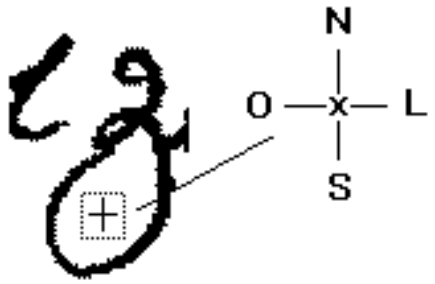


Figura 3.12: Exemplo da detecção das direções de abertura.

Em função deste teste e dependendo da combinação das direções de abertura, os *pixels* do fundo são rotulados pela convenção descrita na Tabela 3.2.

O rótulo 9 é usado para representar caracteres sem traços de ligação. Os componentes do vetor de características para cada sub-região são obtidos contando o número de *pixels* atribuídos à cada rótulo, normalizados pela área da sub-região. Quando não existe *pixels* de um determinado rótulo o valor mapeado para o vetor é $1,0$.

Tabela 3.2: Convenção usada para rotulação de *pixels* no conjunto de características direcionais.

Rótulo	Tipo
0	Fechado
1	Aberto abaixo
2	Aberto acima
3	Aberto à direita
4	Aberto à esquerda
5	Aberto à direita e acima
6	Aberto à esquerda e acima
7	Aberto à esquerda e abaixo
8	Aberto à direita e abaixo
9	Aberto abaixo e acima

3.3.3 Características topológicas (T)

Características topológicas refletem a densidade de *pixels* em diversas regiões da imagem, sendo classificadas como características de baixo nível. Para determinar estas características é feito um zoneamento, dividindo cada sub-região em duas partes, acima e abaixo da linha média da palavra. Depois disto, as partes superior e inferior são divididas em 4 zonas cada uma, como mostra a Figura 3.13.

As componentes do vetor de características $(x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8)$ são obtidas contando o número de *pixels* pretos em cada uma das oito zonas, normalizados pela respectiva área da zona. As componentes (x_9, x_{10}) correspondem às coordenadas do centro de massa do segmento a imagem em cada sub-região, normalizadas pela largura e altura da sub-região, respectivamente. Quando o número de *pixels* pretos é zero, o valor mapeado para o vetor é $1,0$.

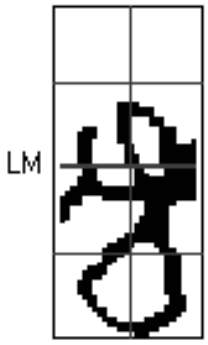


Figura 3.13: Exemplo da divisão em zonas realizada no conjunto de características topológicas.

3.4 Classificador neural

O classificador neural foi escolhido para o desenvolvimento deste sistema pois o dicionário utilizado (meses do ano) apresenta um número pequeno e limitado de classes. Além disso, este classificador tem sido pouco explorado na tarefa de reconhecimento de palavras manuscritas.

Os classificadores neurais se adequam bem a problemas de alta dimensionalidade e que possuam interações complexas entre suas variáveis, o que sugere sua aplicação no reconhecimento de palavras. As principais características desses classificadores são o processo de cálculo, que é inerentemente paralelo, a possibilidade de implementação em *hardware* e a abstração do processo de aprendizagem humano [31].

3.4.1 Redes neurais

A possibilidade de dar às máquinas a habilidade que o ser humano possui de aprender, reter informações, recordá-las, e aplicá-las na solução de diversos tipos de problemas, tem entusiasmado muitos pesquisadores a procurar desenvolver modelos computacionais para os mesmos fins [33]. Uma classe dentre estes modelos é a dos chamados sistemas neurais artificiais ou simplesmente redes neurais.

O que é comumente chamado de redes neurais é um conjunto interconectado de

elementos de processamento (PE), denominados de neurônios, células ou nós, cada qual realizando um simples cálculo. O modelo do neurônio apresentado na Figura 3.14 possui várias entradas (conjunto de *sinapse*), a cada uma das quais é associado um peso, e uma saída, que pode ser usada como entrada de outros elementos de processamento. O valor associado a qualquer neurônio é chamado de sua ativação (net) e representa a soma ponderada das entradas. Ou seja, para um neurônio k:

$$net_k = \sum_{j=1}^{N} x_j w_{kj}, \tag{3.10}$$

em que N é o número de entradas do neurônio, x_j são as entradas do neurônio e w_{kj} são os pesos sinápticos associados a cada entrada.

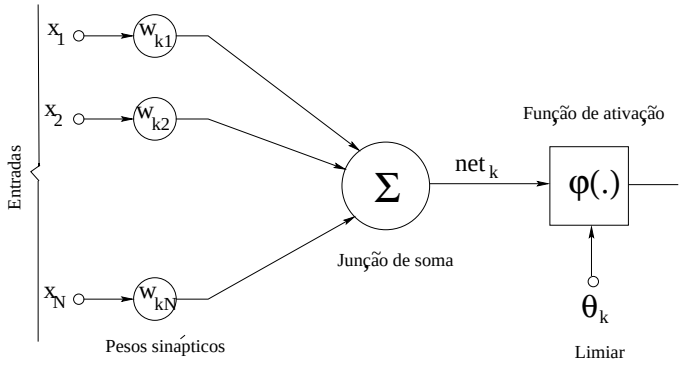


Figura 3.14: Modelo do neurônio utilizado em redes neurais.

A saída de um elemento de processamento pode ser simplesmente o seu valor de ativação. Entretanto, na maioria das redes neurais a saída de um neurônio é dada por uma função de ativação expressa como:

$$y = \varphi(net_k - \theta_k), \tag{3.11}$$

em que θ_k é um valor de limiar.

A função de ativação $\varphi(.)$ garante que o valor de saída do elemento de processamento encontra-se dentro de uma faixa pré-definida. Vários tipos de funções de ativação são usadas para ativar um neurônio artificial, porém o uso particular depende do tipo de dados de saída (contínuo ou discreto) e da faixa de valores assumidos por estes dados

(por exemplo, de -1 a 1). As funções de ativação comumente encontradas na literatura são:

a) Função linear:

$$\varphi(x) = x; \tag{3.12}$$

b) Função degrau unitário:

$$\varphi(x) = \begin{cases} 1, \text{se x} \geq 0 \\ 0, \text{se x} < 0 \end{cases}; \tag{3.13}$$

c) Função bipolar:

$$\varphi(x) = \begin{cases} 1, \text{se x} \geq 0 \\ -1, \text{se x} < 0 \end{cases}; \tag{3.14}$$

d) Função tangente hiperbólica:

$$\varphi(x) = tanh(x); \tag{3.15}$$

e) Função sigmoidal:

$$\varphi(x) = \frac{1}{1 + e^{-ax}}. \tag{3.16}$$

Antes de uma rede neural executar determinada tarefa é necessário que ela seja treinada. O treinamento ou a aprendizagem, no sentido de redes neurais, significa determinar os pesos sinápticos para cada elemento de processamento, através de algoritmos de treinamento. O treinamento de uma rede neural consiste na apresentação de um conjunto de treinamento com propriedades desconhecidas à entrada da rede e em ajustar os seus pesos sinápticos até obter a saída desejada. Este processo é repetido diversas vezes com diferentes classes de dados até que os pesos sinápticos encontrem-se estabilizados. Neste ponto, o processo de aprendizagem fica completo e a rede pode ser usada para classificar as entradas [29].

3.4.2 Caracterização do classificador utilizado

Neste sistema em especial foram utilizadas as redes neurais multicamadas treinadas com o algoritmo de retropropagação do erro, que estão entre os mais difundidos e versáteis modelos de classificadores. Diversos autores mostram que este tipo de rede neural contendo uma camada intermediária e uma função de ativação não-linear é um classificador universal [34, 35]. Isto é, tais redes podem determinar limiares de decisão de complexidade arbitrária.

Tipicamente, a arquitetura de uma rede neural multicamadas consiste de um conjunto de neurônios que constituem a camada de entrada, uma ou mais camadas escondidas e a camada de saída. A Figura 3.15 apresenta um exemplo da rede neural contendo três camadas.

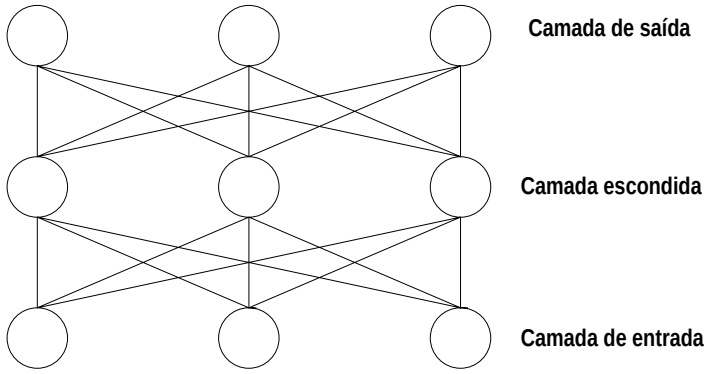


Figura 3.15: Arquitetura de uma rede neural com três camadas.

O número de neurônios na camada de entrada é determinada pela dimensão do vetor de características. A quantidade de neurônios na saída é determinada pelo número de classes existentes. A quantidade de camadas escondidas e o número de neurônios em cada uma destas camadas são determinadas pelo projetista da rede, em geral empiricamente. No sistema em questão a rede neural é composta por 80 neurônios na camada de entrada, uma camada escondida de tamanho variável e 12 neurônios na camada de saída. Como função de ativação foi empregada a função sigmoidal expressa

pela Equação 3.17,

$$\psi(net_i) = \frac{1}{1 + e^{net_i}} \tag{3.17}$$

que permite uma aproximação probabilística da saída da rede.

No treinamento da rede foi utilizado o algoritmo de retropropagação do erro com momento, que consiste dos seguintes passos:

1. Inicializar os pesos sinápticos e limiares (*thresholds*). Os pesos sinápticos da rede e os limiares devem ser inicializados com pequenos números aleatórios, com o intuito de prevenir, por exemplo, que a rede fique saturada com grandes valores de peso.

2. Apresentar os valores das entradas e das saídas desejadas.

3. Ativar a rede para produzir as saídas.

4. Calcular o erro entre a saída produzida pela rede e a saída desejada. Esta função de erro (E_p) é definida como sendo proporcional ao erro quadrático entre a saída atual e a saída desejada, para todos os padrões a serem treinados.

5. Ajustar os pesos sinápticos da rede visando minimizar o erro, de acordo com a seguinte equação:

$$w_{ij}(n+1) = w_{ij}(n) + \eta\delta_j O_j + \alpha[w_{ij}(n) - w_{ij}(n-1)], \tag{3.18}$$

em que η representa o termo de ganho, δ_j representa o gradiente local da rede, O_j representa a saída atual do j-ésimo neurônio e α representa o momento. Para a camada de saída, o gradiente é dado por

$$\delta_j = O_j(1 - O_j)(t_j - O_j), \tag{3.19}$$

em que t_j representa a saída desejada do neurônio j. E para a camada escondida, tem-se

$$\delta_j = O_j(1 - O_j)\sum_k \delta_k w_{kj}. \tag{3.20}$$

6. Repetir os passos 2 à 6 até que o critério de parada estabelecido seja satisfeito. Como foi utilizado treinamento seguido de validação para evitar problemas de super aprendizagem (quando a rede passa a *decorar* os padrões), o critério de parada utilizado foi o erro obtido no conjunto de validação.

O parâmetro de aprendizagem determina a variação do ajuste dos pesos sinápticos da rede. Para um pequeno valor de η o ajuste é lento e a rede demora para convergir. Por outro lado, aumentando o valor de η demasiadamente, pode provocar instabilidade na rede, uma vez que o ajuste é feito bruscamente. Dessa forma, o momento α foi adicionado visando aumentar a convergência, sem no entanto, tornar a saída da rede instável [33].

3.5 Conclusão

Neste capítulo, foi descrito o sistema de referência desenvolvido para a avaliação de características proposta nesta dissertação. Este é composto por três estágios: pré-processamento, extração de características e classificação por rede neural multicamadas. No capítulo seguinte, os testes efetuados e os resultados obtidos serão apresentados.

Capítulo 4

Testes Efetuados e Resultados Obtidos

Neste capítulo, os conjuntos de características descritos anteriormente são avaliados considerando os resultados experimentais obtidos usando o sistema de referência. Os testes foram efetuados utilizando os conjuntos tanto de forma isolada como combinados de acordo com diversas estratégias de combinação. Também foi feita uma avaliação comparativa do sistema proposto neste trabalho em relação ao sistema desenvolvido no LARDOC/PUC-PR pela Profa. Cinthia Freitas em seu trabalho de tese [1], sendo avaliadas diferentes combinações dos dois sistemas. As avaliações são feitas considerando a taxa de reconhecimento obtida e a análise dos erros encontrados.

As técnicas de pré-processamento e de extração de características foram implementadas em linguagem C ANSI e a rede neural foi desenvolvida no ambiente de simulação criado por pesquisadores da Universidade de Stuttgart [36].

A seguir, são apresentados os resultados dos testes efetuados com o sistema de referência, com o sistema de Freitas [1] e por fim uma comparação com outros resultados descritos na literatura.

4.1 Testes efetuados com o sistema de referência

4.1.1 Análise dos conjuntos isolados

Para cada conjunto de características apresentado na Seção 3.3 é treinada uma rede neural distinta. A classe que apresenta o máximo valor de saída da rede é a classe reconhecida. Como a definição do número de neurônios na camada escondida é empírica, foram testadas diversas configurações. Os melhores resultados obtidos nos conjuntos de características perceptivas (*RN-P*), direcionais (*RN-D*) e topológicas (*RN-T*) utilizaram 75, 80 e 85 neurônios na camada escondida, respectivamente. Também foram avaliados diversos valores para os parâmetros η e α, que são a taxa de aprendizagem e o momento, respectivamente. Os melhores desempenhos da rede foram obtidos considerando $\eta = 0,01$ e $\alpha = 0,3$.

O primeiro parâmetro de desempenho utilizado para avaliar o sistema foi a taxa de reconhecimento, que representa o percentual de palavras classificadas corretamente. A Tabela 4.1 mostra a taxa de reconhecimento obtida pelo sistema considerando cada conjunto isoladamente. O resultado mostra que o conjunto que apresenta melhores resultados é o *RN-P*. Embora o conjunto *RN-D* apresente percentuais similares para algumas classes, na média o conjunto *RN-P* se comporta melhor. O conjunto *RN-T* na média tem o pior desempenho, com destaque para os maus resultados obtidos para as classes *Janeiro* e *Maio*.

O desempenho do sistema também pode ser avaliado por meio de uma matriz de classificação, denominada matriz de confusão. Nesta matriz, cada linha e cada coluna corresponde a uma classe. A entrada da matriz para a linha A e coluna B fornece o número de elementos da classe A que foram classificados como pertencentes à classe B. A distribuição dos dados na matriz de confusão fornece informações valiosas que podem ser usadas para avaliar o comportamento do sistema. As Tabelas 4.2, 4.3 e 4.4 mostram as matrizes de confusão obtidas para cada conjunto.

Tabela 4.1: Taxa de reconhecimento média obtida por classe para cada conjunto de características.

Conjunto	Perceptivas (P)	Direcionais (D)	Topológicas (T)
Janeiro	**75** %	74 %	53 %
Fevereiro	**77** %	77 %	71 %
Março	**84** %	86 %	82 %
Abril	**88** %	89 %	87 %
Maio	**82** %	78 %	56 %
Junho	**82** %	69 %	72 %
Julho	**87** %	65 %	79 %
Agosto	**88** %	79 %	74 %
Setembro	**76** %	64 %	76 %
Outubro	**85** %	79 %	81 %
Novembro	**82** %	86 %	67 %
Dezembro	**76** %	74 %	78 %
Média	**81,8** %	76,6 %	73,0 %

Para facilitar a análise das matrizes de confusão, as classes são agrupadas em superclasses definidas pela presença de terminações ou letras em posições comuns:

- Jan**eiro**-Fever**eiro**: Os resultados mostram que a maior parte das confusões encontradas nessas classes ocorre entre elas mesmas. Isto já era esperado pois elas apresentam descendentes na mesma posição, dificultando a tarefa de classificação. Contudo, para o conjunto *RN-T*, a palavra *Janeiro* produziu uma confusão considerável com outras classes, como *Junho*, por exemplo.

- **Ma**rç**o-Ma**i**o**: Apesar da similaridade existente entre estas palavras, a confusão entre elas é relativamente pequena, talvez por causa da presença do descendente em *Março*. Porém, considerando os conjuntos *RN-D* e *RN-T* percebe-se uma confusão considerável da palavra *Maio* com outras classes, consequência possivelmente do procedimento de segmentação utilizado que super segmenta esta palavra.

- **Junho-Julho**: As principais confusões para estas classes ocorrem entre elas mesmas, como esperado já que elas são muito similares. Como no caso anterior, isso pode decorrer do processo de segmentação, pois dependendo do seu resultado os ascendentes correspondentes às letras *lh* da palavra *Julho* podem se localizar na mesma sub-região, dificultando assim a separação das classes. Contudo para o conjunto *RN-T*, uma grande confusão também é observada entre as palavras *Junho* e *Janeiro*.

- **A**bril-**A**gosto: Estas palavras não têm similaridades, portanto a ocorrência de confusões entre elas é pequena. A palavra *Abril*, não mostra confusão com nenhuma classe em especial. Para *Agosto* porém, um alto nível de confusão com *Dezembro* pode ser observado para os conjuntos *RN-D* e *RN-T*, causados talvez pela localização próxima dos ascendentes/descendentes, ou pelo processo de segmentação.

- Set**embro**-Outu**bro**-Nov**embro**-Dez**embro**: As principais confusões dessas palavras são encontradas em sua própria super-classe, principalmente entre *Setembro* e *Outubro* para o conjunto *RN-D* e *RN-T*.

Esta análise mostra que os principais problemas do sistema ocorrem geralmente entre classes que pertencem a uma mesma super-classe, ou seja, classes que apresentam terminações ou letras em posições comuns. Para melhorar os resultados uma seleção criteriosa de características discriminantes dentro de cada super-classe se torna necessária. Os resultados mostram que os conjuntos definidos são válidos, mas o conjunto de características perceptivas apresentou melhores resultados em relação aos outros. Este fato mostra que a incorporação do conhecimento de leitura humano é significativo para a obtenção de bons resultados no reconhecimento de palavras manuscritas. A seguir, os resultados dos conjuntos de características combinados são apresentados.

Tabela 4.2: Matriz de confusão para o conjunto RN-P.

Mês	J	F	M	A	M	J	J	A	S	O	N	D
Janeiro	**75**	10	2	1	2	4	4				2	
Fevereiro	7	**77**	1	1	5	1	2			1	4	1
Março		1	**84**	5	4	1	1	2			1	1
Abril		1	3	**88**	2	1		2	1	1	1	
Maio	2	1		2	**82**	6	1	2	2	2		
Junho	3		1	1	1	**82**	4	2	2	2	1	1
Julho				2	3	6	**87**	1	1			
Agosto	1		2	2	1		2	**88**			1	3
Setembro	1	3	1	1	1	4			**76**	6	4	3
Outubro		3		3		3	1		5	**85**		
Novembro	1	1	2	3		2			6	1	**82**	2
Dezembro	3	2		1	1	3	1	3	1	5	4	**76**

Tabela 4.3: Matriz de confusão para o conjunto RN-D.

Mês	J	F	M	A	M	J	J	A	S	O	N	D
Janeiro	**74**	10		1	2	5	2		2	3		1
Fevereiro	8	**77**	4	1	3	2	1	2			2	
Março		2	**86**	1	3	2		3		1		2
Abril		1	1	**89**	5			1	1	1	1	
Maio	1		2	2	**78**	3	1	2	2	5	2	2
Junho	5		1	1		**69**	14	1	2	2	4	1
Julho	1	1	1		3	17	**65**	3	1	5	2	1
Agosto	2	3		3	1		2	**79**			1	9
Setembro	1	3		1	1	3		1	**64**	19	6	1
Outubro		2	1		2		1		13	**79**	2	
Novembro	2	1		2		3		1	4		**86**	1
Dezembro	4	3			1			8	3	5	2	**74**

Tabela 4.4: Matriz de confusão para o conjunto RN-T.

Mês	J	F	M	A	M	J	J	A	S	O	N	D
Janeiro	**53**	14	2	1	5	11	5		3		1	5
Fevereiro	7	**71**	6	1		4	1	2	1	3	2	2
Março	1	2	**82**	3	3	2	1	3				3
Abril	2	2	1	**87**	3	1			2	1	1	
Maio	4		2	8	**56**	2	3	3	7	6	8	1
Junho	8	1		1	1	**72**	4	1	5	1	5	1
Julho	1		1	2	1	10	**79**	2	2	1		1
Agosto		3	1	2	1	3		**74**		1	4	11
Setembro	1	2		1		3			**76**	7	9	1
Outubro	1		1			1	1	1	8	**81**	5	1
Novembro	4	5	1		2	4		1	9	2	**67**	5
Dezembro	1	1	1	1	2	2	2	5	2		5	**78**

4.1.2 Análise da combinação de conjuntos

No intuito de superar os problemas obtidos com os conjuntos isolados, foram definidas três estratégias de combinação dois-a-dois dos classificadores que utilizam estes conjuntos, procurando assim avaliar indiretamente o seu potencial quando combinados. Temos portanto a união de classificadores homogêneos, ou seja, classificadores que possuem a mesma metodologia de classificação mas diferentes vetores de características. Para facilitar a descrição, são feitas inicialmente algumas definições: Considere $\Psi = \{1, 2, 3, ..., V\}$ um conjunto de classes de palavras, F o vetor de características do padrão desconhecido a ser classificado e $f(n_v, F_i)$ denotando a saída do neurônio n_v associado com a classe v dado o vetor de características F_i.

A partir disso, descreve-se a seguir as estratégias de combinação:

- **Média Aritmética** - O valor atribuído pelo classificador a cada classe, é obtido pela média aritmética dos valores de saída dos neurônios correspondentes dos classificadores homogêneos. A classe que produz a maior média é a reconhecida.

Isto é representado na Equação 4.1:

$$v^* = argmax_{v \in \Psi}(\frac{f(n_v, F_i) + f(n_v, F_j)}{2}) \quad (4.1)$$

em que v^* é a classe reconhecida e F_i, F_j são conjuntos de características diferentes.

- **Multiplicação** - A palavra é reconhecida considerando-se o produto das saídas dos neurônios, de acordo com a Equação 4.2:

$$v^* = argmax_{v \in \Psi}(f(n_v, F_i) \times f(n_v, F_j)) \quad (4.2)$$

- **Agrupamento Neural** - Nesta estratégia, ao contrário das outras que utilizam as redes individuais treinadas, é construída uma nova rede tendo como entrada a concatenação de conjuntos diferentes. De modo que a entrada da nova rede é constituída de (80+80) = 160 neurônios.

A taxa média de reconhecimento obtida considerando cada uma das estratégias descritas anteriormente é mostrada na Tabela 4.5. Os melhores resultados foram obtidos para a combinação por multiplicação dos conjuntos *RN-P* e *RN-D*, onde ocorreu um aumento da taxa média de reconhecimento. Esses resultados sugerem que os conjuntos são complementares. A análise da tabela mostra também que para cada par de conjuntos o melhor desempenho ocorreu para estratégias de combinação distintas. Deste modo, não se pode afirmar qual a melhor dentre elas, pois os resultados apontam que não existe uma regra única para todos os conjuntos.

As Tabelas 4.6, 4.7 e 4.8 apresentam as matrizes de confusão considerando a melhor combinação para cada par de conjuntos. A análise dos erros permite observar que a confusão fora de super-classes diminui consideravelmente, mostrando que a combinação de conjuntos permite uma melhor representação das classes, embora os problemas internos às super-classes descritos na seção anterior ainda persistam.

Uma vez avaliados os conjuntos com o sistema de referência, buscando uma validação dos resultados obtidos, utiliza-se a base de dados em outro sistema com o intuito de comparar os desempenhos. Esses resultados são mostrados a seguir.

Tabela 4.5: Taxa de reconhecimento média obtida usando diferentes estratégias de combinação dos conjuntos.

	Tipo de Combinação		
Conjuntos	Média aritmética	Multiplicação	Agrupamento neural
P e D	**87,2**%	87,0%	86,0%
P e T	85,2%	**85,7**%	84,2%
D e T	81,6%	82,2%	**82,6**%

Tabela 4.6: Matriz de confusão para a melhor combinação dos conjuntos RN-P e RN-D.

Mês	J	F	M	A	M	J	J	A	S	O	N	D
Janeiro	**84**	5	1	2	1	3	2			1		1
Fevereiro	11	**80**	1	1	4		1	1			1	
Março		2	**92**	1	3			1			1	
Abril			1	**96**	3							
Maio	1				**92**	3		1	2	1		
Junho	5			1		**79**	7	1	1	2	3	1
Julho					1	9	**88**	1				1
Agosto	2	1	1	2				**90**				4
Setembro	1	2		1	1	2			**81**	7	4	1
Outubro		1		1			1		5	**92**		
Novembro	2			1				1	5		**91**	
Dezembro		4			1			4		5	4	**82**

Tabela 4.7: Matriz de confusão para a melhor combinação dos conjuntos RN-P e RN-T.

Mês	J	F	M	A	M	J	J	A	S	O	N	D
Janeiro	**73**	9	1	2	1	7	3		2			2
Fevereiro	6	**81**	2		3	4			1	1	1	1
Março		2	**91**	1	4				1			1
Abril				**98**	1	1						
Maio	1			1	**84**	3	2	2	4	2	1	
Junho	2		1	1	1	**86**	3		4	1		1
Julho	3			1	1	5	**89**		1			
Agosto	1	1	1	4			1	**85**				7
Setembro	1	2			1	3			**86**	2	3	2
Outubro		3							5	**91**		1
Novembro	3	3	1						7	1	**81**	4
Dezembro	1	4			1	1		4	3		3	**83**

Tabela 4.8: Matriz de confusão para a melhor combinação dos conjuntos RN-D e RN-T.

Mês	J	F	M	A	M	J	J	A	S	O	N	D
Janeiro	**76**	8			2	5	4	1	1	1		2
Fevereiro	5	**83**	3	1	1	1		3		1	1	1
Março	1	1	**88**	2	3			3				2
Abril	1	2	1	**94**		2						
Maio	2	1	2	1	**83**	2	1	2	2	1	2	1
Junho	6					**82**	3	1	3		4	1
Julho	1		1	1	1	11	**81**	1	1			2
Agosto	2	3	1	1		2		**83**		1		7
Setembro	1	4		1	1	4			**77**	8	4	
Outubro	1	1	1		1	1	1		6	**81**	4	3
Novembro	2	3			2	5			3	1	**83**	1
Dezembro	3	2				1		7	2	1	4	**80**

4.2 Testes efetuados com o sistema de Freitas [1] e abordagens híbridas

A avaliação comparativa foi feita em parceria com a Profa. Cinthia Freitas do LARDOC/PUC-PR que desenvolveu um sistema para reconhecimento do extenso de cheques bancários brasileiros utilizando Modelos Escondidos de Markov (*MEM*). Os *MEM* possuem a capacidade de oferecer um modelo probabilístico à um enfoque estrutural, como normalmente utilizado no reconhecimento de palavras manuscritas, além de possibilitar o modelamento eficaz de diferentes fontes de conhecimento, tanto à nível sintático (modelo para cada palavra) quanto morfológico (forma a reconhecer) [1].

Este sistema utiliza um conjunto de características, descrito anteriormente na revisão bibliográfica [1], que é bastante similar ao conjunto de características perceptivas definido neste trabalho. A presença das características é codificada por símbolos, que formam conjuntos de observações de tamanho variável. A partir desses conjuntos é gerado um modelo para cada palavra pertencente ao dicionário. O treinamento dos modelos é feito utilizando o algoritmo de *Baum-Welch*, juntamente com um procedimento de validação cruzada para determinar o modelo ótimo. Uma vez os modelos treinados, na fase de teste cada conjunto de observações é apresentado ao modelo de cada palavra sendo determinada uma probabilidade condicional $P(O|\lambda)$, por meio do algoritmo *Forward*, em que λ representa o modelo em questão. O modelo que tiver a maior probabilidade associada, representa a palavra reconhecida [37].

A adaptação do sistema de Freitas [1] ao dicionário utilizado neste trabalho foi simples. As imagens utilizadas na entrada do sistema seguiram o mesmo procedimento de pré-processamento descrito no capítulo anterior. A taxa média de reconhecimento obtida foi de 75,9% e a matriz de confusão é mostrada na Tabela 4.9.

A análise da matriz aponta resultados semelhantes aos obtidos com o conjunto *RN-P* (Tabela 4.2), com algumas diferenças: ocorrência de alta confusão da classe *Abril* com *Janeiro*; a palavra *Maio* apresenta uma confusão considerável com *Janeiro* e *Julho*; *Setembro* apresenta um aumento das confusões dentro da sua super-classe e um erro

Tabela 4.9: Matriz de confusão obtida pelo sistema de Freitas [1].

Mês	J	F	M	A	M	J	J	A	S	O	N	D
Janeiro	**72**	13	1	1	5	2	2		2		1	1
Fevereiro	9	**75**		3	1	5	3				2	2
Março	3	3	**80**	3	6		1		3		1	
Abril	14			**82**	1		3					
Maio	10	1	6	2	**67**		10	1	1	1	1	
Junho	3	2		4	3	**75**	3	2	3	1		4
Julho	4	2		7	1	4	**80**	1		1		
Agosto	4			5			1	**80**		2		8
Setembro	1	1	1	5	2	5	4		**61**	8	4	8
Outubro	1	1		2			2	1	5	**87**	1	
Novembro	1	1		5	1				17	1	**70**	4
Dezembro	1	1		1	1	3	2	4	3	1	1	**82**

considerável ocorre entre *Novembro* e *Setembro*. Por outro lado, a palavra *Dezembro* apresenta melhor desempenho, diminuindo a confusão na sua super-classe.

Estes resultados mostram que para o problema em questão, o classificador neural apresenta um melhor mapeamento entrada-saída que os Modelos Escondidos de Markov. Para complementar a avaliação, foram efetuados testes considerando combinações dos dois sistemas. Como a probabilidade $P(O|\lambda)$ assume valores muitos pequenos, foi feita uma normalização prévia, de acordo com a Equação 4.3:

$$P^*(O|\lambda_i) = \frac{P(O|\lambda_i)}{\sum_j P(O|\lambda_j)} \tag{4.3}$$

em que $P^*(O|\lambda_i)$ é a probabilidade normalizada para o modelo λ_i.

A combinação foi realizada de duas maneiras:

- **Média Aritmética** - A palavra é reconhecida considerando-se a média arimética das saídas dos neurônios com a probabilidade normalizada de cada classe, de acordo com a Equação 4.4:

$$v^* = argmax_{v \in \Psi} \left(\frac{f(n_v, F_i) + f(n_v, F_j) + P^*(O|\lambda_v)}{3} \right) \tag{4.4}$$

em que v^* é a classe reconhecida e F_i, F_j são conjuntos de características diferentes.

- **Multiplicação** - A palavra é reconhecida considerando-se o produto das saídas dos neurônios pelas probabilidades normalizadas, de acordo com a Equação 4.5:

$$v^* = argmax_{v \in \Psi} \left(f(n_v, F_i) \times f(n_v, F_j) \times P^*(O|\lambda_v) \right) \tag{4.5}$$

Os testes efetuados consideraram combinações do sistema de Freitas [1] (*MEM*) com todos os classificadores testados anteriormente. A Tabela 4.10 descreve as taxas de reconhecimento médias obtidas para diferentes combinações de *MEM* e *RNs*. O melhor resultado ocorreu para a combinação por multiplicação de *MEM* com os conjuntos *RN-P* e *RN-D*. A matriz de confusão considerando este resultado é mostrado na Tabela 4.11.

Tabela 4.10: Taxa de reconhecimento média obtida usando diferentes combinações de MEM e RNs .

Combinação	Média Aritmética (%)	Multiplicação (%)
MEM e RN-P	88,1	88,7
MEM e RN-D	87,0	87,6
MEM e RN-T	85,8	86,2
MEM, RN-P e RN-D	90,2	**90,4**
MEM, RN-P e RN-T	89,6	89,9
MEM, RN-D e RN-T	87,6	89,0

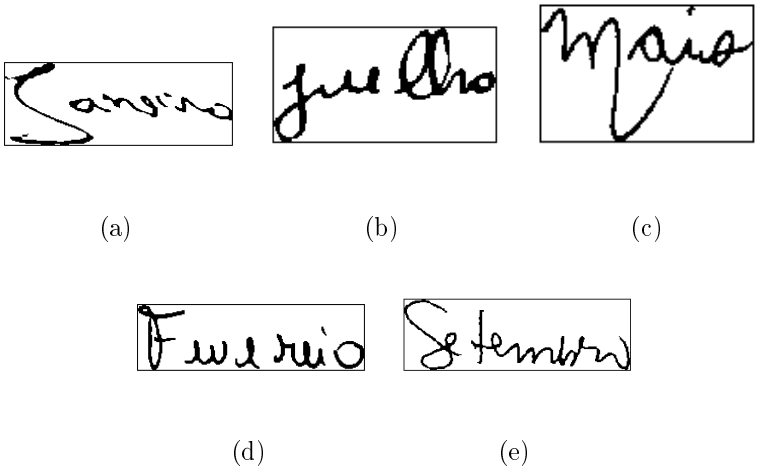
Comparando a matriz apresentada na Tabela 4.11 com a obtida para a combinação RN-P e RN-D (Tabela 4.6), pode-se ver uma melhora no reconhecimento da maioria das classes, havendo um aumento para as classes: *Fevereiro*, *Junho*, *Setembro*, *Outubro* e *Dezembro*. Estes resultados mostram que classificadores diferentes quando unidos numa abordagem híbrida podem fornecer resultados melhores do que aqueles obtidos ao considerá-los isoladamente.

Tabela 4.11: Matriz de confusão para a combinação *MEM*, *RN-P* e *RN-D*.

Mês	J	F	M	A	M	J	J	A	S	O	N	D
Janeiro	**86**	7			1	3	2		1			
Fevereiro	7	**87**	1	1	1	1	1			1		
Março		1	**92**	1	3		1	1			1	
Abril		1		**96**	2					1		
Maio				2	**91**	3	2	2				
Junho	1				2	**94**	1	1		1		
Julho	1			1	3	7	**88**					
Agosto	2		1	3	1	1		**91**				1
Setembro	1	2				1			**86**	9	1	
Outubro							1		3	**96**		
Novembro	1	1	1	1		1			5		**89**	1
Dezembro		1		1	2	1		2	2	2		**89**

Buscando uma melhor compreensão dos erros obtidos foi feita uma análise visual subjetiva das imagens correspondentes aos erros apontados na Tabela 4.11. Algumas conclusões importantes foram retiradas desta análise: os classificadores mapeam com muita fidelidade a presença e posição de ascendentes/descendentes e laços, portanto imagens com problemas na representação dessas características, como a ausência de laços no corpo da palavra, prejudicam a classificação. Considerando os estilos de escrita, o que apresentou maiores erros de classificação foi o do tipo *caracteres disjuntos*, o que pode ter como causa o processo de segmentação implícita adotado. Exemplos dessas confusões são mostrados na Figura 4.1.

A seguir, são descritos alguns resultados encontrados na literatura para o mesmo dicionário.



(a) (b) (c)

(d) (e)

Figura 4.1: Exemplos de erros de classificação. (a) palavra **janeiro** classificada como *fevereiro*, (b) palavra **julho** classificada como *junho*, (c) palavra **maio** classificada como *agosto*, (d) palavra **fevereiro** classificada como *julho* e (e) palavra **setembro** classificada como *fevereiro*.

4.3 Resultados descritos na literatura

Poucos estudos foram encontrados em relação ao reconhecimento das palavras dos meses do ano. Em um deles, Morita *et alli* [38, 39] apresenta uma estratégia de segmentação explícita usando MEM aplicada à base de dados do LARDOC/PUC-PR [26]. Esta base contém 2000 imagens, sendo 1188, 408 e 402 imagens para treinamento, validação e teste, respectivamente. São utilizados dois conjuntos de características, um baseado na análise de concavidades e outro utilizando características perceptivas e a taxa média de reconhecimento obtida usando-se a base de meses do ano foi de 90%.

No trabalho de Kim *et alli* [40, 41] foram feitos experimentos usando palavras dos meses em inglês. Este dicionário se assemelha muito ao português inclusive no problema de terminações iguais para classes distintas. Nesse trabalho, são utilizados dois conjuntos de características: no primeiro conjunto divide-se a imagem em diversas zonas e determina-se características direcionais, de cruzamento e distâncias, além da distribuição dos pixels, enquanto no segundo são utilizadas características de ângulos. A seguir, emprega-se uma metodologia de combinação entre classificadores neurais

e um classificador utilizando MEM. As taxas médias de reconhecimento individuais foram de 86,2% e 77,6%, para o classificador neural e os MEM, respectivamente e a combinação por multiplicação ponderada obteve uma taxa média de reconhecimento de 87,3%. A base de dados do CENPARMI (*Centre for Pattern Recognition and Machine Intelligence*) da *University of Concordia* no Canadá foi usada nestes experimentos e apresenta 4413 imagens para treinamento e 2152 imagens para teste.

Devido às particularidades próprias de cada bases de dados, é dificil fazer comparações dos resultados descritos neste trabalho, com outros descritos na literatura. Como também este não é o objetivo central do presente trabalho, que trata da avaliação de características. Porém, as taxas de reconhecimento obtidas são compatíveis com as de outros autores, comprovando a validade do sistema de referência, bem como das estratégias híbridas de combinação.

4.4 Conclusão

Neste capítulo foram apresentados os testes efetuados e os resultados obtidos no decorrer deste trabalho. A análise dos resultados mostrou que o conjunto de características perceptivas apresentou melhor desempenho para o problema em questão. Porém, ele nem sempre é suficiente, especialmente para palavras que não tenham ascendentes/descendentes ou em que eles estejam localizados na mesma sub-região. Para estes casos, taxas melhores são obtidas com uso conjunto de outras características. Em relação ao classificador, os resultados mostram que as redes neurais atingem um bom desempenho, mas a incorporação de classificadores heterogêneos, ou seja, que possuem estratégias de classificação diferentes, formando um classificador híbrido, pode melhorar estes resultados.

Capítulo 5

Conclusão

O reconhecimento de palavras manuscritas é um problema de difícil solução devido à grande variedade de formas apresentadas pela escrita manual. Com o objetivo de aumentar a discriminação entre as classes são definidas representações que procuram abstrair das imagens das palavras informações únicas à cada classe. Este procedimento é realizado normalmente na etapa de extração de características. Como não existe um consenso sobre a melhor representação, surge uma questão fundamental: **Qual o melhor tipo de característica para representar palavras manuscritas numa dada aplicação?** Alguns autores buscam a sua solução incorporando em seus sistemas informações provenientes de estudos sobre o processo de leitura humano, surgindo então outra questão: **A introdução do conhecimento relativo à leitura humana no modelamento de sistemas de reconhecimento de palavras manuscritas é realmente eficiente e necessário?**

Para responder estas questões foi desenvolvido neste trabalho um sistema de referência para avaliação das características extraídas de palavras manuscritas. A aplicação escolhida foi o reconhecimento das palavras dos meses do ano, por apresentarem similaridades entre as classes, o que ajuda no desenvolvimento da análise em questão. Este sistema é composto por três estágios: pré-processamento, extração de características e classificação. No pré-processamento é feita uma padronização das imagens a partir das normalizações da inclinação média das letras e do declive da palavra. Na etapa de

extração de características, são definidos três conjuntos distintos que incorporam, em diferentes níveis, abstrações da forma da palavra. Estes conjuntos são denominados: conjunto de características perceptivas (P), direcionais (D) e topológicas (T). Por fim, o classificador toma a decisão final de à qual classe pertence a palavra, através de uma rede neural multicamadas treinada com o algoritmo de retropropagação do erro.

Os resultados mostram que o conjunto de características perceptivas produziu a melhor taxa de reconhecimento média, indicando que o classificador utiliza um processamento de informação similiar ao utilizado no sistema de leitura humano para discriminar as diferentes palavras. Portanto, os resultados obtidos mostram que a incorporação do conhecimento do processo de leitura humano na definição das características é válida. Porém, isto não é suficiente para uma boa representação das classes, sendo necessário a utilização conjunta de outras características que venham complementar este conhecimento. Mesmo assim, qualquer sistema de reconhecimento de palavras não deve ignorar na sua etapa de extração de características o potencial das características perceptivas.

Além de avaliar as características, este trabalho analisou também o desempenho do classificador neural interagindo com o classificador Markoviano numa abordagem híbrida, mostrando que estes classificadores agem de forma complementar e permitem um aumento na taxa média de reconhecimento quando utilizados em conjunto. Por fim, utilizando o sistema híbrido se obteve uma taxa de reconhecimento de 90,4%, o que valida às diversas proposições feitas ao longo do trabalho, tanto na definição das características como no sistema de referência em geral.

5.1 Contribuições

Podem ser citadas como contribuições originais deste trabalho:

- Construção de uma base de dados regional *omni-escritor* composta de 6000 imagens provenientes de 500 escritores de diferentes níveis sociais e educacionais, sendo assim bastante heterogênea;

- Avaliação e discussão do potencial de diferentes conjuntos de características em relação à um sistema comum. Na literatura não foi encontrado nenhum estudo que procurasse fazer este tipo de avaliação considerando a metodologia comparativa empregada neste trabalho;

- Desenvolvimento de um novo sistema para o reconhecimento de palavras manuscritas, aplicado na leitura das palavras que representam os meses do ano. O principal destaque do sistema desenvolvido foi a estratégia de segmentação implícita adotada, pois dentro da literatura pesquisada não foi encontrada estratégia semelhante. Apesar de não ser o objetivo central deste trabalho, o sistema de referência descrito apresentou bons resultados, sendo aplicável em tarefas de reconhecimento de palavras com dicionário limitado.

5.2 Perspectivas de trabalhos futuros

Como sugestões de trabalhos futuros, para dar continuidade ao trabalho já desenvolvido podem ser citadas:

- Investigação da utilização de técnicas de seleção de características, buscando a otimização dos vetores a partir da análise da representatividade das características utilizadas nos conjuntos desenvolvidos;

- Aprofundamento do estudo da abordagem híbrida de classificação de modo a melhorar sua interação, otimizando o desempenho conjunto dos classificadores;

- Testar o sistema com outras bases de dados, afim de uma melhor comparação dos resultados com outros descritos na literatura;

- Avaliar o potencial do sistema de referência para outras aplicações, por exemplo, o reconhecimento do extenso manuscrito dos cheques;

- Utilização de um procedimento de validação cruzada para obtenção de classificadores neurais com maior potencial de generalização;

- Avaliar a incorporação de medidas de rejeição evitando que palavras espúrias prejudiquem o desempenho global do sistema.

Bibliografia

[1] Freitas, C. O. de A. *Uso de Modelos Escondidos de Markov para Reconhecimento de Palavras Manuscritas*. Tese de doutorado, Pontifícia Universidade Católica do Paraná, 2001.

[2] Gader, P., Whalen, M., Ganzberger, M. e Hepp, D. Handprinted Word Recongition on a NIST Data Set. *Machine Vision and Application*, 8:31–40, 1995.

[3] Gader, P., Mohamed, M. e Chiang, J.-H. Comparison of Crisp and Fuzzy Character Neural Networks in Handwritten Word Recognition. *IEEE Transactions on Fuzzy Systems*, 3(3):357–363, 1995.

[4] El-Yacoubi, A., Gilloux, M., Sabourin, R. e Suen, C. Y. An HMM-Based Approach for Off-Line Unconstrained Handwritten Word Modeling and Recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 21(8):752–760, 1999.

[5] Côté, M., Lecolinet, E., Cheriet, M. e Suen, C. Y. Automatic Reading of Cursive scripts using a Reading Model and Perceptual Concepts. *International Journal on Document Analysis and Recognition*, 1:3–17, 1998.

[6] Tappert, C. C., Suen, C. Y. e Wakahara, T. The State of Art in On-line Handwritting Recognition. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 12(8):787–808, 1990.

[7] Bartneck, N. The Role of Handwriting Recognition in Future Reading Systems. In *Progress in Handwriting Recognition*, 1996.

[8] Madhvanath, S. e Govindaraju, V. The Role of Holistic Paradigms in Handwritten Word Recognition. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 23(2):149–164, 2001.

[9] Heutte, L. *Reconnaissance de Caractères Manuscrits: Application à la Lecture Automatique des Chèques et des Enveloppes Postales.* Thèse de doctorat, Université de Rouen, 1994.

[10] Côté, M. *Utilisation d'un Modèle d'Accès et de Concepts Perceptifs pour la Reconnaissance d'Images de Mots Cursifs.* Thèse de doctorat, École Nationale Supérieure des Télécommunications, France, 1997.

[11] Freitas, C., El-Yacoubi, A., Bortolozzi, F. and Sabourin, A. Brazilian Bank Check Handwritten Legal Amount Recognition. In *SIBGRAPI'2000*, Gramado - RS, 2000. Simpósio Brasileiro de Computação Gráfica e Processamento de Imagens.

[12] Mohamed, M. A. e Gader, P. Generalized Hidden Markov Models - Part I: Theoretical Frameworks. *IEEE Transactions on Fuzzy Systems*, 8(1):67–81, 2000.

[13] Trier, O. D., Jains, A. K. e Taxt, T. . Feature Extraction Methods for Character Recognition - A Survey. *Pattern Recognition*, 29(4):641–662, 1996.

[14] Schomaker, L. e Segers, E. A Method for the Determination of Features used in Human Reading of Cursive Handwriting. In *IWFHR'98*, The Netherlands, 1998. International Workshop on Frontiers for Handwritten Recognition.

[15] Gillies, A.M. Cursive Word Recognition Using Hidden Markov Models. *Proceedings of the Advanced Technology Conference - United States Postal Service*, 1, 1992.

[16] Chen, M.-Y., Kundu, A., Zhou, J. e Srihari, S. N. Off-Line Handwritten Word Recognition using Hidden Markov Models. *Proceedings of the Advanced Technology Conference - United States Postal Service*, 1, 1992.

[17] Chen, M.-Y., Kundu, A. e Zhou, J. Off-Line Handwritten Word Recognition Using Hidden Markov Model Type Stochastic Network. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 16(5):481–496, 1994.

[18] Kundu, M. e Bahl, P. Recognition of Handwritten Script: A Hidden Markov Model Based Approach. *Relatório Técnico*, 1988.

[19] Kundu, A., He, Y. and Chen, M.-Y. . Alternatives to Variable Duration HMM in Handwriting Recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20(11):1275–1280, November 1998.

[20] Bunke, H., Roth, M. e Schukatt-Talamazzini, E. G. Off-line Cursive Handwriting Recognition using Hidden Markov Models. *Relatório Técnico, IAM-94-008, Institut für Informatik und angewandle Mathematic, Universität Bern*, 1994.

[21] Wang, W., Brakensiek, A., Kosmala and Rigoll G. . HMM Based High Accuracy Off-Line Cursive Handwriting Recognition by a Baseline Detection Error Tolerant Feature Extraction Approach. In *IWFHR'2000*, Amsterdam - The Netherlands, 2000. International Workshop on Frontiers for Handwriting Recognition.

[22] Brakensiek, A., Kosmala, A., Willet, D., Wang, W. e Rigoll G. Performance Evaluation of a New Modeling Technique for Handwriting Recognition Using Identical On-Line and Off- Line Data. In *ICDAR'99*, Bangalore - India, 1999. International Conference on Document Analysis and Recognition.

[23] Brakensiek, A., Rottland, A., Kosmala, A., e Rigoll G. Off-Line Handwriting Recognition Using Various Hybrid Modeling Techniques and Character N-Grams. In *IWFHR'2000*, Amsterdam - The Netherlands, 2000. International Workshop on Frontiers for Handwriting Recognition.

[24] Guillevic, D. and Suen, C. Y. HMM Word Recognition Engine. In *ICDAR'97*, Ulm - Germany, 1997. International Conference on Document Analysis and Recognition.

[25] Gonzalez, R. C. e Woods, R. E. *Digital Image Processing.* Addison-Wesley, 1992.

[26] Freitas, C. O. A., Morita, M., Soares, L. E. O., Justino, E., El Yacoubi, A., Lethelier, E., Bortolozzi, F. e Sabourin, R. Brazilian Bank Check Databases. In *CLEI'2000*, México, 2000. Conferência Latino-Americana de Informática.

[27] HP ScanJet 5200c Series. http://www.pandi.hp.com/pandi-db/prodinfo.main?-product=scanjet5200c&Region=non_us.

[28] Veloso, L. R. Sistema de Reconhecimento de Palavras Manuscritas Dependente do Usuário para a Língua Portuguesa. Proposta de Tese, 2001.

[29] Veloso, L. R. Reconhecimento de Caracteres Numéricos Manuscritos. Dissertação de mestrado, Universidade Federal da Paraíba - Centro de Ciências e Tecnologia - Departamento de Engenharia Elétrica, 1998.

[30] Oliveira Jr., J. J., Veloso, L. R. e Carvalho, J. M. Interpolation/Decimation Scheme Applied to Size Normalization of Characters Images. In *ICPR'2000.* International Conference on Pattern Recognition, 2000.

[31] Schalkoff, R. *Pattern Recognition - Statistical, Structural and Neural Approaches.* Jonh Wiley & Sons, 1992.

[32] Parker, J. R. *Algorithms For Image Processing and Computer Vision.* Jonh Wiley & Sons, 1997.

[33] Haykin, S. *Neural Networks - A Comprehensive Foundation.* Prentice Hall, 1996.

[34] Pandya, A. S. e Mancy, R. B. *Pattern Recognition with Neural Networks on C++.* CRC Press and IEEE Press, 1995.

[35] Bishop, C. M. *Neural Networks for Pattern Recognition.* Oxford University Press, 1995.

[36] A. Zell et al. *SNNS - Stuttgart Neural Network Simulator, User Manual, Version 4.2.* University of Stuttgart, 1994.

[37] Rabiner, L. e Juang, B-H. *Fundamentals of Speech Recognition*. Englewood Cliffs - Prentice Hall, 1993.

[38] Morita, M. E., El Yacoubi, A., Bortolozzi, F. e Sabourin, R. Handwritten Month Word Recognition on Brazilian Bank Cheques. In *ICDAR'2001*, Seattle - USA, Setembro 2001. International Conference on Documents Analysis Recognition.

[39] Morita, M. E., Lethelier, E., El Yacoubi, A., Bortolozzi, F. e Sabourin, R. An HMM-based Approach for Date Recognition. In *DAS'2000*, Rio de Janeiro - Brazil, Dezembro 2000. International Workshop on Document Analysis Systems.

[40] Kim, J. H., Kim, K. K. e Suen, C. Y. Hybrid Schemes of Homogeneous and Heterogeneous Classifiers for Cursive Word Recognition. In *IWFHR'2000*, Amsterdam - Netherlands, Setembro 2000. International Workshop on Frontiers in Handwriting Recognition.

[41] Kim, J. H., Kim, K. K., Nadal, C. P. e Suen, C. Y. A Methodology of Combining HMM and MLP Classifiers for Cursive Word Recognition. In *ICPR'2000*, Barcelona - Spain, Setembro 2000. International Conference on Pattern Recognition.