Reconhecimento de Palavras Manuscritas Usando Análise Multi-Vistas

José Josemar de Oliveira Júnior

Tese de Doutorado submetida ao Programa de Pós-Graduação em Engenharia Elétrica da Universidade Federal de Campina Grande como parte dos requisitos necessários para obtenção do grau de Doutor em Ciências no Domínio da Engenharia Elétrica.

Área de Concentração: Processamento da Informação

João Marques de Carvalho, PhD.
Orientador

Cinthia Obladen de Almendra Freitas, DSc.
Co-orientadora

Robert Sabourin, PhD.
Co-orientador

Campina Grande, Paraíba, Brasil

Reconhecimento de Palavras Manuscritas Usando Análise Multi-Vistas

José Josemar de Oliveira Júnior

Tese de Doutorado apresentada em Outubro de 2006

João Marques de Carvalho, PhD.
Orientador

Cinthia Obladen de Almendra Freitas, DSc.
Co-orientadora

Robert Sabourin, PhD.
Co-orientador

Campina Grande, Paraíba, Brasil, Outubro de 2006

FICHA CATALOGRÁFICA ELABORADA PELA BIBLIOTECA CENTRAL DA UFCG

O48r Oliveira Junior, José Josemar de
2006 Reconhecimento de palavras manuscritas usando análise multi-vistas/ José Josemar de Oliveira Júnior. João Pessoa, 2006.
81.:il.

Referências.
Tese (Doutorado em Engenharia Elétrica) Universidade Federal de Campina Grande, Departamento de Engenharia Elétrica e Informática.
Orientadores: João Marques de Carvalho, Cinthia Obladen de Almendra Freitas e Robert Sabourin.

1 Processamento de Imagens – Reconhecimento de Padrões 2 Modelos Perceptivos – Prcessamento de Imagens 3 Reconhecimento de Manuscritos – Processamento de Imagens I Título

CDU 004.383.3:004.352.242

RECONHECIMENTO DE PALAVRAS MANUSCRITAS USANDO ANÁLISE MULTI-VISTAS

JOSÉ JOSEMAR DE OLIVEIRA JÚNIOR

Tese Aprovada em 30.10.2006

JOÃO MARQUES DE CARVALHO, Ph.D., UFCG
Orientador

CINTHIA OBLADEN DE ALMENDRA FREITAS, Dr., PUC-PR
Orientadora

ROBERT SABOURIN, Dr., ETS-CANADÁ (Ausência Justificada)
Orientador

CARLOS EDUARDO PEDREIRA, Dr., UFRJ
Componente da Banca

RONEI MARCOS DE MORAES, Dr., UFPB
Componente da Banca

HERMAN MARTINS GOMES, Dr., UFCG
Componente da Banca

JOSEANA MACEDO FECHINE, D.Sc., UFCG
Componente da Banca

FRANCISCO MARCOS DE ASSIS, Dr., UFCG
Componente da Banca

CAMPINA GRANDE – PB
OUTUBRO - 2006

Dedicatória

Dedico este trabalho a minha avó, Gentila, que sempre lutou pelos meus sonhos como se fossem seus.

Agradecimentos

A Deus, que durante essa longa caminhada sempre me deu o que precisava e não simplesmente o que pedia;

Ao Professor João Marques, que despertou meu interesse pela pesquisa e pela docência;

À Professora Cinthia Freitas, por sua amizade e paciência e pelo acolhimento durante minha estada em Curitiba;

Ao Professor Robert Sabourin, pelas discussões e trocas de idéias;

Aos membros da banca examinadora, pelas critícas e sugestões, em particular ao Professor Ronei e à Professora Joseana, pela amizade construída ao longo desses anos de trabalho;

À Suzete, Luciana e Vânia, por caminharem ao meu lado compartilhando as agruras e alegrias dessa jornada;

À Katucha e Vanessa que sempre acreditaram no meu potencial, mesmo quando eu duvidava;

Ao Departamendo de Engenharia Elétrica da UFCG pela formação adquirida, sobretudo à COPELE, na pessoa do ex-coordenador Professor Antônio Marcos, e dos seus funcionários, Ângela, Pedro e Eleonora;

À CAPES, pelo suporte financeiro;

Enfim, seria impossível citar o nome de todas as pessoas que contribuíram para o desenvolvimento deste trabalho, mas agradeço a todos os meus familiares, amigos, alunos e voluntários.

"Há pessoas que vêem as coisas como elas são e que perguntam a si mesmas: *Porquê?*. E há pessoas que sonham as coisas como elas jamais foram e que perguntam a si mesmas: ***Por que não?***"

Bernard Shaw

Resumo

Este trabalho propõe uma metodologia de reconhecimento de palavras manuscritas usando diferentes arquiteturas que são inspiradas nas conclusões obtidas em relação aos mecanismos perceptivos e o processo de leitura humano. Como estudo de caso, a abordagem é aplicada ao problema do reconhecimento de palavras manuscritas que representam os meses do ano. Este problema é relevante pois ocorre com frequência no processamento de cheques bancários, dentre outras aplicações. O sistema de análise multi-vistas proposto é formado pelas seguintes arquiteturas: pseudo-segmentação de radical, pseudo-segmentação fixa e pseudo-segmentação variável. Cada arquitetura é formada por um módulo de extração de primitivas, inspirado em modelos perceptivos e específico para o tipo de segmentação utilizado e por um classificador apropriado. Os testes foram realizados com uma base de palavras construída especificamente para este fim, também descrita neste trabalho.

Abstract

This work presents a multiple classifier system applied to the handwritten word recognition (HWR) problem. The goal is to investigate the use of perceptual models in the development of recognition systems. The handwritten words are analyzed considering different approximation levels, in order to get a computational approach of the reading human process. The application proposed is the recognition of the Portuguese handwritten names of the months. The considered system is formed by the following architectures: 2 fixed sub-regions, 8 fixed sub-regions and N variable sub-regions. Each architecture is formed by a module of features extraction, based on perceptual models and specific for each type of segmentation, and an appropriate classifier. The experimental tests have performed on a database specifically built for this problem, also described in this work.

Conteúdo

Lista de Tabelas

Lista de Figuras

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. Poderia se pensar que neste enfoque, documentos em papel constituem relíquias de um tempo distante, principalmente quando se falam em documentos manuscritos. Esta idéia porém não é correta, já que muita informação ainda circula de forma manuscrita, como por exemplo, cartas, cheques bancários e formulários. Por outro lado, na era da informação tecnológica, as vantagens dos computadores e a sua superioridade no armazenamento, transferência e processamento de informações não pode ser desperdiçada. Para resolver esse dilema 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 (SCHOMAKER; SEGERS, 1998).

Hoje em dia, as principais aplicações dos sistemas de leitura automática podem ser encontradas em grandes organizações, em que um grande número de documentos similares devem 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 inicialmente exploraram 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 tem 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 a melhor forma de representar estas palavras considerando a grande variação existente entre elas quando provenientes de escritores diferentes. Uma vez definida a representação tem-se, como última etapa no projeto de sistemas de reconhecimento de palavras manuscritas, a caracterização do classificador cuja função é atribuir um rótulo à palavra desconhecida.

1.1 Motivação

Segundo Correia (2005), o homem é um reconhecedor nato de padrões. Seu sistema de visão o permite distinguir, analisar e reconhecer o que está sendo visto com bastante facilidade. Como exemplo, pode-se citar sua capacidade para reconhecer letras e dígitos. Desde criança, os caracteres pequenos, grandes, impressos em várias fontes, manuscritos ou até mesmo parcialmente incompletos podem ser reconhecidos sem maiores dificuldades.

Sendo assim, diversas linhas de pesquisa na área de sistemas automáticos procuram estabelecer uma possível dualidade homem-computador, incorporando em seus sistemas o conhecimento existente sobre os processos biológicos (SCHOMAKER, 2004). Diversos autores (MADHVANATH; GOVINDARAJU, 2001; FREITAS, 2002) buscam, na análise dos mecanismos de percepção visual e leitura humanos, determinar primícias que auxiliem no desenvolvimento de sistemas de leitura automática.

Por outro lado, no problema de reconhecimento de palavras manuscritas, diferentes métodos de extração de características e técnicas de classificação foram intensivamente estudados nas últimas décadas (PLAMONDON; SRIHARI, 2000; TRIER; JAIN; TAXT, 1996). Muitos métodos de reconhecimento foram propostos, mas nenhum deles isoladamente apresentou resultados suficientes que os validasse como uma solução completa para o problema. Esse fato instigou os pesquisadores a buscarem soluções híbridas, formadas pela combinação de vários métodos.

Deste modo, a principal motivação deste trabalho é definir uma aproximação computacional dos modelos perceptivos, que possa servir de base para a definição de um sistema de reconhecimento de palavras manuscritas.

1.2 Objetivos

O objetivo principal deste trabalho é desenvolver um sistema de leitura automático, de modo a obter uma aproximação computacional inspirada nos mecanismos de percepção usados no processo de leitura humano. Para tanto, é definida uma arquitetura de análise multi-vistas das palavras, de modo que uma mesma amostra seja representada por diferentes arranjos estruturais, que conservam as relações espaciais presentes na palavra. Esses arranjos são implementados utilizando diferentes primitivas e diferentes métodos de reconhecimento, de modo que ao final é concebido um sistema de múltiplos classificadores para reconhecimento de palavras manuscritas.

1.3 Evolução do trabalho

O trabalho desenvolvido nesta tese é fruto de um projeto de cooperação acadêmica (PROCAD/CAPES) estabelecido entre a Universidade Federal de Campina Grande (UFCG) e a Pontifícia Universidade Católica do Paraná (PUC-PR) entre os anos de 2001 e 2006.

O marco inicial do projeto foi o desenvolvimento de um estudo comparativo sobre conjuntos de características aplicados no reconhecimento de palavras manuscritas realizado pelo autor durante seus estudos de mestrado (OLIVEIRA Jr., 2002), sob orientação da Profa. Cinthia Freitas. Essa interação possibilitou o compartilhamento de experiências e o conhecimento do trabalho de tese da professora, que discutia a utilização de modelos escondidos de Markov aplicados no contexto do extenso de cheques bancários (FREITAS, 2001). Alguns anos depois, Marcelo Kapp, sob a mesma orientação, apresentou um sistema com novo conjunto de características e analisou duas arquiteturas de redes neurais nesse processo (KAPP, 2004). Essa sequência de trabalhos possibilitou um aprofundamento no problema em questão que culminou com esse trabalho de tese.

1.4 Organização do texto

O texto deste trabalho encontra-se dividido da seguinte forma:

O Capítulo 2 apresenta um resumo das principais contribuições dos estudos sobre os processos de percepção visual e leitura humanos que dão suporte ao trabalho. Também apresenta-se uma discussão sobre os diferentes níveis de aproximação que os sistemas de leitura automática utilizam na simulação desses processos.

O Capítulo 3 define a arquitetura de análise multi-vistas proposta neste trabalho, discutindo as três estratégias diferentes de pseudo-segmentação utilizadas.

O Capítulo 4 descreve o sistema de reconhecimento de palavras manuscritas, desenvolvido durante os estudos desta tese. Os classificadores utilizados no processo são apresentados juntamente com as representações de características adaptadas à cada esquema de pseudo-segmentação.

O Capítulo 5 apresenta a base de dados utilizada na simulação do sistema, bem como o procedimento experimental aplicado na análise individual e combinada dos classificadores. Também se discute uma nova medida de diversidade para avaliar os classificadores e se faz uma comparação dos resultados obtidos com outros sistemas descritos na literatura.

O Capítulo 6 discute as considerações finais sobre a arquitetura proposta e apresenta as principais contribuições do trabalho.

Capítulo 2

Leitura automática de palavras e modelos perceptivos

Segundo Matos (2004), ao longo de milhares de anos os seres vivos aprenderam e aperfeiçoaram os mecanismos biológicos de reconhecimento de padrões. Embora, para os seres humanos a tarefa de reconhecer um rosto ou a voz de alguém seja trivial, esta é uma tarefa bastante difícil de ser realizada por computadores. Nos sistemas de leitura automática isto não é diferente, desde que as palavras manuscritas são padrões 2D complexos de grande variabilidade devido às variações no estilo de escrita inerentes a cada escritor.

Para solucionar este problema, alguns autores (MADHVANATH; GOVINDARAJU, 2001; CÔTÉ, 1997) procuram modelar computacionalmente os processos cognitivos envolvidos na leitura e reconhecimento de palavras manuscritas no desenvolvimento de sistemas automáticos. Isto não é simples, pois não existe uma teoria universal que determine com exatidão os mecanismos neurais envolvidos no processo de percepção e leitura de palavras. Além disso, a natureza dinâmica desta atividade, bem como a grande utilização de conhecimento prévio, dificulta a determinação de uma teoria única e completa (SEKULER; BLAKE, 1994; FREITAS, 2001).

Neste Capítulo, analisa-se resumidamente os mecanismos de percepção visual, geração e leitura de palavras, bem como faz-se uma análise computacional deste processo através de diferentes níveis de abstração. Os trabalhos de Freitas (2001, 2002) e Correia (2005) serviram de base no desenvolvimento deste capítulo.

2.1 Percepção visual

A percepção visual, segundo a Psicologia, é um processo cognitivo, uma forma de se conhecer o mundo, uma atividade elementar que atribui significados aos estímulos nervosos recebidos pelo cérebro, de modo a construir uma imagem compreensível do objeto para o qual se olha. A percepção visual é considerada um processo de cárater ativo, construtivo e relacionado com processos cognitivos superiores que transcorrem no tempo, ou seja, é um processo completo que depende tanto da informação captada, quanto da fisiologia e das experiências anteriores de quem percebe. Como a percepção depende em parte do que está sendo visto e, em parte, do que foi aprendido, esta é relacionada com processos de aprendizagem e memorização. O processo de aprendizagem nos humanos tem início na infância e se torna cada vez mais rápido e exato com o aumento da idade. Outros fatores que podem influenciar a percepção visual são motivações e expectativas, uma vez que, a motivação sugestiona a visão para ver apenas o que se deseja e a expectativa acentua a prontidão para responder apenas a um tipo de estímulo (CORREIA, 2005).

O ato de perceber consta de duas fases. Na primeira fase, denominada pré-atencional, o indíviduo detecta a informação sensorial e a analisa. Na segunda fase, denominada construção pessoal, ele produz o objeto perceptual específico. Esse processo perceptivo é cíclico, uma vez que há uma constante antecipação do que sucederá, baseada na informação que acaba de entrar pelos olhos e em esquemas que selecionam a informação a ser processada, com base em processos probabilísticos extraídos da experiência anterior. Pode-se dizer que as experiências perceptuais seguintes tendem a influenciar as anteriores, eliminando a possibilidade de haver duas experiências idênticas (CORREIA, 2005).

Estas observações comprovam a complexidade do mecanismo de percepção e justificam a dificuldade de se determinar uma aproximação computacional desse processo.

2.1.1 Teorias perceptivas: sintética e analítica

A maneira como o homem percebe visualmente as formas não é totalmente entendida e no intuito de compreender este mecanismo, duas abordagens são aceitas pelos psicológos: a sintética e a analítica (SEKULER; BLAKE, 1994; FREITAS, 2001). A primeira nasceu com o trabalho de Max Wertheimer em 1912, cujos seguidores acreditam que as formas são percebidas como um todo, tendo como lema principal a frase: *O todo é diferente da soma das partes individuais*. A abordagem analítica, também conhecida como teoria estruturalista nasceu no século XIX e apóia a idéia principal de que processos mentais complexos são criados combinando-se componentes fundamentais, deste modo sensações simples formam os blocos construtores da forma percebida.

A abordagem sintética, também conhecida como teoria da forma ou teoria Gestalt (palavra alemã que significa forma), considera a imagem projetada na retina como um todo indissociável, não analisável, contendo toda a informação necessária para a sua percepção. Em outras palavras, cada imagem na retina provoca uma percepção global única, em que o cérebro não executa elementos isolados, mas sim a relação entre eles, fazendo com que se enxergue o todo e não suas partes constituintes. Nesta teoria, a percepção inseparável das formas é marcada por princípios de organização, como a separação figura/fundo e por estruturas regulares da forma.

A outra abordagem, denominada analítica, é formada por teorias que, ao contrário da teoria Gestalt, enfatizam que a imagem não é suficiente para a percepção exata dos objetos e que outras variáveis decorrentes de uma análise são necessárias. A teoria afirma que a forma não é percebida como um todo, mas pelas partes que a constitui. Os movimentos oculares são os mecanismos fundamentais, para integrar os elementos básicos da percepção. Resultados semelhantes são apresentados por Piaget (VERNON, 1974), que garante que, a busca sistemática pelo movimento dos olhos e fixações sucessivas de aspectos importantes da forma é um fator essencial da atividade perceptiva.

É importante ressaltar, que o processo de percepção é individual e as teorias não podem ser generalizadas, uma vez que as pessoas não reagem da mesma maneira ante a uma imagem. A percepção pode variar até para um mesmo indivíduo, dependendo da sua idade e do tempo de exposição à imagem (SEKULER; BLAKE, 1994). Assim, é possível que para adultos, a busca visual seja desnecessária devido à sua maturidade e que, neste caso, o padrão é percebido como um todo, enquanto que uma criança precisa investigar melhor o padrão antes de tomar uma decisão. Essas teorias traduzem as abordagens discutidas anteriormente, sobre o mecanismo de percepção visual, para o contexto de leitura de palavras manuscritas.

2.2 Geração e leitura de palavras manuscritas

Manuscritos consistem de marcas gráficas cujo propósito é comunicar algo que pode ser uma idéia ou pensamento utilizando, para tanto, uma convenção que relaciona essas marcas à linguagem. A geração de manuscritos envolvem diversas funções. A partir de uma intenção de comunicação, uma mensagem é preparada com base em conceitos semânticos, sintáticos e léxicos, que são convertidos de algum modo em um conjunto de alógrafos e grafos constituídos por traços (PLAMONDON; SRIHARI, 2000).

Os processos perceptivos envolvidos na leitura têm sido estudados em profundidade pela Psicologia Cognitiva. Tais estudos são pertinentes desde que formam a base dos algoritmos que emulam o processo de leitura humano. Embora muitos estudos tratem da leitura de caracteres

impressos, algumas conclusões são igualmente válidas para textos manuscritos. Por exemplo, os movimentos oculares fixam-se em pontos discretos no texto e em cada ponto de fixação o cerébro usa o campo periférico visual para inferir a forma do texto (CORREIA, 2005).

Segundo experimentos psicológicos realizados em seres humanos, o processo de reconhecimento de caracteres isolados é realizado de dois modos distintos (MADHVANATH; GOVINDARAJU, 2001):

1. caractere que possua uma estrutura simples ou que ocorra frequentemente é processado como uma unidade única sem qualquer decomposição em estruturas mais simples, sendo denominado reconhecimento holístico;

2. caractere que não ocorre frequentemente ou que tenha uma estrutura complexa consome mais tempo no seu reconhecimento, sendo este proporcional ao número de traços que formam o caractere, intitulado reconhecimento analítico.

Deste modo, a teoria holística sugere que as palavras são identificadas diretamente a partir de sua forma global enquanto a teoria analítica afirma que o reconhecimento é feito a partir da identificação das letras componentes (MADHVANATH; GOVINDARAJU, 2001).

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 (SCHOMAKER; SEGERS, 1998; CÔTÉ, 1997) apresentam conclusões sobre este processo, descritas a seguir:

- Em um primeiro nível do processo de leitura, 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 das vogais ($handwriting = hndwrtng$);

- O processo de leitura das vogais (*a,e,i,o*) não apresenta confusões entre si, 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 em sua parte final;

- 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 àquelas 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 início 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 (MADHVANATH; GOVINDARAJU, 2001). 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.

Experimentos apontam também para a existência de um *efeito da superioridade da palavra*, com base no princípio de que as palavras são lidas mais facilmente ao se apresentar de uma única vez todas as letras que formam a palavra inteira do que ao serem apresentadas individualmente (SEKULER; BLAKE, 1994). Isto demonstra que ao ver uma palavra, o leitor não processa cada letra isoladamente das demais.

As discussões sobre o processo de leitura humano apontam para a supremacia dos métodos holísticos, porém no desenvolvimento de sistemas automáticos isso nem sempre é verdade, dada a importância do conhecimento prévio do leitor durante o processo de leitura. Desse modo, os sistemas de leitura automática buscam uma aproximação desse processo para superar as dificuldades, como será discutido a seguir.

2.3 Níveis de abstração

A modelagem por níveis de abstração, definida por Freitas (2002), auxilia na compreensão do processo de percepção de palavras manuscritas, sobretudo no que diz respeito às limitações dos sistemas automáticos, em relação ao processo humano de leitura. Deste modo, as imagens são analisadas em 4 níveis de abstração, a saber:

- **Nível objeto**: a palavra e seu contexto;
- **Nível global**: a palavra;
- **Nível local**: as letras que compõem a palavra;
- **Nível *pixel***: a palavra em meio digital.

A decomposição do problema de reconhecimento de palavras manuscritas em 4 níveis de abstração permite que se observe, de maneira hierárquica, a complexidade inerente ao problema, bem como, as dificuldades do processamento de imagens, em contraposição ao objetivo do reconhecimento. Isto, devido ao fato da imagem estar em formato digital e, portanto, na representação em nível dos *pixels*, enquanto a visão do observador trabalha em nível dos objetos. Uma análise de cada um desses níveis é realizada a seguir.

2.3.1 Nível objeto: a palavra e seu contexto

A palavra manuscrita em geral vem acompanhada de um entorno, ou seja, do contexto em que ela é empregada. Na área de Reconhecimento de Palavras Manuscritas (RPM) os mais variados contextos podem ser encontrados, como: cheques bancários (HEUTTE, 1994; GUILLEVIC, 1995; ÂVILA, 1996; CÔTÉ et al., 1998; FREITAS, 2001; MORITA et al., 2004), envelopes postais (EL-YACOUBI; SABOURIN; SUEN, 1999), formulários diversos e fichas de cadastramento, dentre outros. Na Figura 2.1 são exemplificadas imagens de palavras manuscritas nos contextos de cheques bancários e envelopes postais.

O contexto influencia no processo de reconhecimento de palavras manuscritas, uma vez que pode implicar em um vocabulário finito de palavras, como no caso de valores escritos por extenso ou datas de cheques bancários. Ou não, como no caso de endereços em envelopes postais, em que o léxico é de grande dimensão em geral, pois envolve os nomes de ruas de uma ou várias cidades.

(a) (b)

Figura 2.1: Exemplos de palavras manuscritas e o contexto: (a) cheque bancário e (b) envelope postal.

O contexto também influencia na segmentação das palavras manuscritas. Como exemplo, na Figura 2.1-a pode-se ressaltar a influência do padrão de fundo na percepção das informações manuscritas no cheque bancário. Na Figura 2.1-b pode-se observar que o selo e o carimbo

sobrepõem-se às palavras manuscritas. Estas influências fazem com que as etapas de processamento dos objetos em questão sejam prejudicadas ou dificultadas.

Assim, muitos esforços computacionais podem ser realizados com o intuito de atender ao princípio da Gestalt conhecido como figura/fundo (FILHO, 2000). A separação figura/fundo designa a divisão do campo visual em duas regiões, separadas por um contorno. No interior do contorno encontra-se a figura, com característica de objeto, mesmo que não reconhecível, e no fundo encontra-se a textura, a qual é percebida como se estendendo atrás da figura. A teoria Gestalt propõe que a separação figura/fundo é uma propriedade organizadora do sistema visual em que toda figura é percebida em seu ambiente, em seu contexto (CORREIA, 2005).

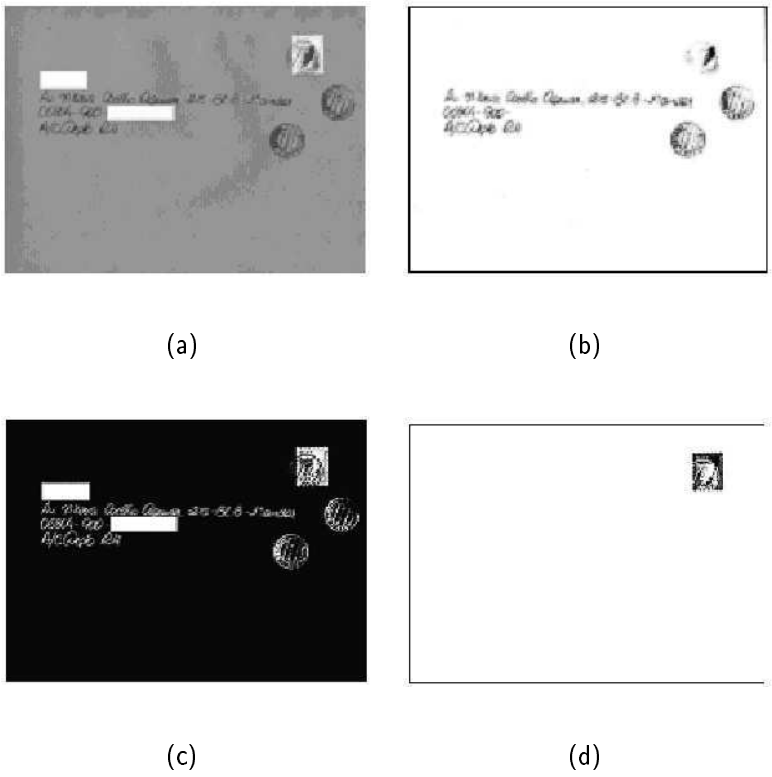
No que se refere à segmentação das palavras manuscritas em relação ao seu contexto, muitos estudos são encontrados na literatura que aplicam diferentes técnicas, tais como:

- segmentação por análise dos componentes conectados da imagem do documento, que associa aos elementos um valor de confiança (YU; JAIN; MOHIUDDIN, 1997);
- segmentação por contexto, que busca regiões em que a informação manuscrita é esperada *a priori* (LII; PALUMBO; SRIHARI, 1993);
- segmentação por multi-binarização, que avalia as faixas de níveis de cinza presentes na imagem do objeto (TSAI, 1985; YAN, 1996);
- segmentação por textura, que elimina o fundo complexo de certos documentos (RUZON, 1997);
- segmentação morfológica *Watershed*, que extrai informações em imagens complexas (FACON, 1996; BEUCHER; MEYER, 1992).

Como exemplo, pode-se citar o trabalho proposto por Yonekura e Facon (2003) em que uma metodologia de segmentação automática de envelopes postais com fundo complexo busca separar o conteúdo da imagem dos envelopes em três classes: bloco endereço manuscrito e carimbo, fundo e selo. Os autores utilizam pouco conhecimento *a priori* dos envelopes e a metodologia faz uma mescla da abordagem *Watershed* morfológico com o conceito de matriz de co-ocorrência. Na Figura 2.2 são exemplificados os resultados obtidos com a metodologia proposta.

2.3.2 Nível global: palavra

Uma vez que o conteúdo manuscrito seja extraído das imagens, é necessário separá-lo em palavras, dando início ao processo de reconhecimento de palavras manuscritas, que utiliza uma



(a) (b)

(c) (d)

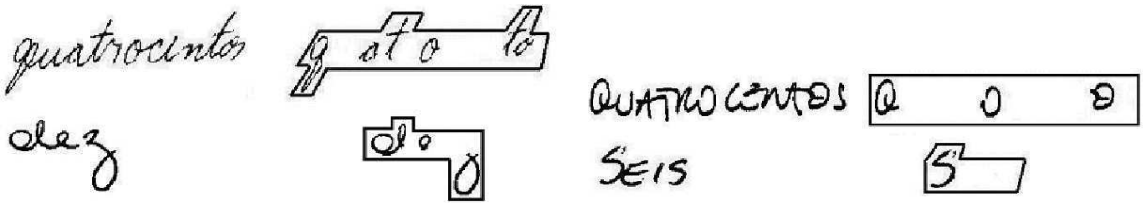
Figura 2.2: Exemplo de segmentação figura-fundo (extraída de Freitas (2002)): (a) imagem original, (b) bloco endereço manuscrito e carimbo, (c) fundo e (d) selo.

de duas possíveis abordagens: global (*global approach*) realizada em nível de palavras (GUILLEVIC, 1995; CÔTÉ et al., 1998; FREITAS, 2001; OLIVEIRA Jr. et al., 2002; KAPP; FREITAS; SABOURIN, 2003) ou local (*analytical approach*) realizada em nível de caracteres (LECOLINET, 1990; GILLOUX; LEROUX; BERTILLE, 1995; KIM, 1996; MORITA et al., 2004).

A abordagem global evita a etapa de segmentação explícita, extraindo primitivas globais diretamente das palavras. Esta abordagem procura explorar as informações do contexto, permitindo que características obtidas de modelos psicológicos possam ser inseridas (GUILLEVIC, 1995; CÔTÉ et al., 1998). Porém, é restrita às aplicações com léxicos pequenos, devido à necessidade de se obter modelos individuais para cada uma das palavras do léxico estudado, bem como, por requerer uma base de dados de tamanho satisfatório para treinamento dos modelos. Estes requisitos podem ser, na prática, difíceis de satisfazer.

Uma vantagem em trabalhar com as palavras de forma global está em não depender da segmentação das palavras em letras ou pseudo-letras. Adicionalmente, a utilização deste nível de abstração tenta incorporar princípios do processo de leitura humano ao método computacional. Na Figura 2.3 são exemplificadas imagens de palavras da base de dados de cheques bancários

da PUC-PR (FREITAS, 2001), demonstrando a complexidade do problema. Nesta figura, a forma da palavra cursiva apresenta informação discriminante que pode ser usada no reconhecimento (MADHVANATH; GOVINDARAJU, 2001; TAPPERT; SUEN; WAKAHARA, 1990). Isto não ocorre quando a palavra é escrita em estilo letra de forma ou caixa-alta, que impossibilita a análise global através da forma do traçado como um todo.



(a) (b)

Figura 2.3: Exemplos de análise global de palavras manuscritas por meio da extração do contorno (extraída de Freitas (2002)): (a) estilo cursivo e (b) estilo caixa-alta.

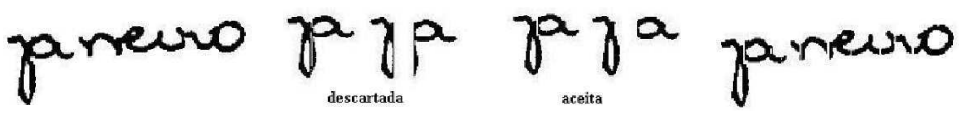
2.3.3 Nível local: letra

Na abordagem local existe a necessidade de segmentar a palavra em elementos básicos, que podem ser letras (caracteres) ou segmentos de letras (pseudo-letras). Deste modo, procura-se modelar os caracteres que compõem as palavras. Esta abordagem caracteriza-se pela dificuldade em definir a fronteira entre os caracteres e pelo fato de que o desempenho do método de reconhecimento dependerá do sucesso do processo de segmentação. Estes métodos podem ser desenvolvidos considerando etapas distintas para segmentação e reconhecimento, ou associando ambas em uma única etapa.

Morita et al. (2004) descreve um método de segmentação de palavras manuscritas e exemplifica a dificuldade da definição da fronteira entre os caracteres de uma mesma palavra. No método proposto busca-se a minimização da largura vertical do traço no provável ponto de segmentação, como ilustrado na Figura 2.4.

2.3.4 Nível *pixel*: palavra digital

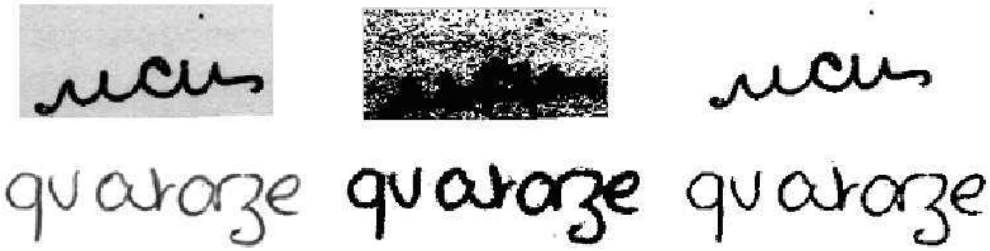
O nível mais baixo de abstração de uma imagem é o *pixel*, no qual são aplicados grande parte dos métodos e técnicas de processamento de imagens. Neste nível, todos os demais se tornam desconhecidos, seguindo o princípio de que quanto mais baixo o nível de representação da imagem, menos informação se têm sobre o todo. Um exemplo de método clássico de processamento de



(a) (b) (c) (d)

Figura 2.4: Exemplos de segmentação de palavras manuscritas em letras ou em pseudo-letras (extraída de Freitas (2002): (a) imagem original, (b) hipótese de segmentação descartada, (c) hipótese de segmentação aceita e (d) palavra segmentada.

imagem em nível de *pixel* é a operação de limiarização ou *thresholding* (GONZALEZ; WOODS, 1992). Por meio dela obtém-se uma nova representação da imagem original, utilizando outra composição de níveis de cinza, como em uma imagem binária. Na Figura 2.5 são exemplificados o resultado da aplicação de dois algoritmos de limiarização (FREITAS, 2002) para imagens de palavras manuscritas em níveis de cinza.



(a) (b) (c)

Figura 2.5: Exemplos de processo de binarização (extraída de Freitas (2002)): (a) imagens originais em níveis de cinza, (b) imagens limiarizadas pelo método de anisotropia e (c) imagens limiarizadas pelo método de Otsu (1979).

Todos esses aspectos mostram que a capacidade humana de ler desafia os pesquisadores a entender o funcionamento dos mecanismos humanos utilizados nesta tarefa, bem como de solucionar os problemas que dificultam o desenvolvimento de sistemas computacionais de leitura automática.

2.4 Discussão

Muitos trabalhos procuram nos estudos sobre a percepção visual humana estabelecer fundamentos para o desenvolvimento de sistemas de leitura automática. As discussões baseiam-se em

duas teorias perceptivas: abordagens sintética e analítica. Na primeira, argumenta-se que as formas são percebidas como um todo, enquanto na outra afirma-se que o processo de percepção depende da detecção de elementos fundamentais. A maioria dos estudos experimentais sobre o processo de leitura humano descritos na literatura sustenta a hipótese sintética. Entretanto, os mecanismos envolvidos nesse processo são subjetivos e não existem estudos conclusivos sobre o assunto.

Do ponto de vista computacional, os sistemas holísticos possuem limitações práticas, principalmente em relação ao tamanho do léxico utilizado. Para um grande número de palavras, é difícil reconhecê-las unicamente utilizando características perceptivas. Deste modo, a solução mais coerente seria desenvolver sistemas híbridos que busquem explorar os pontos positivos das duas abordagens. Uma solução seria determinar uma modelagem que representasse da melhor forma possível as características perceptivas, que são mais discriminantes, mas que também considerasse paralelamente uma análise mais refinada dos segmentos constituintes da palavra. Neste trabalho é proposta uma nova arquitetura de reconhecimento que implementa esta solução.

Em termos de representação computacional, o bom funcionamento de um sistema de leitura automática necessita de um correto mapeamento entre as primitivas (características) extraídas da imagem e o processo de tomada de decisão (classificador). Também é conhecido que a integração de múltiplos classificadores descorrelacionados determina uma solução melhor do que um único classificador. Essa característica também é explorada na modelagem proposta.

Capítulo 3

Análise Multi-Vistas

No Capítulo 2 foram discutidos os diferentes níveis de abstração que podem ser utilizados na representação de palavras manuscritas do ponto de vista computacional e suas implicações em termos de processamento digital de imagens. Dentre eles, o principal foco de interesse deste trabalho está no nível global, que procura interpretar a palavra como um todo, simulando o processo natural em que o leitor utiliza conhecimento prévio juntamente com codificações da forma da palavra no processo de reconhecimento.

Embora a análise global não se baseie em segmentação, um processo de pseudo-segmentação pode ser utilizado para produzir um sistema de reconhecimento robusto. O objetivo é representar partes da palavra isoladamente considerando o arranjo espacial dessas partes, sem perder a visão do todo. Essa estratégia facilita a representação computacional do método e difere da abordagem analítica que procura reconhecer os caracteres isoladamente para então determinar a palavra.

Neste trabalho define-se uma nova arquitetura de reconhecimento, denominada de **análise multi-vistas**, que busca diferentes aproximações simultâneas da mesma amostra. Esse objetivo é alcançado ao se particionar a palavra segundo diferentes arranjos espaciais. O intuito é obter perspectivas diferentes da mesma imagem que forneçam informações complementares, facilitando a tomada de decisão. Essa modelagem computacional determina uma arquitetura global que incorpora aspectos estruturais da palavra. Deste modo, alia-se o poder discriminante dos elementos perceptivos inerentes ao reconhecimento global à uma análise local dos segmentos da palavra.

Os processos de pseudo-segmentação utilizados são definidos a seguir:

- **Pseudo-segmentação de radical (PR)** - A imagem é dividida em duas zonas, a partir do seu centro de gravidade, com o objetivo de analisar separadamente os prefixos e os

sufixos (KAPP; FREITAS; SABOURIN, 2003; KAPP, 2004);

- **Pseudo-segmentação fixa (PF)** - A imagem é dividida em oito partes iguais, valor que corresponde ao número médio de caracteres que formam as palavras do léxico em análise que será descrito posteriormente (OLIVEIRA Jr. et al., 2002; OLIVEIRA Jr., 2002);

- **Pseudo-segmentação variável (PV)** - A imagem é dividida em um número variável de segmentos de tamanhos diferentes, a partir de uma análise da linha central da palavra. Este processo permite uma forma de representação mais refinada em relação aos métodos anteriores, além de se adequar a casos em que mais segmentos são necessários, determinando uma representação local dos segmentos (FREITAS; BORTOLOZZI; SABOURIN, 2004b).

A partir da análise multi-vistas são definidas estratégias de extração de características e classificação, porém com o mesmo princípio de interpretação global da palavra e incorporação de primitivas perceptivas. Ao final, os classificadores são combinados de modo a ressaltar a complementariedade dos diferentes modelos de pseudo-segmentação.

Na Figura 3.1 é apresentado um diagrama representativo da arquitetura. A imagem de entrada é analisada de acordo com cada tipo de pseudo-segmentação. Para cada estratégia é definido um processo de extração de características e uma estratégia de reconhecimento (rede neural ou modelo escondido de Markov), respectivamente. Ao final é feita uma integração das diferentes análises por meio da combinação dos classificadores, resultando na palavra reconhecida.

Essa arquitetura realça diferentes visões do problema indo da análise macroscópica, representada pela análise de prefixos e sufixos, passando por uma análise intermediária que procura representar os caracteres isolados da palavra, até uma análise microscópica que procura determinar segmentos importantes da estrutura da palavra. Estas informações complementares auxiliam no processo de tomada de decisão dos classificadores, principalmente no reconhecimento de amostras confusas.

Considerando a arquitetura de análise multi-vistas apresentada foi desenvolvido um sistema de reconhecimento de palavras manuscritas que é apresentado no próximo capítulo.

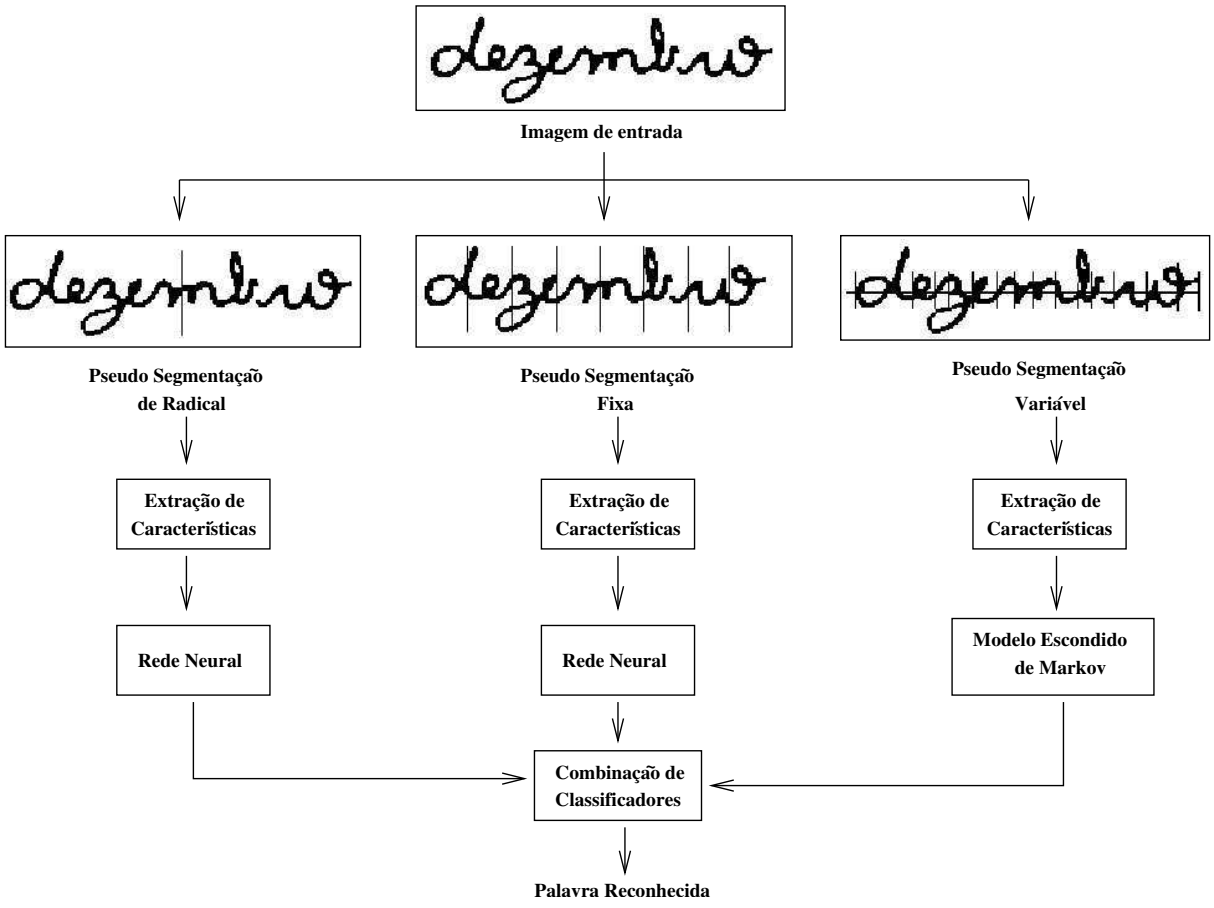


Figura 3.1: Diagrama em blocos representativo da arquitetura de análise multi-vistas.

Capítulo 4

Descrição do sistema

Neste capítulo é apresentado o sistema desenvolvido neste trabalho com base na arquitetura de Análise Multi-Vistas discutida anteriormente. Inicialmente, é definido o dicionário que será utilizado e em seguida serão apresentadas as partes constituintes do sistema. Inicialmente, são mostradas as operações de pré-processamento (conjunto de algoritmos para eliminação do ruído e normalização das imagens) aplicadas à base de dados utilizada. Em seguida, definem-se os métodos de reconhecimento para cada processo de pseudo-segmentação definido na análise multi-vistas, sendo discutidos a formação do vetor de características e o método de reconhecimento utilizado. Como as saídas dos classificadores são combinadas de modo a produzir uma decisão final sobre a amostra em análise, também é feita uma discussão sobre o processo de combinação de múltiplos classificadores ao final do capítulo.

4.1 Descrição do dicionário

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

Por exemplo, pode-se considerar um sistema para processamento automático de cheques bancários. O sistema deve considerar diferentes tipos de dados, tais como dígitos e palavras, escritos em diferentes estilos. Neste exemplo, localiza-se o estudo de caso realizado neste trabalho, que é o reconhecimento das palavras dos meses do ano, representado por um léxico de 12 classes: *Janeiro*, *Fevereiro*, *Março*, *Abril*, *Maio*, *Junho*, *Julho*, *Agosto*, *Setembro*, *Outubro*, *Novembro* e *Dezembro*.

Apesar do número pequeno de classes envolvidas neste problema, algumas características das mesmas contribuem para aumentar a sua complexidade. Como pode ser observado na Figura 4.1, há classes muito semelhantes e/ou com mesma terminação no léxico. Além disso, em alguns nomes a primeira letra coincide, sendo a mesma muito importante no reconhecimento de palavras, como observado por Schomaker e Segers (1998). Estes e outros fatores afetam o desempenho de qualquer método de reconhecimento, o que torna esta aplicação um desafio interessante para a arquitetura proposta.

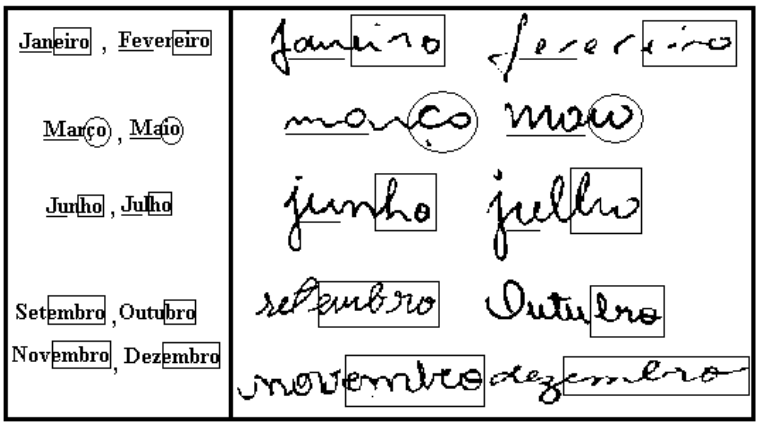


Figura 4.1: Complexidade do problema de reconhecimento em estudo: semelhança entre prefixos e sufixos.

4.2 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 (2001) que consistem de três etapas, descritas a seguir:

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

As etapas de normalização são necessárias pois os formulários utilizados no processo de construção da base de dados descrito no próximo capítulo não fornecem 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 dos formulários ou ocasionados pelas operações de normalização.

4.2.1 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 (θ), 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. Uma vez obtidos os perfis, é calculada a entropia associada a cada perfil de projeção, segundo a Equação 4.1:

$$H_\theta = -\sum_{v=1}^{L} P_v(\theta) \log P_v(\theta); \tag{4.1}$$

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. A probabilidade é estimada pela razão entre o número de *pixels* pretos e o número total de *pixels* em cada coluna.

O ângulo que proporciona a menor entropia é considerado o ângulo de inclinação média α 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, de tamanho $M \times N$, com coordenadas (i, j) são calculadas as suas novas coordenadas $(i\prime, j\prime)$ na imagem normalizada, utilizando a Equação 4.2:

$$\begin{aligned} j' &= \lfloor j - (M - i).tan\alpha \rfloor, \\ i' &= i. \end{aligned} \tag{4.2}$$

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.

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

$$\begin{aligned} i' &= i + j \tan \alpha \\ j' &= j. \end{aligned} \tag{4.3}$$

Nesta equação, α é o ângulo pelo qual deseja-se rotacionar os *pixels* da imagem.

4.2.3 Suavização

O algoritmo de suavização utilizado baseia-se no deslocamento de máscaras sobre a imagem. Estas máscaras, definidas por Veloso (2001), 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 4.2. 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 4.3. 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, os elementos centrais tem seu valor modificado (0 para 1 ou 1 para 0).

A seguir, são apresentados os métodos de reconhecimento utilizados no processo de análise multi-vistas. Inicialmente é apresentado o processo de extração de características aplicado em cada mecanismo de pseudo-segmentação e em seguida é feita uma discussão teórica sobre os classificadores utilizados: redes neurais e modelos escondidos de Markov.

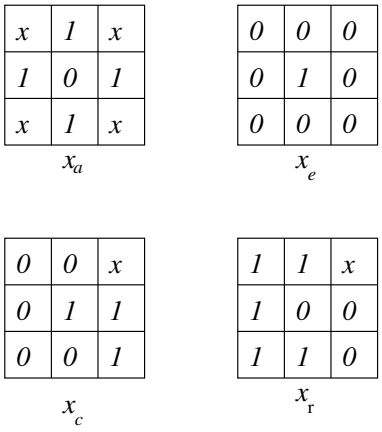


Figura 4.2: Máscaras utilizadas no processo de suavização - primeiro procedimento.

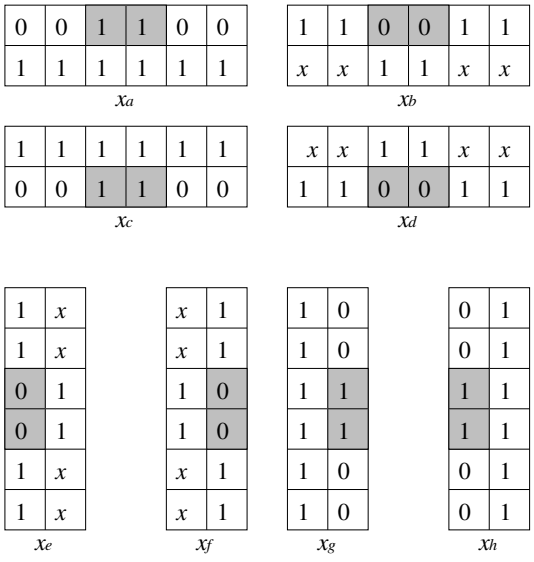


Figura 4.3: Máscaras utilizadas no processo de suavização - segundo procedimento.

4.3 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 amostra de entrada (FREITAS et al., 2000; MOHAMED; GADER, 2000). 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.

Neste trabalho, para cada método de pseudo-segmentação é definido um conjunto de características apropriado, que é mostrado a seguir.

4.3.1 Pseudo-segmentação de radical (PR)

O método de reconhecimento descrito nesta seção foi desenvolvido por Kapp (2004). O mecanismo de pseudo-segmentação utilizado separa a palavra, a partir do seu centro de gravidade, em duas regiões, a da esquerda e a da direita, como mostrado na Figura 4.4. O objetivo é explorar a ocorrência das características em cada região, obtendo assim a informação sobre o posicionamento das mesmas na palavra, o que dá mais precisão na classificação das formas.

Figura 4.4: Exemplo do zoneamento utilizado.

Neste método são utilizadas características perceptivas (MADHVANATH; GOVINDARAJU, 2001) e geométricas, representadas pela contagem de suas ocorrências na imagem da palavra. As características são extraídas de cada palavra gerando um vetor de características de dimensão 24. As características perceptivas são consideradas características de alto nível, de acordo com a classificação de Madhvanath e Govindaraju (2001) e sua utilização é justificada pelo processo de leitura humano que usa os traços ascendentes, descendentes e a estimação do comprimento das palavras para ler sentenças manuscritas.

O primeiro passo na extração de características perceptivas é determinar as zonas da palavra, que de acordo com a definição de Freitas (2001) se divide em:

- 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). A partir do histograma, as linhas superior (LS) e inferior (LI) do corpo da palavra 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 (2001) 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. Exemplo deste processo é apresentado na Figura 4.5.

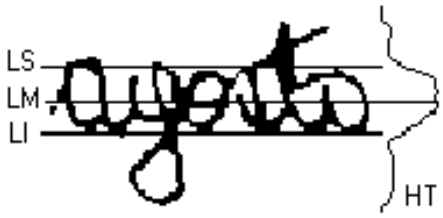


Figura 4.5: Exemplo do processo de detecção das zonas da palavra.

O conjunto de características extraído de cada uma das regiões (esquerda e direita) da palavra, é descrito a seguir:

- Número de laços, ascendentes e descendentes. 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;
- Número de semicírculos côncavos e convexos no corpo da palavra, Figura 4.6-a e Figura 4.6-b, respectivamente. Inicialmente, a palavra é esqueletizada através do algoritmo de Holt (HOLT et al., 1987; PARKER, 1997). Em seguida, os pontos côncavos e convexos

são extraídos do corpo das palavras esqueletizadas por morfologia matemática, utilizando elementos estruturantes diferentes. As concavidades e convexidades constituem primitivas complementares, ou seja, auxiliam na representação das curvaturas das letras e ligações entre letras, ou ainda, de laços abertos existentes no corpo das palavras;

- Número de pontos de cruzamento, de ramificação e finalizadores, Figuras 4.6-c, 4.6-d, 4.6-e, respectivamente. Esses pontos também são obtidos por morfologia matemática considerando além do corpo da palavra, as regiões ascendente e descendente;
- Número de cruzamentos com o eixo horizontal da palavra (NCH), Figura 4.6-f. O eixo horizontal corresponde à linha média obtida, como descrito anteriormente e ilustrado na Figura 4.5;
- Proporção de *pixels* que fazem parte do traçado em relação ao contexto da palavra (NPP), Figura 4.6-g. O menor retângulo envolvente da palavra é utilizado no cálculo da proporção, sendo obtida usando a Equação 4.4:

$$prop = \frac{tp - tpp}{tp} \tag{4.4}$$

  em que tp e tpp são, o número total de *pixels* e a quantidade de *pixels* pretos dentro do retângulo, respectivamente;
- Número de traços verticais, Figura 4.6-h e número de traços horizontais, obtidos por morfologia matemática usando elementos estruturantes que representam linhas verticais e horizontais, respectivamente;
- Número de laços ascendentes e descendentes.

4.3.2 Pseudo-segmentação fixa (PF)

Neste método foi aplicado um processo de pseudo-segmentação dividindo cada imagem em 8 sub-regiões de mesmo tamanho (OLIVEIRA Jr. et al., 2002; OLIVEIRA Jr., 2002). Este número de sub-regiões corresponde ao número médio de letras presentes nas palavras que formam o léxico em análise, descrito posteriormente. Para cada sub-região é extraído um vetor x com 10 padrões, $x = x_i$ em que $i = 1, \cdots, 10$. Deste modo, é formado um vetor de características de dimensão 80 para cada imagem. Exemplo desse procedimento é mostrado na Figura 4.7.

Para se obter um conjunto de características invariante à escala, todos os componentes do vetor de características foram normalizados no intervalo [0,1], em função da própria definição dos

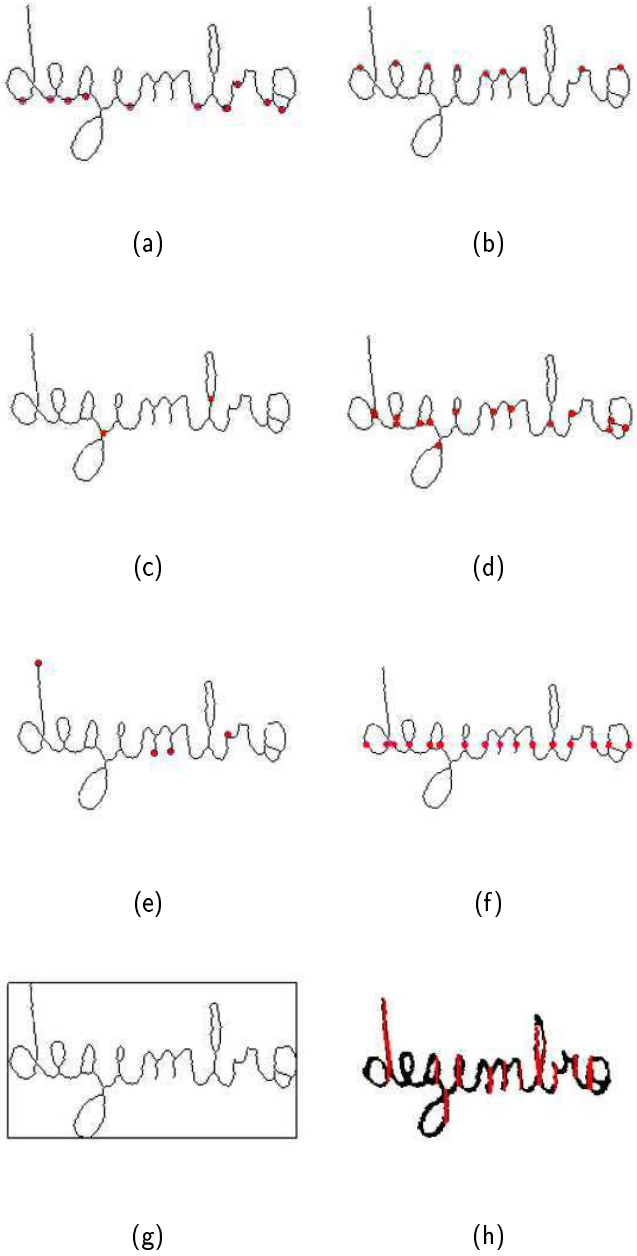


Figura 4.6: Exemplo do processo de extração de características aplicado no processo de pseudo-segmentação de radical (PR): (a) semicírculos côncavos, (b) semicírculos convexos, (c) pontos de cruzamento, (d) pontos de ramificação, (e) pontos finalizadores, (f) NCH, (g) NPP e (h) traços verticais.

Figura 4.7: Exemplo do processo de pseudo-segmentação fixa (PF).

padrões. A partir disso, foram definidos três conjuntos de características diferentes, denominados de conjuntos de características perceptivas, direcionais e topológicas, descritos a seguir.

Características perceptivas (PF-P)

Na extração de características perceptivas, é necessário inicialmente se determinar as zonas da palavra. Estas zonas são definidas da mesma forma como apresentado na Seção 4.3.1. Uma vez determinadas as zonas da palavra e após o processo de pseudo-segmentação, são extraídos os 10 padrões de cada uma das 8 sub-regiões, que são definidos a seguir (OLIVEIRA Jr. et al., 2002; OLIVEIRA Jr., 2002):

- 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 (2001), 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. Esta posição é determinada calculando a posição média entre os pontos extremos do ascendente em relação à horizontal, que é normalizada pelo número de pixels da sub-região na 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 pelo número de pixels do corpo da palavra na vertical;

- 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;

- Localização do laço: É dada pelas coordenadas do centro de massa do laço, que são definidas de acordo com a Equação 4.5:

$$(X_{cm}, Y_{cm}) = (\frac{\sum_{i=1}^{M}\sum_{j=1}^{N} i.f_{ij}}{M}, \frac{\sum_{i=1}^{M}\sum_{j=1}^{N} j.f_{ij}}{N}), \tag{4.5}$$

  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 pelo número de pixels da sub-região na horizontal e na vertical, respectivamente;

- Concavidades: Inicialmente, são extraídos os pontos extremos do contorno externo da parte da palavra em análise. Em seguida, os ângulos definidos por dois segmentos de reta, traçados entre o ponto inferior mínimo e os pontos mais à direita e mais à esquerda mínimos 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 palavra na 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 do classificador podem afetar o seu desempenho, deste modo preferiu-se atribuir o valor $0, 001$.

Características direcionais (PF-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 (1997). Neste método, para cada *pixel* do fundo da imagem, é verificado em cada uma das quatro direções principais (Norte, Sul, Leste e Oeste) se um *pixel* preto pode ser encontrado, como ilustrado na Figura 4.8 (OLIVEIRA Jr. et al., 2002; OLIVEIRA Jr., 2002).

Em função deste teste e dependendo da combinação das direções de abertura, os *pixels* do fundo da imagem são rotulados pela convenção apresentada na Tabela 4.1. 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 a cada rótulo, normalizados pela área da sub-região. Quando não existe *pixels* de um determinado rótulo o valor mapeado para o vetor é 0,001.

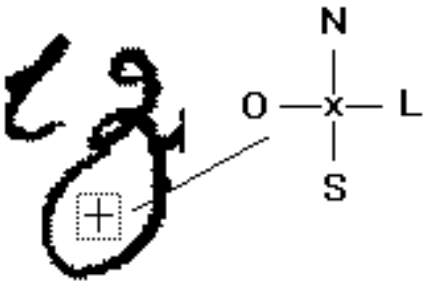


Figura 4.8: Exemplo da detecção das direções de abertura.

Tabela 4.1: 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

Características topológicas (PF-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 disso, as partes superior e inferior são divididas em 4 zonas cada uma, como apresentado na Figura 4.9 (OLIVEIRA Jr. et al., 2002; OLIVEIRA Jr., 2002).

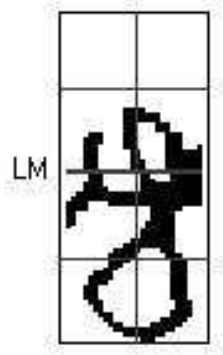


Figura 4.9: Exemplo da divisão em zonas realizada no conjunto de características topológicas.

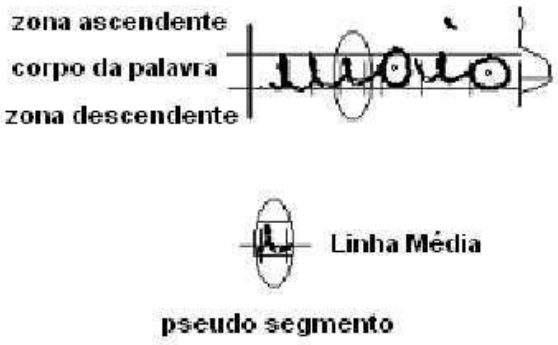
As componentes do vetor de características $(x_1, \cdots, x_8)$ são obtidas contando o número de *pixels* pretos em cada uma das oito zonas, normalizados pela respectiva área da zona. As componentes (x_9, x_{10}) correspondem às coordenadas do centro de massa do segmento da imagem em cada sub-região, normalizadas pelo número de pixels da sub-região na horizontal e na vertical, respectivamente. Quando o número de *pixels* pretos é zero, o valor mapeado para o vetor é 0,001.

4.3.3 Pseudo-segmentação variável (PV)

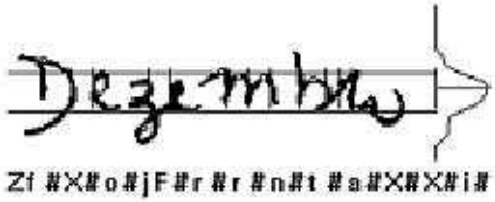
O método de reconhecimento descrito nesta seção foi desenvolvido por Freitas (2001). O processo de pseudo-segmentação aplicado neste método define um número variável de segmentos de tamanho variado, isto é feito considerando-se as transições branco-preto presentes na linha média (LM) da palavra. De modo que um segmento é definido a partir de duas transições branco-preto consecutivas, desde que não se encontrem internas a laços no corpo da palavra. A linha média é determinada seguindo o mesmo procedimento apresentado na Seção 4.3.1. Na Figura 4.10-a é ilustrado um exemplo do processo de pseudo-segmentação utilizado.

As características utilizadas neste método são as mesmas empregadas nos conjuntos de características perceptivas e direcionais apresentadas na Seção 4.3.2, embora usando uma representação diferente, adaptada a este método. O conjunto de características definido considera características direcionais baseadas na análise de concavidades e convexidades bem como características perceptivas (ascendentes, descendentes e laços). Concavidades e convexidades no corpo da palavra são extraídas analisando-se as direções de abertura dos *pixels* do fundo da imagem. Deste modo, após o processo de pseudo-segmentação, as características são extraídas dos segmentos, que são então rotuladas de acordo com a Tabela 4.2, sendo estabelecido grafemas

(sequências de símbolos que representam as características extraídas de um pseudo-segmento) para cada segmento, que formam uma sequência de observações para cada imagem. No caso de nenhuma característica ser extraída do segmento analisado, um símbolo vazio denotado por X é emitido. Na Figura 4.10-b é ilustrado um exemplo de geração da sequência de grafemas pelo processo de extração de características.



(a)



(b)

Figura 4.10: Exemplo do processo de extração de características utilizado no processo de pseudo-segmentação variável (PV): (a) Determinação das zonas e pseudo-segmentação e (b) geração dos grafemas.

De modo a selecionar um sub-conjunto ótimo de grafemas, um critério de Informação Mútua foi usado para definir o alfabeto de símbolos. Este critério baseia-se no conteúdo de informação de cada característica extraída e na ocorrência de combinações dessas características no mesmo pseudo-segmento. O alfabeto completo é composto de 29 símbolos diferentes selecionados usando o critério de Informação Mútua, levando em consideração todas as possíveis combinações dos símbolos (FREITAS; BORTOLOZZI; SABOURIN, 2001, 2004a).

A Informação Mútua é uma medida da informação que uma variável aleatória X têm sobre

Tabela 4.2: Convenção utilizada no processo de extração de características utilizado no método de pseudo-segmentação variável (PV).

Conjunto de Características	Característica [Simbolo]
Características perceptivas	Ascendente pequeno [t] e grande [T]
	Descendente pequeno [f] e grande [F]
	Laço superior [l] e inferior [j]
	Laço no corpo da palavra grande [O] e pequeno [o]
Análise de concavidades e convexidades	Concavidade aberta à direita [(] e aberta à esquerda [)]
	Convexidade aberta à direita [C] e aberta à esquerda [Z]
	Convexidade aberta abaixo [n] e aberta acima [u]
	Falso laço [a] no corpo da palavra
	Ligação abaixo [i]
	Ligação acima [r]
Pseudo-segmento vazio	Vazio [X]

uma segunda variável aleatória Y. Isto significa uma redução no conteúdo de informação de uma variável aleatória devido ao conhecimento da outra. Deste modo, a informação mútua denotada por $I(X,Y)$, descrita em Cover e Thomas (1991), é a entropia relativa entre a distribuição conjunta $p(x,y)$ e o produto das distribuições marginais $p(x)p(y)$ das variáveis aleatórias X e Y, definida como:

$$I(X,Y) = \sum_{i=1}^{N}\sum_{j=1}^{M} p(x_i,y_j) log_2(\frac{p(x_i,y_j)}{.} p(x_i)p(y_j)) \tag{4.6}$$

No caso em análise, a informação mútua mede a quantidade de informação distribuída sobre o conjunto de características extraído das imagens das palavras. A seguir, será feita uma discussão teórica sobre os classificadores utilizados na concepção do sistema.

4.4 Caracterização dos classificadores

Os métodos tradicionais de reconhecimento de padrões dividem a tarefa de reconhecimento em duas partes: inicialmente, um conjunto de características previamente definidas são extraídas das imagens, sendo em seguida aplicadas a um classificador que determina probabilidades condicionais relativas a cada classe. Deste modo, diferentes conjuntos de características podem ser propostos, bem como diferentes classificadores podem ser desenvolvidos. Ou seja, o problema em

questão é determinar o melhor conjunto de características que se adeque ao melhor classificador para o problema em questão. Deste modo foram usadas, neste trabalho, as redes neurais e os modelos escondidos de Markov como ferramentas de classificação, que são discutidas a seguir.

4.4.1 Redes neurais

Os classificadores neurais se adequam bem a problemas não-lineares, que possuam interações complexas entre suas variáveis e com um número limitado de classes, o que sugere a sua aplicação no sistema de reconhecimento desenvolvido neste trabalho (ZHANG, 2000; MARINAI; GORI; SODA, 2005). 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 (SCHALKOFF, 1992).

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 cálculo simples (HAYKIN, 1996). O modelo do neurônio ilustrado na Figura 4.11 possui várias entradas (conjunto de *sinapses*), 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:

$$net_k = \sum_{j=1}^{N} x_j w_{kj}, \tag{4.7}$$

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.

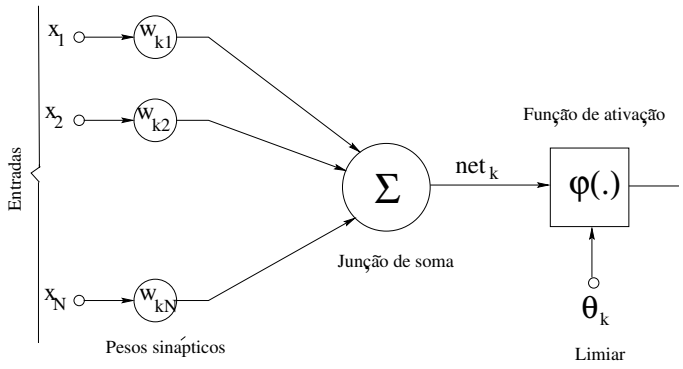


Figura 4.11: Modelo do neurônio utilizado em redes neurais artificiais.

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:

$$y = \varphi(net_k - \theta_k), \tag{4.8}$$

em que θ_k é um valor de limiar.

A função de ativação $\varphi(.)$ é responsável pela característica não-linear do classificador neural e 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).

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. Na Figura 4.12 é apresentado um exemplo da rede neural contendo três camadas.

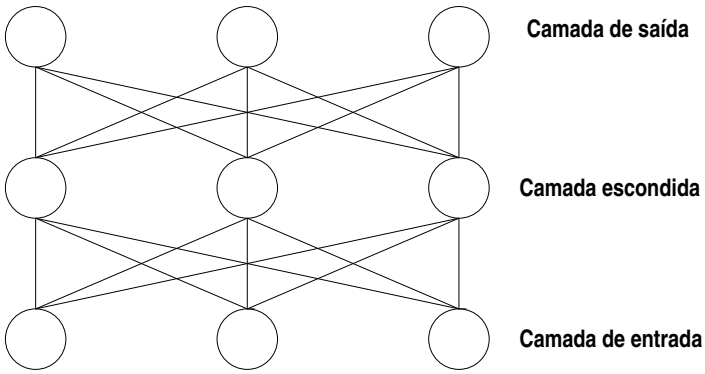


Figura 4.12: Arquitetura de uma rede neural com três camadas.

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 (HAYKIN, 1996; CORREIA, 2005).

Neste trabalho, em especial, foram utilizadas redes neurais multicamadas (MLP) treinadas com o algoritmo de retropropagação do erro com momento, que estão entre os mais difundidos e versáteis modelos de classificadores neurais. Este tipo de rede neural contendo uma camada intermediária e usando uma função de ativação não-linear é considerado um classificador universal (SUSSMANN, 1992), isto é, tais redes podem determinar limiares de decisão de complexidade

arbitrária. De modo que foi empregada a função sigmoidal como função de ativação expressa pela Equação 4.9,

$$\psi(net_i) = \frac{1}{1 + e^{net_i}} \tag{4.9}$$

que permite uma aproximação probabilista da saída da rede (RICHARD; LIPPMANN, 1991).

O algoritmo de retropropagação do erro com momento utilizado no treinamento da rede consiste dos seguintes passos (HAYKIN, 1996):

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:

   $$w_{ij}(n+1) = w_{ij}(n) + \eta\delta_j O_j + \alpha[w_{ij}(n) - w_{ij}(n-1)], \tag{4.10}$$

   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

   $$\delta_j = O_j(1 - O_j)(t_j - O_j), \tag{4.11}$$

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

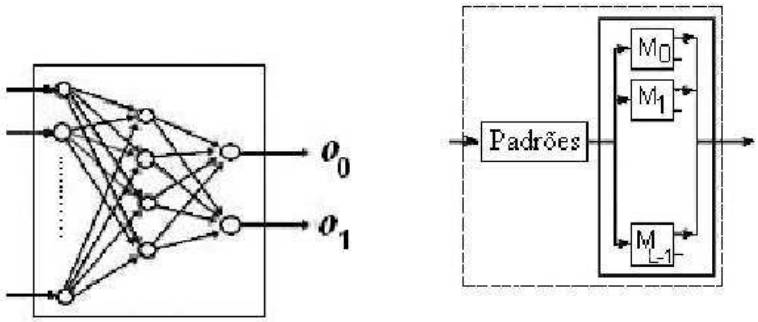
   $$\delta_j = O_j(1 - O_j)\sum_k \delta_k w_{kj}. \tag{4.12}$$

6. Repetir os passos 2 a 5 até que o critério de parada estabelecido seja satisfeito.

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 (HAYKIN, 1996).

Neste trabalho do mesmo modo como em Kapp (2004), além da utilização das redes neurais numa arquitetura convencional, como ilustrado na Figura 4.12, foi investigada uma estratégia denominada de arquitetura classe-modular. Nesta estratégia, uma tarefa única é decomposta em múltiplas sub-tarefas e cada uma dessas sub-tarefas é alocada para uma rede especialista. Nesta estratégia, como discutido em Oh e Suen (2002), um problema de L variáveis é decomposto em L subproblemas de 2 variáveis. Para cada uma das L classes, um classificador binário é especificamente projetado. Deste modo, o classificador binário discrimina esta classe das outras $L-1$ classes. No contexto classe-modular, L classificadores binários resolvem o problema original de L variáveis cooperativamente e um módulo de decisão integra as saídas dos L classificadores binários.

Na Figura 4.13-a, pode-se ver a arquitetura MLP para um classificador binário. O classificador modular MLP consiste de L sub-redes M_i, com $0 \leq i \leq L-1$, cada uma sendo responsável apenas por uma classe específica. A tarefa específica de cada M_i é selecionar entre dois grupos de classe $\Omega_0 = \{i\}$ e $\Omega_1 = \{l | 0 \leq l < L \text{ e } l \neq i\}$, ou seja, com apenas duas saídas, classificando se determinado exemplo pertence a classe ou não. As três camadas são totalmente conectadas. A camada de entrada tem d nós para aceitar o vetor de características d-dimensional e a camada de saída tem dois nós de saída, denotados por O_0 e O_1 para Ω_0 e Ω_1, respectivamente. A arquitetura completa da rede é mostrada na Figura 4.13-b. Após a extração de características, o vetor de padrões é usado por todas as L classes, sendo aplicado na camada de entrada de todas as sub-redes produzindo um vetor de saída $D = (O_0, O_1)$. Depois os valores de O_0 formam o vetor de saída final, sobre o qual é determinado o resultado do reconhecimento (KAPP, 2004).



(a) (b)

Figura 4.13: Arquitetura Classe-Modular, extraída de Oh e Suen (2002): (a) sub-rede e (b) rede completa com L módulos.

4.4.2 Modelos escondidos de Markov

Os Modelos Escondidos de Markov (Hidden Markov Models - HMM) são métodos estatísticos (redes estocásticas) que tem sido extremamente usados para modelamento sequencial de dados de comportamento variável, como encontrados em aplicações ligadas ao reconhecimento de voz e manuscritos.

Formalmente, um modelo escondido de Markov, como definido por Rabiner (RABINER, 1989), é um processo estocástico duplamente constituído, formado por um processo fundamental que não é observável (ou seja, escondido), mas que pode ser monitorado através de outro conjunto de processos estocásticos que produzem as sequências de observações. Isto significa que uma função probabilística de uma cadeia escondida de Markov é um processo estocástico gerado por dois mecanismos co-relacionados, uma cadeia de Markov que serve de base contendo um número finito de estados, e um conjunto de funções aleatórias, onde cada uma delas está associada a um estado específico. Em intervalos de tempo discretos, assume-se que o processo está em algum estado e uma observação é gerada por funções aleatórias correspondendo a este determinado estado. A cadeia de Markov interna modifica então seus estados de acordo com sua matriz de probabilidade de transições. O observador vê apenas a saída das funções aleatórias associadas a cada estado e não pode observar diretamente os estados da parte interna da cadeia de Markov; por isso, o termo modelo escondido de Markov.

Em princípio, a cadeia de Markov interna pode ser de qualquer ordem. Entretanto, ao longo do texto, as considerações irão se restringir à cadeias de Markov de primeira ordem, isto é, aquelas cuja a probabilidade de transição para qualquer estado depende apenas do referido estado e do seu antecessor.

A seguir serão feitas definições sobre HMMs discretos e os principais problemas associados a eles.

HMMs discretos: Definições e problemas associados

Muito do material e notações apresentadas nesta seção são adaptadas de Rabiner (RABINER, 1989). Considerando a existência de um número finito, digamos N, de estados no modelo. Em cada instante de tempo, temos a presença de um novo estado baseado na distribuição de probabilidades de transições que dependem do estado anterior (propriedade markoviana). Depois de cada transição, um simbolo de observação é produzido de acordo com uma distribuição de probabilidades, que depende do estado atual. A distribuição de probabilidade é mantida fixa independente de como ou quando o estado seja iniciado. Isto significa que as propriedades do

processo podem ser consideradas permanentes, exceto por pequenas flutuações, durante um certo intervalo de tempo e então, em certos instantes, ocorre uma mudança gradual para outro conjunto de propriedades.

Definimos a seguir a notação do modelo para um HMM de primeira ordem com observações discretas:

- T Tamanho da sequência de observação.
- N Número de estados no modelo.
- M Número de simbolos de observação.
- S $\{S_1, S_2, ..., S_N\}$, estados.
- Q $\{q_1 q_2 ... q_T\}$, sequência de estados.
- V $\{v_1, v_2, v_3, ..., v_M\}$, conjunto discreto de observações possíveis.
- q_t Estado presente no tempo t.
- A $\{a_{ij}\}$,$a_{ij} = P(q_{t+1} = S_j | q_t = S_i)$, distribuição das probabilidades de transição dos estados.
- B $\{b_j(k)\}$,$b_j(k) = P(v_k \ em \ t | q_t = S_j)$, distribuição de probabilidades de observação de simbolos no estado j.
- π $\{\pi_i\}$,$\pi_i = P(q_1 = S_i)$, distribuição do estado inicial.

Será usado a partir de agora, a notação compacta $\lambda = (A, B, \pi)$ para indicar o conjunto completo de parâmetros do modelo. Caracterizado a forma do modelo de Markov escondido λ, existem três problemas principais que precisam ser resolvidos, de modo que o modelo possa ser usado em aplicações do mundo real. Estes problemas são descritos a seguir.

Problema da classificação

A probabilidade de uma sequência observada $O = O_1, O_2, \cdots, O_T$ dado um modelo λ, $P(O|\lambda)$ pode ser usada para classificação. O caminho direto de se calcular $P(O|\lambda)$ é feito enumerando-se as sequências de estado possíveis. Assumindo a independência estatística entre as observações, têm-se que:

$$\begin{aligned} P(O|\lambda) = & \sum_{todo\ Q} P(O|Q, \lambda) = \\ = & \sum_{q_1, q_2, q_3, \cdots, q_T} \pi_{q_1} b_{q_1}(O_1) a_{q_1 q_2} b_{q_2}(O_2) \cdots a_{q_{T-1} q_T} b_{q_T}(O_T) \end{aligned} \tag{4.13}$$

Este método de calcular $P(O|\lambda)$ requer $O(TN^T)$ operações. Um método, denominado de processo de avanço-retrocesso requer $O(TN^2)$ operações.

Considere a variável de avanço $\alpha_t(i)$ definida como

$$\alpha_t(i) = P(O_1 O_2 \cdots O_t, q_t = S_i|\lambda). \tag{4.14}$$

Pode-se determinar $\alpha_t(i)$ indutivamente pelo algoritmo a seguir, denominado de algoritmo de $Forward$:

Inicialização: Para $1 \leq i \leq N$, faça $\alpha_1(i) = \pi_i b_i(O_1)$

Indução: Para $1 \leq t \leq T-1$ e $1 \leq j \leq N$, faça $\alpha_{t+1}(j) = [\sum_{i=1}^{N} \alpha_t(i) a_{ij}] b_j(O_{t+1})$

Término: $P(O|\lambda) = \sum_{i=1}^{N} \alpha_T(i)$

De modo similar, considera-se uma variável de retrocesso $\beta_t(i)$ definida como

$$\beta_t(i) = P(O_{t+1} O_{t+2} \cdots O_T|q_t = S_i, \lambda) \tag{4.15}$$

e novamente pode-se resolver $\beta_t(i)$ indutivamente pelo algoritmo a seguir, denominado de algoritmo de $Backward$.

Inicialização: Para $1 \leq i \leq N$, faça $\beta_T(i) = 1$.

Indução: Para $1 \leq t \leq T-1$ e $1 \leq i \leq N$, faça $\beta_t(i) = \sum_{j=1}^{N} a_{ij} b_j(O_{t+1}) \beta_{t+1}(j)$.

Término: Para $1 \leq t \leq T-1$, faça $P(O|\lambda) = \sum_{i=1}^{N} \sum_{j=1}^{N} \alpha_t(i) a_{ij} b_j(O_{t+1}) \beta_{t+1}(j)$.

Problema da sequência de estados ótima

Existem diversas formas de encontrar a sequência de estados ótima associada com uma dada sequência de observação. As dificuldades consistem na definição do que seria uma sequência ótima, isto é, existem vários critérios de otimização possíveis. Um deles seria escolher os estados que são individualmente mais prováveis. Este critério de otimização maximiza o número esperado de estados individualmente corretos.

Para implementar esta solução, define-se a variável

$$\gamma_t(i) = P(q_t = S_i|O, \lambda) \tag{4.16}$$

que pode ser calculada a partir de

$$\gamma_t(i) = \frac{\alpha_t(i)\beta_t(i)}{P(O|\lambda)} = \frac{\alpha_t(i)\beta_t(i)}{\sum_{j=1}^{N} \alpha_t(j)\beta_t(j)}. \tag{4.17}$$

Usando $\gamma_t(i)$, pode-se determinar qual o estado q_t mais provável no tempo t, $1 \leq t \leq T$, como

$$q_t = \arg\max\{\gamma_t(i)\}, \quad 1 \leq i \leq N. \tag{4.18}$$

O principal problema com o critério acima e sua solução ocorre quando existem transições que não são permitidas. Neste caso, a sequência de estados ótima obtida pode, de fato, ser uma sequência de estados impossível. Este fato mostra a necessidade de se definir limitações globais para a sequência de estados ótima obtida. Um critério de otimização deste tipo é encontrar a sequência de estados com maior probabilidade, isto é, maximizar $P(O, Q|\lambda)$. Uma técnica formal para encontrar esta solução existe e denomina-se de algoritmo de Viterbi.

Este algoritmo define uma variável

$$\delta_t(i) = \max_{q_1, q_2, \cdots, q_{t-1}} P(q_1, q_2, \cdots, q_t = S_i, O_1 O_2 \cdots O_t|\lambda). \tag{4.19}$$

Similarmente $\delta_{t+1}(i)$ pode ser determinada indutivamente de acordo com o seguinte procedimento:

Inicialização: Para $1 \leq i \leq N$, faça $\delta_1(i) = \pi_i b_i(O_1)$ e $\psi_1(i) = 0$.

Recursão: Para $2 \leq t \leq T$ e $1 \leq j \leq N$, faça $\delta_t(j) = \max_{1 \leq i \leq N}[\delta_{t-1}(i)a_{ij}]b_j(O_t)$ e $\psi_t(j) = \arg\max_{1 \leq i \leq N}[\delta_{t-1}(i)a_{ij}]$.

Término: $P^* = \max_{1 \leq i \leq N}[\delta_T(i)]$;
$q_T^* = \arg\max_{1 \leq i \leq N}[\delta_T(i)]$.
Para todo $1 \leq t \leq T-1$, faça $q_t^* = \psi_{t+1}(q_{t+1}^*)$.

Problema de treinamento

Dada qualquer sequência de observação finita como dados de treinamento, não se pode treinar o modelo de forma ótima. Pode-se contudo, escolher A, B, e π tal que $P(O|\lambda)$ seja localmente maximizado. O método de Baum-Welch é um algoritmo iterativo que usa as probabilidades de avanço e retrocesso para resolver o problema de treinamento por estimação dos parâmetros.

Para implementar a solução, define-se inicialmente a variável $\gamma_t(i)$, como sendo a probabilidade de estar no estado S_i no tempo *t* e então define-se $\xi_t(i,j)$ a probabilidade de se estar no estado S_i no tempo *t* e no estado S_j no tempo *t+1*, dado o modelo e a sequência de observação, isto é,

$$\gamma_t(i) = P(q_t = S_i|O,\lambda) = \frac{\alpha_t(i)\beta_t(i)}{P(O|\lambda)}; \tag{4.20}$$

$$\xi_t(i,j) = P(q_t = S_i, q_{t+1} = S_j|O,\lambda) = \frac{\alpha_t(i)a_{ij}b_j(O_{t+1})\beta_{t+1}(j)}{P(O|\lambda)}. \tag{4.21}$$

Agora, têm-se que $\sum_{t=1}^{T-1}\xi_t(i,j)$ é igual ao número esperado de transições feitas de S_i para S_j e $\sum_{t=1}^{T-1}\gamma_t(i)$ é o número esperado de transições de S_i.

As formulas de re-estimação de Baum-Welch para A, B e π são:

$$\bar{\pi}_i = \gamma_1(i) \tag{4.22}$$

$$\bar{a}_{ij} = \frac{\sum_{t=1}^{T-1}\xi_t(i,j)}{\sum_{t=1}^{T-1}\gamma_t(i)} \tag{4.23}$$

$$\bar{b}_j(k) = \frac{\sum_{t=1,O_t=k}^{T}\gamma_t(j)}{\sum_{t=1}^{T}\gamma_t(j)}. \tag{4.24}$$

Aplicações iterativas destas fórmulas convergem para o máximo local de $P(O|\lambda)$.

Tipos de HMMs

Até o momento, se tem considerado somente o caso especial de HMM's ergódicos ou completamente conectados, no qual qualquer estado do modelo pode ser alcançado (num único passo) de qualquer outro estado do modelo. (Estritamente falando, um modelo ergódico tem a propriedade de que qualquer estado pode ser alcançado de qualquer outro estado em um número finito de passos).

Um outro tipo de modelo de HMM muito utilizado é o modelo *left-right* (esquerda-direita) ou modelo de Bakis, que tem esta denominação porque a sequência associada com o modelo tem a propriedade de que quando o tempo passa, o índice do estado também aumenta, ou permanece o mesmo.

No modelo do HMM *left-right*, os coeficientes de transições de estado possuem a seguinte propriedade:

$$a_{ij} = 0, \quad j < i \tag{4.25}$$

isto é, não são permitidas transições para estados cujos índices são menores do que o estado atual.

Além disso, as probabilidades de estado inicial tem as seguintes propriedades:

$$\pi_i = \begin{cases} 0, & i \neq 1 \\ 1, & i = 1 \end{cases} \tag{4.26}$$

desde que a sequência de estado começa no estado 1 e termina no estado N.

Geralmente restrições adicionais são impostas aos coeficientes de transição dos estados para garantir que grandes alterações nos índices dos estados não ocorram, assim uma restrição da forma

$$a_{ij} = 0, \quad j > i + \Delta \tag{4.27}$$

é geralmente utilizada.

Embora tenha-se dividido os HMMs nestes dois modelos, há diversas variações e combinações possíveis. Vale salientar ainda que as limitações do modelo esquerda-direita não afetam o processo de reestimação.

Considerações sobre a implementação dos HMMs

As discussões anteriores tiveram como enfoque a teoria de HMMs e várias variações na forma do modelo. Será feita agora uma discussão em relação as dificuldades de implementação destes modelos, incluindo escalonamento, estimação dos valores iniciais, e escolha do tipo e tamanho do modelo.

Escalonamento

Para se compreender porque requere-se o escalonamento para a implementação do processo de reestimação dos HMMs, considere as definições de $\alpha_t(i)$. Pode se ver que $\alpha_t(i)$ consiste da soma de um grande número de termos, da forma

$$\left(\prod_{s=1}^{t-1} a_{q_s q_{s+1}} \prod_{s=1}^{t} b_{q_s}(\mathbf{O}_s) \right)$$

com $q_t = S_i$. Como a e b são menores que 1, vê-se que quando t se torna muito grande (em geral, 10 ou mais), cada termo de $\alpha_t(i)$ decaí exponencialmente para zero. Para t suficientemente grande, da ordem de 100 ou mais, a faixa dinâmica do cálculo de $\alpha_t(i)$ excede a faixa de precisão de qualquer máquina. Assim, um modo de fazer o cálculo é incorporando-se um procedimento de escalonamento.

Para compreender melhor este processo de escalonamento, considere a fórmula de reestimação para os coeficientes de transição de estado a_{ij}. Escrevendo a equação de reestimação diretamente

em termos das variáveis de avanço e retrocesso tem-se

$$\bar{a}_{ij} = \frac{\sum_{t=1}^{T-1} \alpha_t(i) a_{ij} b_j(O_{t+1}) \beta_{t+1}(j)}{\sum_{t=1}^{T} \sum_{j=1}^{N} \alpha_t(i) a_{ij} b_j(O_{t+1}) \beta_{t+1}(j)}. \tag{4.28}$$

Considere o cálculo de $\alpha_t(i)$. Para cada t, inicialmente determinamos $\alpha_t(i)$ de acordo com a fórmula de indução mostrada anteriormente, e então multiplica-se pelo coeficiente de escala c_t, onde

$$c_t = \frac{1}{\sum_{i=1}^{N} \alpha_t(i)}. \tag{4.29}$$

Então, para um t fixo, inicialmente determina-se

$$\alpha_t(i) = \sum_{j=1}^{N} \hat{\alpha}_{t-1}(j) a_{ij} b_j(O_t). \tag{4.30}$$

Então, o conjunto de coeficientes escalonados $\hat{\alpha}_t(i)$ é determinado como

$$\hat{\alpha}_t(i) = \frac{\sum_{j=1}^{N} \hat{\alpha}_{t-1}(j) a_{ij} b_j(O_t)}{\sum_{i=1}^{N} \sum_{j=1}^{N} \hat{\alpha}_{t-1}(j) a_{ij} b_j(O_t)}. \tag{4.31}$$

Por indução pode-se escrever $\hat{\alpha}_{t-1}(j)$ como

$$\hat{\alpha}_{t-1}(j) = \left(\prod_{\tau=1}^{t-1} c_\tau \right) \alpha_{t-1}(j). \tag{4.32}$$

Então, escreve-se $\hat{\alpha}_t(i)$ como

$$\hat{\alpha}_t(i) = \frac{\sum_{j=1}^{N} \alpha_{t-1}(j) \left(\prod_{\tau=1}^{t-1} c_\tau \right) a_{ij} b_j(O_t)}{\sum_{i=1}^{N} \sum_{j=1}^{N} \alpha_{t-1}(j) \left(\prod_{\tau=1}^{t-1} c_\tau \right) a_{ij} b_j(O_t)} = \frac{\alpha_t(i)}{\sum_{i=1}^{N} \alpha_t(i)} \tag{4.33}$$

isto é, cada $\alpha_t(i)$ é efetivamente escalonado pela soma de todos os estados de $\alpha_t(i)$.

Em seguida, determina-se os termos $\beta_t(i)$. A diferença neste caso é que se utiliza os mesmos fatores de escala para cada tempo t para os β's como os que foram usados para os α's. Assim, os β's são da forma

$$\hat{\beta}_t(i) = c_t \beta_t(i) \tag{4.34}$$

Desenvolvendo-se as expressões anteriores prova-se que apesar do escalonamento, as fórmulas de reestimação são preservadas. Este processo também se aplica aos coeficientes π ou B.

A única mudança no processo do HMM que ocorre pela utilização do escalonamento é no cálculo de $P(O|\lambda)$. Pois, não se pode meramente somar $\hat{\alpha}_t(i)$, desde que estão escalonados.

Contudo, pode se usar a propriedade que

$$\prod_{t=1}^{T} c_t \sum_{i=1}^{N} \alpha_T(i) = C_T \sum_{i=1}^{N} \alpha_T(i) = 1. \tag{4.35}$$

Então se tem

$$\prod_{t=1}^{T} c_t \cdot P(O|\lambda) = 1 \tag{4.36}$$

ou

$$P(O|\lambda) = \frac{1}{\prod_{t=1}^{T} c_t} \tag{4.37}$$

ou

$$\log[P(O|\lambda)] = -\sum_{t=1}^{T} \log c_t. \tag{4.38}$$

Então o logaritmo de $P(O|\lambda)$ pode ser determinado, porém não se pode calcular seu valor exato pois ele é muito pequeno.

Estimativas iniciais dos parâmetros do HMM

Na teoria, as equações de reestimação fornecem valores dos parâmetros do HMM que correspondem ao máximo local de uma função de verossimilhança. Uma questão chave é como escolher estimativas iniciais dos parâmetros do HMM tal que o máximo local seja na verdade o máximo global da função de verossimilhança.

Basicamente, não existe uma maneira simples e direta de responder esta questão. Ao invés disso, se tem mostrado que estimativas iniciais com distribuições aleatórias ou uniformes dos parâmetros π e A são adequadas na maioria dos casos. Contudo, para os parâmetros B, nota-se a necessidade de boas estimativas, que ajudam no caso de simbolos discretos, e que são essenciais no caso de distribuições contínuas. Para isso utiliza-se diversas técnicas de segmentação dos vetores de observação, como por exemplo, segmentação das observações usando máxima verossimilhança pela média, entre outras.

Escolha do modelo

O último problema na implementação dos HMMs é a escolha do tipo do modelo (ergódico ou esquerda-direita), escolha do tamanho do modelo (número de estados), e escolha dos simbolos de observação (discretos ou contínuos). Infelizmente, não existe um caminho simples, ou teoricamente correto, de fazer estas escolhas, pois elas dependem basicamente do sinal a ser modelado. Com estes comentários, encerra-se a discussão dos aspectos teóricos dos modelos escondidos de Markov.

4.5 Combinação de classificadores

O objetivo final de qualquer sistema de reconhecimento de padrões é obter a melhor classificação possível para a tarefa em questão. Deste modo, são avaliados diferentes esquemas de classificação buscando a melhor solução para o problema. O resultado experimental da aplicação desses esquemas forma uma base para que se escolha o classificador que mais se adeque ao problema. Foi observado em estudos de classificação (XU; KRZYZAK; SUEN, 1992; KITTLER et al., 1998), que embora um dos projetos atinja o melhor desempenho individual, os conjuntos de padrões rotulados erroneamente pelos diferentes classificadores não são necessariamente sobrepostos. Isto sugere que diferentes esquemas de classificação oferecem informações complementares, sobre os padrões a serem classificados, que podem melhorar o desempenho do classificador selecionado.

Estas observações motivaram o interesse na combinação de classificadores. A idéia não é obter um esquema único de decisão. Ao invés disso, todos os projetos, ou seus subconjuntos, são usados para a tomada de decisão combinando-se suas opiniões individuais de modo a se obter uma decisão de consenso. Diversos esquemas de combinação de classificadores foram desenvolvidos e demonstrou-se experimentalmente que alguns deles conseguem superar o desempenho obtido por um único classificador ótimo (HO; HULL; SRIHARI, 1994; LAM; SUEN, 1995; KITTLER et al., 1998).

Basicamente, pode-se definir o problema de combinação de classificadores em dois cenários distintos. No primeiro cenário, todos os classificadores usam a mesma representação dos padrões de entrada. Um exemplo típico desta categoria é um conjunto de classificadores k-NN, cada um usando o mesmo vetor de características, mas diferentes parâmetros de classificação. Outro exemplo é um conjunto de classificadores neurais de arquitetura fixa, mas apresentando conjuntos de pesos distintos obtidos por diferentes estratégias de treinamento. Neste caso, cada classificador, para um dado padrão de entrada, pode ser considerado como produtor de uma estimativa da mesma probabilidade *a posteriori* da classe.

No segundo cenário, cada classificador usa sua própria representação dos padrões de entrada. Em outras palavras, as características extraídas dos padrões são únicas para cada classificador. Uma aplicação importante da combinação de classificadores neste cenário é a possibilidade de integrar medidas/características fisicamente diferentes. Neste caso, não é possível considerar o cálculo de probabilidades *a posteriori* como estimativas do mesmo valor funcional, já que os sistemas de classificação operam em diferentes espaços de características.

4.5.1 Definição da combinação de múltiplos classificadores

Xu, Krzyzak e Suen (1992) apresentam a seguinte definição da combinação de múltiplos classificadores. Dado um espaço de padrões P consistindo de M conjuntos mutuamente exclusivos $P = L_1 \cup \cdots \cup L_M$ em que cada L_i, $\forall i \in \Lambda = \{1, 2, \cdots M\}$ representa um conjunto de padrões específicos denominado classe. Para uma amostra x de P, a tarefa do classificador e é atribuir a x um índice $j \in \Lambda \cup \{M+1\}$ como um rótulo para representar que x é observado como sendo da classe L_j se $j \neq M+1$, em que $j = M+1$ denota que x é rejeitado por e. Desconsiderando a estrutura interna do classificador, bem como a teoria e metodologia em que se baseia, pode-se observar um classificador como uma função que recebe uma amostra de entrada x e retorna um rótulo j, ou $e(x) = j$.

Embora j seja a informação desejada no estágio final de classificação, muitos dos algoritmos de classificação existentes são capazes de fornecer outras informações complementares. De fato, o rótulo final j é o resultado da melhor seleção entre os M valores e esta seleção certamente descarta alguma informação que é desconsiderada no resultado final quando há somente um classificador. Contudo, algumas informações podem ser úteis para a combinação de múltiplos classificadores. Deste modo a informação de saída pode ser dividida em três níveis:

- Nível abstrato - O classificador e apresenta na saída apenas o rótulo ao qual o padrão apresentado fora classificado, considerando um subconjunto $J \subset \Lambda$.
- Nível posto - O classificador e ordena todos os rótulos em Λ ou um subconjunto $J \subset \Lambda$ em uma dada ordem de modo que o rótulo de melhor posição representa a melhor escolha.
- Nível de medição - O classificador e atribui a cada rótulo em Λ um valor que representa a verossimilhança de x em relação aquele rótulo.

A partir dos três níveis de saída mencionados anteriormente, pode-se resumir o problema de combinação de múltiplos classificadores em três tipos:

- Tipo 1: A combinação é feita com base na informação de saída do nível abstrato. Dado C classificadores individuais e_c, $c = 1, \cdots, C$, cada um dos quais atribui a entrada x um rótulo j_c, ou seja, $e_c(x) = j_c$. O problema é construir um classificador E que integre esses eventos, dando à x um rótulo definitivo j, isto é $E(x) = j$, $j \in \Lambda \cup \{M+1\}$.
- Tipo 2: A combinação é feita com base na informação de saída do nível posto. Para uma entrada x, cada e_c produz um subconjunto $S_c \subseteq \Lambda$ com todos os rótulos em S_c ordenados em uma lista. O problema é usar estes eventos $e(x) = S_c$, $c = 1, \cdots, C$ para construir um classificador E em que $E(x) = j$, $j \in \Lambda \cup \{M+1\}$.

- Tipo 3: A combinação é feita com base na informação de saída do nível de medição. Para uma entrada x, cada e_c produz um vetor real $M_e(c) = [m_c(1), \cdots, m_c(M)]^t$, em que $m_c(i)$ denota a probabilidade que e_c atribui à x em relação ao rótulo i. O problema é usar estes eventos $e_c(x) = M_e(c)$, $c = 1, \cdots, C$ para construir um classificador E em que $E(x) = j$, $j \in \Lambda \cup \{M + 1\}$.

Quanto à regra de combinação, os combinadores são classificados como baseados em regras fixas ou estáticas e baseados em treinamento. Os primeiros utilizam regras definidas *a priori*, enquanto os outros requerem um treinamento prévio para definição da regra. Uma análise da teoria de múltiplos classificadores é apresentada nos trabalhos de Xu, Krzyzak e Suen (1992), Kittler et al. (1998), Matos (2004) e Webb (2002).

4.5.2 Diversidade *versus* múltiplos classificadores

Um dos principais problemas envolvendo combinação de classificadores é a existência de dependência entre os mesmos. Atualmente, existem muitas discussões sobre técnicas que tem por objetivo gerar classificadores descorrelacionados. Isto introduz um conceito muito debatido recentemente denominado diversidade, que parte do princípio de que não há ganho ao se combinar classificadores idênticos. Zouari (2004) afirma que a combinação de classificadores somente é eficaz se classificadores diferentes forem também independentes. É discutido também que, classificadores de menor desempenho que cometam erros diferentes entre si possuem melhor resultado na combinação do que classificadores de maior desempenho que cometam erros idênticos. Sendo assim, quanto mais diversos são os classificadores, melhor será o resultado da combinação.

Diversos trabalhos (BROWN et al., 2005; TSYMBAL; PECHENIZKIY; CUNNINGHAM, 2005; RUTA; GABRYS, 2005) supõem que a independência de classificadores é uma hipótese necessária e até mesmo obrigatória para se obter uma melhora significativa do desempenho, embora ainda não haja experimentos suficientes que apóiem essa condição. Desde que, é possível obter resultados interessantes e até mesmo melhores ao se combinar classificadores dependentes quando comparados à classificadores independentes. Na verdade, a combinação ideal é aquela que compõe classificadores de alto desempenho e o máximo possível discordantes. Ou seja, não se pode estudar diversidade dentro de um arranjo de classificadores sem levar em consideração seus desempenhos individuais.

É necessário determinar uma solução de compromisso entre diversidade e desempenho, porém ainda não existe nenhum estudo teórico que mostre uma medida confiável que relacione estes dois conceitos. Estudos recentes (KUNCHEVA; WHITAKER, 2003; WINDEATT, 2005; AIRES, 2005) procuram avaliar medidas que possam estabelecer uma solução para este dilema, não

havendo entretanto resultados conclusivos. As medidas de diversidade podem ser divididas em dois tipos:

- Medidas *pairwise* - Estes métodos calculam a média de uma métrica particular entre todos os possíveis pares de classificadores. A métrica utilizada determina as características da medida de diversidade. Geralmente, o cálculo da métrica utiliza a contagem das concordâncias ou discordâncias dos classificadores.

- Medidas *non-pairwise* - Estes métodos usam a entropia ou outras medidas similares para calcular a correlação entre cada classificador individual e o resultado médio da combinação.

Além de se medir a diversidade, ou de se determinar o quanto de diversidade é necessário para se obter um desempenho ótimo da combinação dos classificadores, pode-se discutir também as estratégias para geração de diversidade. Recentemente, Duin (2002) apresentou uma lista das principais estratégias, classificando-as em ordem crescente de prioridade, como mostrado a seguir:

- **Inicializações diferentes** - Inicializar os classificadores de modos diferentes para obter saídas diferentes. Isto pode ser aplicado especialmente em classificadores neurais.

- **Escolha de parâmetros** - Variações no número de neurônios na rede neural, por exemplo.

- **Arquiteturas diferentes** - Considerando classificadores neurais, pode-se usar redes MLP, RBF, entre outras.

- **Estruturas diferentes** - Em certos casos, os pesquisadores necessitam utilizar o mesmo espaço de características ou a mesma base de aprendizagem. A fim de evitar a redundância na tomada de decisão, eles utilizam classificadores de diferentes estruturas. Na literatura existem diversos tipos de classificadores, por exemplo, redes bayesianas, classificador gaussiano, árvores de decisão, modelos escondidos de Markov.

- **Bases de aprendizagem diferentes** - Esta estratégia consiste em formar sub-bases com dados diferentes a partir da base original. As técnicas mais utilizadas para este fim são conhecidas como *bagging* e *boosting* e são amplamente discutidas na literatura (WEBB, 2002; MATOS, 2004). Mas se pode também construir grupos de dados específicos para cada classe usando técnicas de agrupamento.

- **Características diferentes** - Utilização de uma ou mais famílias de características adaptadas a cada classificador pode produzir saídas diferentes.

Em resumo, existem diversas formas de se construir classificadores independentes embora ainda não se possa afirmar qual é a melhor. Do mesmo modo, o conceito de diversidade é muito importante na definição de classificadores individuais dentro de um ambiente de múltiplos classificadores, embora ainda não se tenha uma medida que determine como relacionar corretamente diversidade e desempenho.

Trazendo essas informações para a arquitetura de análise multi-vistas pode-se ver que sua definição é direcionada à geração de diversidade, pois embora utilize a mesma base de dados como será descrito posteriormente, as arquiteturas são definidas usando estruturas e características diferentes. Sendo assim é possível fazer análises diferentes da mesma amostra, que possam ser complementares, de modo que ao final o combinador determine a melhor solução em função dessas diferentes observações.

4.6 Conclusão

Neste capítulo, foi apresentado o sistema de reconhecimento de palavras manuscritas desenvolvido neste trabalho. Este sistema é definido a partir de observações provenientes dos estudos sobre o processo de leitura humano que dão suporte à arquitetura de análise multi-vistas proposta. Esta arquitetura define três diferentes estratégias de pseudo-segmentação que determinam um sistema de múltiplos classificadores. Para cada estratégia foram apresentados o conjunto de características e o método de classificação empregado. As características extraídas são aquelas mais comumente utilizadas no processo de leitura humano, de modo que o sistema é direcionado à uma aproximação computacional dos mecanismos de percepção.

Os sistemas de múltiplos classificadores buscam combinar os resultados de vários classificadores individuais de modo que a combinação apresente um desempenho global melhor do que aquele obtido pelo melhor classificador individual. Vários autores sugerem diversas formas de se obter este classificador ótimo considerando diferentes níveis de aproximações das saídas dos classificadores. Porém ainda não foi estabelecida uma regra única que defina isso. Entretanto, atualmente, o conceito de diversidade e sua relação com a geração de sistemas de múltiplos classificadores tem sido explorada neste intuito.

No capítulo seguinte, a metodologia experimental de avaliação do sistema será apresentada, mostrando a base de dados utilizada, os testes efetuados e os resultados obtidos.

Capítulo 5

Metodologia de testes e resultados obtidos

Neste capítulo é feita uma análise de desempenho do sistema descrito no Capítulo 4, baseado na arquitetura de análise multi-vistas proposta. A avaliação considera o comportamento do sistema tanto isoladamente como a interação dessas partes na concepção do sistema completo. Inicialmente, é feita uma análise da construção da base de dados utilizada e sua caracterização em relação aos estilos de escrita encontrados. Em seguida é feita uma análise dos resultados obtidos na aplicação das técnicas de pré-processamento apresentadas na Seção 4.2. Após isso é discutida a metodologia utilizada nos experimentos e os resultados obtidos considerando a análise individual e combinada dos classificadores. Também apresenta-se uma nova medida de discordância para avaliação da diversidade dos classificadores. Ao final é feita uma comparação com outros sistemas descritos na literatura.

5.1 Base de dados

Para desenvolver os experimentos foi utilizada a base de dados disponível no Laboratório de Análise e Processamento de Sinais (LAPS) da UFCG. Esta base foi construída de modo a representar da melhor forma possível os diferentes estilos de escrita presentes no nosso vocabulário. Aproximadamente 60% dos formulários foram coletados na cidade de Campina Grande - PB, e o restante na cidade de Curitiba - PR como parte do projeto PROCAD UFCG/PUCPR. De modo que foram coletadas 850 observações de cada mês, de escritores de diferentes níveis de educação, localizados em diferentes regiões do país.

Cada escritor preencheu um formulário específico, em papel sulfite branco, em que cada

palavra correspondente a cada mês deveria ser escrita uma vez. Nenhuma restrição foi imposta em relação ao estilo de escrita utilizado, nem foi sugerido qualquer modelo prévio para a escrita da palavra, o que resultou numa base de dados heterogênea. Não se têm conhecimento da existência de outra base para esse dicionário com volume de dados e/ou perfil semelhante.

As palavras escritas nos formulários foram digitalizadas usando um instrumento de aquisição (*scanner*) da marca HP Scanjet 5200 C (SERIES, 1998) na resolução de 200 DPI (*dots per inch*) com dois níveis de cinza, sendo armazenadas em formato PCX. Na Figura 5.1 são ilustradas algumas amostras contidas nesta base de dados.

A base de dados completa é constituída de 10.200 palavras, sendo 850 para cada classe. Para ser utilizada nos experimentos, ela foi dividida aleatoriamente em três conjuntos: Conjunto 1 - Base de treinamento, com 6.120 palavras (60% do total); Conjunto 2 - Base de validação, com 2.040 palavras (20% do total) e Conjunto 3 - Base de teste, com 2.040 palavras (20% do total). Para cada conjunto, as palavras são igualmente distribuídas entre as classes.

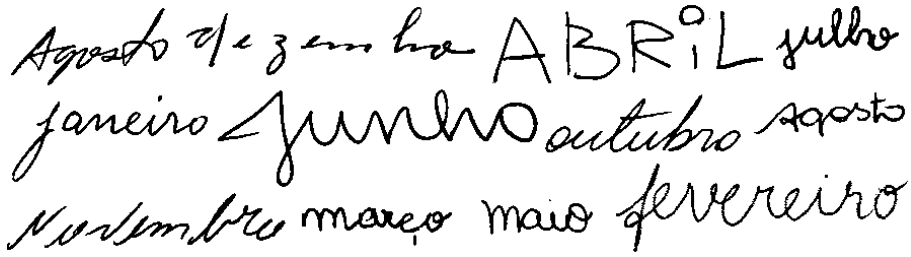


Figura 5.1: Amostras da base de dados utilizada.

5.1.1 Caracterização da base de dados

Para caracterizar a base de dados foi feita uma análise com relação aos estilos de escrita encontrados nela. Segundo Tappert, Suen e Wakahara (1990) pode-se classificar a escrita cursiva em cinco categorias principais, conforme Figura 5.2:

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;
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.

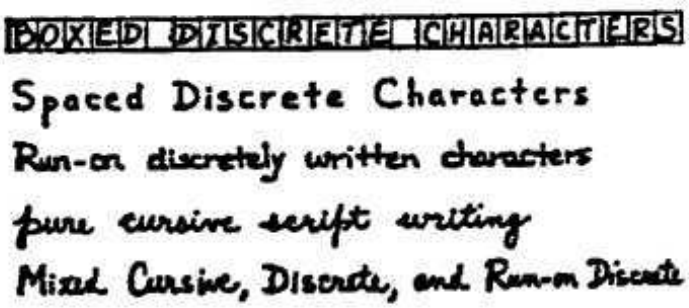


Figura 5.2: Tipos de escrita segundo a classificação de Tappert (extraída de Tappert, Suen e Wakahara (1990)).

Freitas (2001) considera que a categoria 3 insere-se na categoria 5, classificando as palavras em quatro grupos: Cursiva pura, caixa alta, caracteres disjuntos e mista. Seguindo esta classificação, foi realizada uma análise subjetiva da base de dados quanto à distribuição dos tipos de escrita presentes nas bases de treinamento, validação e teste. Esse levantamento é apresentado na Tabela 5.1.

Tabela 5.1: Distribuição dos tipos de escrita nos subconjuntos da base de dados utilizada.

	Treinamento	**Validação**	**Teste**
Cursiva pura	58 %	55 %	53 %
Caixa alta	4 %	8 %	7 %
Caracteres disjuntos	8 %	11 %	8 %
Mista	30 %	26 %	32 %

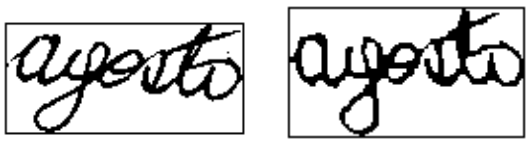
Outro levantamento realizado foi a porcentagem de palavras com a primeira letra maiúscula, sendo determinado um percentual de 28%, 27% e 31% para os conjuntos de treinamento, validação e teste, respectivamente.

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 possuem um nível de confusão elevado.

Essa análise comprova que a base de dados construída apresenta um grau elevado de diversidade em relação aos estilos de escrita, o que atende aos requisitos necessários para uma boa avaliação do sistema.

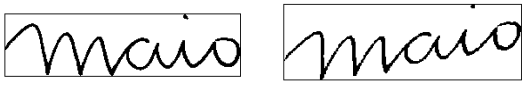
5.2 Análise dos resultados do pré-processamento

Na Seção 4.2 foram apresentadas as técnicas de pré-processamento aplicadas sobre as imagens originais com o objetivo de diminuir a variação no estilo de escrita inerente a cada autor. Deste modo, após uma análise visual, constatou-se que os objetivos esperados pelo pré-processamento foram atingidos em 99% das imagens pré-processadas. Porém, como já previsto por Veloso (2001), 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 na Figura 5.3 que mostra uma imagem em que o pré-processamento obteve bom resultado e na Figura 5.4, um caso em que ocorreu o problema acima descrito.



(a) (b)

Figura 5.3: Exemplo de pré-processamento adequado: (a) imagem original e (b) imagem pré-processada.



(a) (b)

Figura 5.4: Exemplo de pré-processamento incorreto: (a) imagem original e (b) imagem pré-processada.

5.3 Análise dos classificadores isolados

5.3.1 Metodologia de testes

Para cada conjunto de características apresentado nas Seções 4.3.1 e 4.3.2, uma rede neural convencional e uma rede neural classe-modular com uma camada escondida foi treinada e testada. Inicialmente, foi utilizado o MATLAB como ambiente de simulação (DEMUTH; BEALE, 1998), pois dentre as regras de treinamento disponíveis, existe uma que utiliza regularização bayesiana (MACKAY, 1992; FORESEE; HAGAN, 1997), que otimiza automaticamente o número de parâmetros da rede. Porém, devido ao tamanho da base de dados, os recursos computacionais disponíveis não foram suficientes para fazer o treinamento da rede usando esta regra de aprendizagem nesse ambiente. Deste modo, se fez a opção de utilizar o ambiente de simulação SNNS (ALLI, 1994), criado por pesquisadores da Universidade de Stuttgart.

A partir disso, definiu-se empiricamente a quantidade de neurônios na camada escondida, sendo testadas diversas configurações. A melhor configuração escolhida foi aquela que apresentou o menor erro ao se testar a rede treinada sobre o conjunto de validação. Para o sistema baseado no processo de pseudo-segmentação de radical (PR), o número de neurônios foi variado entre 20 e 50, sendo que o melhor resultado foi obtido usando 40 neurônios. No caso do sistema baseado no processo de pseudo-segmentação fixa, o número de neurônios foi variado entre 70 e 120, para o conjunto de características perceptivas (PF-P), direcionais (PF-D) e topológicas (PF-T). As melhores configurações foram obtidas considerando 110, 95 e 110 neurônios, 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 critério de parada utilizado no treinamento foi o erro médio obtido sobre o conjunto de validação. Essa estratégia foi utilizada ao se observar que, a partir de um certo ponto, embora o erro médio sobre o conjunto de treinamento diminuí-se, o mesmo não ocorria sobre o conjunto de validação, que estabilizava num determinado valor. Esse tipo de procedimento é conhecido na literatura como *early stopping* (DEMUTH; BEALE, 1998).

Para as redes MLP classe-modular, cada um dos L classificadores binários foi treinado independentemente usando o algoritmo de retropropagação do erro (HAYKIN, 1996), juntamente com os conjuntos de treinamento e validação. Para treinar um classificador binário para cada classe de palavra é feita uma reorganização dos conjuntos de treinamento e validação em dois subconjuntos, Z_0 e Z_1. Em que Z_0 contêm as amostras da classe corrente e Z_1 as amostras de todas as outras classes, não sendo feita nenhuma espécie de balanceamento. Para reconhecer os padrões de teste, o módulo de decisão considera somente as saídas O_0 de cada subrede

(Figura 4.13) e utiliza um esquema *winner-takes-all* para determinar a classe final.

Com relação ao sistema baseado no processo de pseudo-segmentação variável (PV), os modelos escondidos de Markov usados neste trabalho são baseados em uma topologia discreta do tipo esquerda-direita, em que cada transição pode pular até dois estados. O tamanho do léxico permite um modelo para cada classe. As matrizes A e B foram inicializadas considerando distribuições uniformes. Com base no tamanho médio das sequências de observações estabeleceu-se uma estratégia de busca do número de estados adequado para cada modelo variando de 3 até o tamanho mínimo das sequências de observações durante o processo de treinamento. O treinamento dos modelos baseia-se no algoritmo de Baum-Welch, discutido na Seção 4.4.2, juntamente com um processo de validação cruzada (*cross-validation*), como descrito por Rabiner (1989). O objetivo do processo de validação cruzada é monitorar o resultado final durante o processo de treinamento. Ele é feito sobre dois conjuntos de dados: treinamento e validação. Depois de cada iteração do algoritmo de Baum-Welch sobre os dados de treinamento, a verossimilhança dos dados de validação é calculada usando o algoritmo *Forward*, apresentado na Seção 4.4.2. O processo de treinamento calcula

$$\lambda = argmax(P(\lambda|O_i, i = 1, \cdots, t)) \tag{5.1}$$

em que λ é o modelo reestimado, O_i são sequências de observação na base de dados de treinamento, $P(\lambda|O_t)$ é a probabilidade da sequência de observação O_t dado o modelo λ e O_t denota o símbolo observado no tempo t.

Durante os experimentos, os escores de casamento (*matching*) entre cada modelo λ_i e uma sequência de observações O são calculados usando o algoritmo *Forward*, de modo a determinar

$$P(O|\lambda_l) = argmax(P(O|\lambda_i)) \tag{5.2}$$

em que λ_l é o modelo que determina a máxima probabilidade para a amostra, atribuindo-a à classe l.

Os modelos de Markov foram avaliados considerando as mesmas distribuições de palavras usadas pelo classificador neural. Para cada classe, um modelo foi treinado e validado, de forma similar ao que é feito nos classificadores neurais.

5.3.2 Resultados obtidos

O primeiro parâmetro de desempenho utilizado para avaliar o sistema foi a taxa de reconhecimento, que representa o percentual de palavras classificadas corretamente. Na Tabela 5.2 são apresentadas as taxas de reconhecimento obtidas para cada esquema individualmente. Os

esquemas que utilizam classificadores neurais foram avaliados considerando as arquiteturas convencional e classe-modular. Pode-se observar que na média, a maior taxa de reconhecimento foi obtida pelo processo de pseudo-segmentação fixa usando o conjunto de características direcionais (PF-D) na arquitetura classe-modular.

Tabela 5.2: Taxa de reconhecimento média obtida por cada classificador individualmente para cada classe, sendo PR - Pseudo-segmentação de Radical, PF-P - Pseudo-segmentação Fixa com características Perceptivas, PF-D - Pseudo-segmentação Fixa com características Direcionais, PF-T - Pseudo-segmentação Fixa com características Topológicas e PF-P - Pseudo-segmentação Variável.

Classes	Classificadores								
	PR Conv	PR Modular	PF-P Conv	PF-P Modular	PF-D Conv	PF-D Modular	PF-T Conv	PF-T Modular	PV
Janeiro	60,0 %	60,6 %	76,5 %	80,0 %	92,3 %	**92,9 %**	85,3 %	83,5 %	75,3 %
Fevereiro	74,1 %	70,0 %	82,9 %	83,5 %	90,6 %	**92,3 %**	84,7 %	83,5 %	81,2 %
Março	69,4 %	72,9 %	87,1 %	87,6 %	94,7 %	**95,3 %**	90,0 %	92,3 %	83,5 %
Abril	80,6 %	84,1 %	90,0 %	91,7 %	94,1 %	**94,7 %**	87,6 %	89,4 %	85,9 %
Maio	85,9 %	86,5 %	94,1 %	92,3 %	96,5 %	**97,6 %**	86,5 %	87,6 %	85,3 %
Junho	65,3 %	70,6 %	82,3 %	**85,9 %**	81,8 %	81,8 %	82,9 %	81,2 %	76,5 %
Julho	78,2 %	78,8 %	84,1 %	85,9 %	83,5 %	85,9 %	84,7 %	**87,1 %**	78,2 %
Agosto	75,9 %	78,2 %	90,6 %	**92,3 %**	90,6 %	90,6 %	75,2 %	80,6 %	88,2 %
Setembro	68,8 %	68,8 %	82,3 %	80,6 %	**90,0 %**	87,1 %	85,9 %	84,7 %	74,1 %
Outubro	78,2 %	76,5 %	87,1 %	90,0 %	94,1 %	**95,3 %**	87,6 %	87,6 %	87,1 %
Novembro	81,2 %	77,6 %	85,9 %	86,5 %	91,2 %	**92,3 %**	87,6 %	87,1 %	83,5 %
Dezembro	62,3 %	61,8 %	78,8 %	79,4 %	89,4 %	**91,2 %**	72,9 %	75,2 %	81,2 %
Média	73,3 %	73,9%	85,1 %	86,3 %	90,7 %	**91,4 %**	84,3 %	85,0 %	81,7 %

Analisando o desempenho individual por classe, observa-se que o classificador PF-D Modular obteve o melhor desempenho na maioria das classes, porém naquelas em que os classificadores PF-P Modular e PF-T Modular se sobressaíram, a diferença de desempenho entre eles é considerável, o que indica que eles podem ser complementares. Outra observação é que os classificadores PR e PV não tiveram destaque em nenhuma das classes, tendo na maioria das vezes desempenho inferior aos outros classificadores. Esta observação vista de forma isolada leva a crer que os sistemas PR e PV não são interessantes, porém mais adiante será mostrado que isso não é necessariamente uma verdade.

Comparando-se o desempenho das arquiteturas de rede neural, observa-se que em todos os esquemas, a arquitetura classe-modular obteve uma taxa de reconhecimento em média maior do

que àquela obtida pela arquitetura convencional, embora tenha sido uma diferença mínima em alguns casos. Este resultado mostra que dependendo da aplicação o uso da arquitetura classe-modular pode trazer um aumento do custo computacional em relação ao tempo de treinamento, sem implicar necessariamente num melhor desempenho do classificador. Mesmo assim, os resultados apresentados deste ponto em diante, envolvendo classificadores neurais, se limitarão aos obtidos usando arquitetura classe-modular.

Outra ferramenta de avaliação do desempenho dos classificadores consiste na análise da distribuição dos dados na matriz de confusão que fornece informações valiosas sobre o comportamento do sistema. As colunas dessa matriz representam os dados a serem classificados, e as linhas representam o número de classificações para o conjunto de dados analisado. Nas Tabelas 5.3, 5.4, 5.5, 5.6 e 5.7 são apresentadas as matrizes de confusão obtidas para cada classificador. No caso dos classificadores neurais são mostradas as matrizes correspondentes ao treinamento na arquitetura classe-modular.

A análise das matrizes mostra que os principais pontos de confusão são àqueles já previstos na Figura 4.1, que ocorrem entre classes semelhantes e/ou com a mesma terminação. Entretanto, outros pontos de confusão merecem destaque, são eles:

- **Sistema PR** - Janeiro e Março, Fevereiro e Setembro, Abril e Maio, Agosto e Abril, Setembro e Fevereiro, Dezembro e Fevereiro.
- **Sistema PF-P** - Janeiro e Julho.
- **Sistema PF-D** - Junho e Janeiro, Julho e Janeiro.
- **Sistema PF-T** - Maio e Abril, Junho e Janeiro, Dezembro e Agosto.
- **Sistema PV** - Julho e Maio, Agosto e Abril.

É difícil definir com exatidão as causas dessas confusões, mesmo porque em alguns classificadores ocorrem erros generalizados. Elas podem ser geradas por um mapeamento inadequado características-classificador ou por deficiência do próprio método de reconhecimento, embora provavelmente sejam fruto de uma segmentação inadequada e/ou de um conjunto de características pouco discriminante nesses casos isolados.

Tabela 5.3: Matriz de confusão para o sistema de pseudo-segmentação de radical (PR).

Mês	J	F	M	A	M	J	J	A	S	O	N	D
Janeiro	**103**	*28*	*11*	3	5	4	1	2	2	6		5
Fevereiro	*19*	**119**	1		1	1	1	2	*10*	4	5	7
Março	8		**124**	4	*27*	1	1	2			3	
Abril	1	1	5	**143**	*11*	3	2	4				
Maio	1		*12*	9	**147**	1						
Junho	4	1	1	2	3	**120**	*30*	3		3	1	2
Julho	3	2	3	4	4	*16*	**134**	2	1	1		
Agosto	8		4	9	3	1	2	**133**	1	2		7
Setembro	1	*12*	2		1	5		1	**117**	*12*	8	*11*
Outubro	4	2	4	2	1	3	3	3	9	**130**	6	3
Novembro	1	4	1						*16*	6	**132**	*10*
Dezembro	4	*14*	1			1	1	6	*16*	6	*16*	**105**

Tabela 5.4: Matriz de confusão para o sistema de pseudo-segmentação fixa com características perceptivas (PF-P).

Mês	J	F	M	A	M	J	J	A	S	O	N	D
Janeiro	**136**	*13*		2	2	3	7	2	1		3	1
Fevereiro	*11*	**142**		2	4	2	3		2		3	1
Março	1		**149**	2	*15*		1				1	1
Abril	3	2		**156**	6			2		1		
Maio		6	2		**157**		3	1	1			
Junho	3	5			1	**146**	9	1	3	1	1	
Julho	4	2	1		5	9	**146**		1	2		
Agosto	2	3	1	1	2			**157**			2	2
Setembro		6		2	1	2	2		**137**	4	*13*	3
Outubro		1		1	1	1	2		7	**153**	4	
Novembro	2	2	2	2	1	1	1	2	7	3	**147**	
Dezembro		5	1			2		5	7	3	*12*	**135**

Tabela 5.5: Matriz de confusão para o sistema de pseudo-segmentação fixa com características direcionais (PF-D).

Mês	J	F	M	A	M	J	J	A	S	O	N	D
Janeiro	**158**	5				3	2		1	1		
Fevereiro	9	**157**	1					1	1	1		
Março			**162**		6			1	1			
Abril	1			**161**				1	3	1		3
Maio		1	2		**166**					1		
Junho	9	1				**139**	*16*	1	1		2	1
Julho	9	1		1	1	7	**146**	3				2
Agosto	1	2	2	1				**154**		1	1	8
Setembro	1	1		1			1	1	**148**	2	8	7
Outubro									4	**162**	2	2
Novembro			4			3			4	1	**157**	1
Dezembro		1		1		1	1	2	6	1	2	**155**

Tabela 5.6: Matriz de confusão para o sistema de pseudo-segmentação fixa com características topológicas (PF-T).

Mês	J	F	M	A	M	J	J	A	S	O	N	D
Janeiro	**142**	6	1	1	4	6	2	2	2		2	2
Fevereiro	*11*	**142**	1	3	4	1			2	3	1	2
Março		1	**157**	1	5			2	4			
Abril		1		**152**	6			1	3	5		2
Maio		4		*14*	**149**	1				2		
Junho	*10*	2	1		1	**138**	*10*		1	2	1	4
Julho	1			2	2	9	**148**	3	1	4		
Agosto	4		3	1	7	2		**137**	4	1	4	7
Setembro		2		1	1	1	2	2	**144**	6	7	4
Outubro		5				1	2	1	3	**149**	7	1
Novembro	1	2		1	1			1	5	2	**148**	9
Dezembro		1	1			2		*12*	8	2	*16*	**128**

Tabela 5.7: Matriz de confusão para o sistema de pseudo-segmentação variável (PV).

Mês	J	F	M	A	M	J	J	A	S	O	N	D
Janeiro	**128**	*11*		2	4	4	6	1	7		3	4
Fevereiro	7	**138**	5		2	4	1	1	4	3	3	2
Março		3	**142**	4	*18*		2			1		
Abril	1		2	**146**	3		4	7	1	4		2
Maio	1	1	*14*	5	**145**		4					
Junho		5		4	4	**130**	*18*		4	3		2
Julho	3	2	2	6	*10*	8	**133**	3	1	2		
Agosto	3			*10*	4		1	**150**		1		1
Setembro	1		3	4	6	2	3	1	**126**	7	7	*10*
Outubro		1		4	1		6	1	8	**148**	1	
Novembro	4		2	2		1	1	1	*12*	4	**142**	1
Dezembro	1	2		2		1		4	8	6	8	**138**

5.4 Análise da combinação dos classificadores

5.4.1 Metodologia de testes

Para avaliar o potencial combinado dos classificadores, as saídas individuais das redes e a probabilidade estimada para cada modelo de Markov foram combinadas. Para garantir que esta combinação seja válida do ponto de vista estatístico é feita uma normalização das saídas das redes neurais garantindo assim que elas representam estimativas das probabilidades *a posteriori*.

Como discutido na Seção 4.4.2, a utilização de escalonamento define que o algoritmo de treinamento dos modelos escondidos de Markov estime o logaritmo das probabilidades $P(O|\lambda_i)$ (Equação 4.38), de modo que os valores obtidos são pequenos ($\approx 10^{-8}$), o que pode afetar o resultado final da combinação desde que a ordem dos valores de saída das redes neurais são $\approx 10^{-2}$. Para evitar esse problema, foi determinado um valor normalizado de probabilidade para cada modelo λ_i:

$$P^*(O|\lambda_i) = \frac{P(O|\lambda_i)}{\sum_j P(O|\lambda_j)} \tag{5.3}$$

Deste modo, resolvido o problema da normalização das probabilidades $P(O|\lambda_i)$, resta apenas definir as regras de combinação a serem utilizadas. Para tanto, foi feita uma seleção das principais regras de combinação discutidas na literatura (WEBB, 2002). Levando-se em conta que os classificadores definidos nesta aplicação podem ser considerados estáveis pelo tamanho do conjunto de treinamento utilizado, se fez a opção de utilizar regras fixas de combinação.

De modo que, sendo Z um objeto que se deseja classificar, L o número de classes envolvidas no problema e tendo C classificadores com entradas $x_1, \cdots, x_C$, as regras de combinação são definidas como:

- Regra do produto - Atribua Z à classe w_j se

$$\prod_{i=1}^{C} p(w_j|x_i) > \prod_{i=1}^{C} p(w_k|x_i); \quad k = 1, \cdots, L, \quad k \neq j. \tag{5.4}$$

- Regra da soma - Atribua Z à classe w_j se

$$\sum_{i=1}^{C} p(w_j|x_i) > \sum_{i=1}^{C} p(w_k|x_i); \quad k = 1, \cdots, L, \quad k \neq j. \tag{5.5}$$

- Regra da soma ponderada - Atribua Z à classe w_j se

$$\sum_{i=1}^{C} \alpha_i \cdot p(w_j|x_i) > \sum_{i=1}^{C} \alpha_i \cdot p(w_k|x_i); \quad k = 1, \cdots, L, \quad k \neq j, \tag{5.6}$$

  em que α_i, $i = 1, \cdots, C$ são pesos específicos para cada um dos C classificadores, que satisfazem à condição de $\sum_{i=1}^{C} \alpha_i = 1$.

O desenvolvimento completo dessas regras é encontrado em literatura específica, como em Webb (2002), Matos (2004) e Kittler et al. (1998).

5.4.2 Resultados obtidos

Definidas as regras, apresentam-se na Tabela 5.8 as taxas de reconhecimento médias obtidas considerando cada uma dessas regras de combinação em diferentes configurações de classificadores. Os pesos usados na regra da soma ponderada foram obtidos através de um procedimento de busca que consistiu na geração aleatória dos pesos num total de 2000 iterações para cada combinação, sendo então determinado a *n-upla* que obteve o melhor resultado.

A análise da Tabela 5.8 mostra que a melhor taxa de reconhecimento foi obtida utilizando a regra da soma ponderada na combinação dos classificadores PR, PF-P, PF-D, PF-T e PV. Comparando este resultado àquele obtido pelo melhor classificador individual (PF-D) obteve-se um ganho de $\frac{97,8\%-91,4\%}{91,4\%} = 7,0\%$ na taxa de reconhecimento. Neste caso, em especial, foram determinados os seguintes pesos na combinação: $\alpha_1 = 0,261$ para PR, $\alpha_2 = 0,138$ para PF-P, $\alpha_3 = 0,244$ para PF-D, $\alpha_4 = 0,089$ para PF-T e $\alpha_5 = 0,268$ para PV.

O resultado anterior mostra que apesar do classificador PF-D obter o melhor desempenho individual, na ponderação, o classificador ao qual foi associado o maior peso foi PV, levando à

Tabela 5.8: Taxa de reconhecimento média obtida usando diferentes combinações de classificadores.

Classificadores	Regras de fusão		
	Produto (%)	Soma (%)	Soma ponderada (%)
PF-P e PF-D	93,0	92,9	93,6
PF-P e PF-T	90,5	89,4	90,0
PF-D e PF-T	92,1	91,7	93,1
PF-P e PR	90,8	90,3	90,5
PF-D e PR	94,0	93,9	94,4
PF-T e PR	90,0	88,9	89,3
PF-P e PV	93,9	93,2	93,5
PF-D e PV	95,7	95,0	95,6
PF-T e PV	93,6	93,4	93,5
PR e PV	91,0	89,9	90,5
PF-P, PF-D e PF-T	93,5	93,4	94,0
PF-P, PF-D e PR	95,5	94,8	95,4
PF-P, PF-D e PV	96,3	96,8	96,9
PF-P, PF-T e PR	92,9	92,5	93,2
PF-P, PF-T e PV	94,9	95,3	95,7
PF-D, PF-T e PR	93,8	94,0	95,0
PF-D, PF-T e PV	95,7	96,0	96,5
PF-P, PR e PV	95,4	95,4	95,8
PF-D, PR e PV	96,6	96,8	97,2
PF-T, PR e PV	95,1	95,1	95,5
PF-P, PF-D, PF-T e PR	95,0	95,1	95,8
PF-P, PF-D, PF-T e PV	96,2	96,3	97,2
PF-P, PF-D, PR e PV	97,2	97,1	97,7
PF-P, PF-T, PR e PV	96,0	96,2	96,9
PF-D, PF-T, PR e PV	96,5	96,8	97,4
PF-P, PF-D, PF-T, PR e PV	97,0	97,3	**97,9**

conclusão que os classificadores utilizados são complementares, o que justifica a utilização da combinação de classificadores no problema em questão. Embora este tenha sido o melhor resultado, é interessante observar que a combinação dos classificadores PR, PF-P, PF-D e PV obteve desempenho muito próximo, ou seja, o classificador PF-T não tem muita influência na combinação, o que já era indicado pelo baixo valor do peso associado à ele na melhor combinação. Já o classificador PR têm um resultado diferente, pois mesmo não tendo destaque em nenhuma classe isoladamente e com um desempenho individual bem inferior aos demais, na combinação tem um papel mais discriminante que o classificador PF-T. Outro ponto interessante é a combinação dos classificadores PF-D, PR e PV que obteve uma taxa de reconhecimento elevada, próxima do melhor resultado da combinação. Estas observações mostram a complementariedade dos três sistemas de reconhecimento baseados em processos diferentes de pseudo-segmentação, o que valida o processo de análise multi-vistas proposto neste trabalho.

Embora a melhor taxa de reconhecimento tenha sido obtida na combinação de todos os classificadores usando a regra da soma ponderada, percebe-se que outras configurações tiveram desempenho similar. De modo que para fazer uma melhor avaliação foi calculado o coeficiente *Kappa* nos casos em que a taxa de reconhecimento era próxima dos 97,9%.

O coeficiente *Kappa* foi proposto no ínicio dos anos 60 por Cohen (1960) como medida alternativa de desempenho aplicada à testes psicológicos para diagnose médica. Desde então, diversos pesquisadores, citando por exemplo os trabalhos de Carletta (1996), Mangabeira, Azevedo e Lamparelli (2003), Breve, Jr. e Mascarenhas (2005) tem utilizado este coeficiente como medida auxiliar para avaliação de desempenho de classificação. Ele é calculado, a partir da matriz de confusão, segundo a Equação 5.7 e sua variância também pode ser determinada de acordo com a Equação 5.8:

$$K = \frac{N \cdot \sum_{i=1}^{r} x_{ii} - \sum_{i=1}^{r}(x_{i+} \cdot x_{+i})}{N^2 - \sum_{i=1}^{r}(x_{i+} \cdot x_{+i})}; \tag{5.7}$$

$$\sigma_K^2 = \frac{\frac{\sum_{i=1}^{r} x_{ii}}{N}(1 - \frac{\sum_{i=1}^{r} x_{ii}}{N})}{N(1 - \frac{\sum_{i=1}^{r}(x_{i+} \cdot x_{+i})}{N^2})^2}; \tag{5.8}$$

em que N é o número total de amostras, r é o número total de classes do problema em questão, x_{ii} é o número de amostras classificadas corretamente em cada classe, x_{i+} e x_{+i} é o número de amostras da linha i e da coluna i, respectivamente. O coeficiente *Kappa* varia no intervalo de $[-1, 1]$, em que -1 significa discordância total e 1 significa concordância total. De modo que, quanto mais próximo de 1 estiver o valor de K, melhor é o desempenho do classificador. Na Tabela 5.9 é apresentado um comparativo entre a taxa de reconhecimento, o coeficiente *Kappa* e sua variância calculados para algumas das configurações de classificadores mostradas anteriormente.

Tabela 5.9: Comparação entre a taxa de reconhecimento obtida, o coeficiente *Kappa* e sua variância calculados para algumas combinações de classificadores.

Classificadores	Taxa de Reconhecimento(%)	Coeficiente *Kappa*	Variância
PF-P, PF-T, PR e PV	96,9	0,9657	1,7796e-05
PF-P, PF-D e PV	96,9	0,9663	1,7460e-05
PF-D, PR e PV	97,2	0,9699	1,5635e-05
PF-P, PF-D, PF-T e PV	97,2	0,9700	1,5575e-05
PF-D, PF-T, PR e PV	97,4	0,9716	1,4805e-05
PF-P, PF-D, PR e PV	97,7	0,9743	1,3442e-05
PF-P, PF-D, PF-T, PR e PV	97,9	0,9759	1,2634e-05

Fazendo uma análise da tabela, pode-se ver que o maior coeficiente *Kappa* corresponde à maior taxa de reconhecimento obtida pela configuração que utiliza todos os classificadores. Um detalhe interessante de ser observado é a de que configurações com mesma taxa de reconhecimento, possuem coeficientes *Kappa* ligeiramente diferentes. Pode-se ver, por exemplo, que apesar das configurações *PF-P, PF-T, PR e PV* e *PF-P, PF-D e PV* terem mesma taxa de reconhecimento, a segunda possui um coeficiente Kappa mais alto e uma variância menor, usando um número menor de classificadores na combinação. Ou seja, a segunda tem um desempenho melhor com um menor custo computacional. Desse modo, o coeficiente *Kappa* é usado como medida auxiliar na avaliação da combinação.

Como parte do sistema proposto é constituída por quatro classificadores neurais treinados com diferentes conjuntos de características, é interessante investigar o desempenho do sistema ao se utilizar uma única rede. Deste modo, foi construída uma rede única na arquitetura classe-modular, usando a mesma metodologia descrita na Seção 5.3.1, com 264 neurônios na camada de entrada e uma camada escondida variando de 250 a 300 neurônios. A melhor configuração foi obtida usando 264 neurônios na camada escondida, com uma taxa de reconhecimento média de 93,4%.

Comparando esse valor com a combinação dos classificadores *PF-P, PF-D, PF-T e PR* apresentada na Tabela 5.8, que obteve uma taxa de reconhecimento média de 95,8%, pode-se ver que a rede única tem desempenho inferior, além de apresentar maior custo computacional de treinamento devido ao seu tamanho. Esse fato reforça a vantagem de se dividir a tarefa de reconhecimento entre vários classificadores mais simples ao invés de se utilizar um classificador único.

Para finalizar, na Tabela 5.10 é apresentada a matriz de confusão obtida para a melhor combinação. Comparando-a com a matriz de confusão do melhor classificador individual mostrada na Tabela 5.4, observa-se um aumento na taxa de reconhecimento para todas as classes, principalmente para as classes $Junho$ e $Julho$ que apresentam um aumento de aproximadamente 15%. Este resultado mostra que a combinação dos diferentes classificadores definidos pela arquitetura de análise multi-vistas melhora o desempenho geral do sistema quando comparados aos classificadores isolados.

Tabela 5.10: Matriz de confusão para o melhor resultado de combinação.

Mês	J	F	M	A	M	J	J	A	S	O	N	D
Janeiro	**165**	1				1	1		2			
Fevereiro	3	**165**						1			1	
Março			**166**		4							
Abril				**168**	2							
Maio			1		**169**							
Junho	1					**160**	*9*					
Julho						1	**169**					
Agosto				2				**168**				
Setembro				1			1		**163**		3	2
Outubro									1	**169**		
Novembro									1		**169**	
Dezembro						1		2	3			**164**

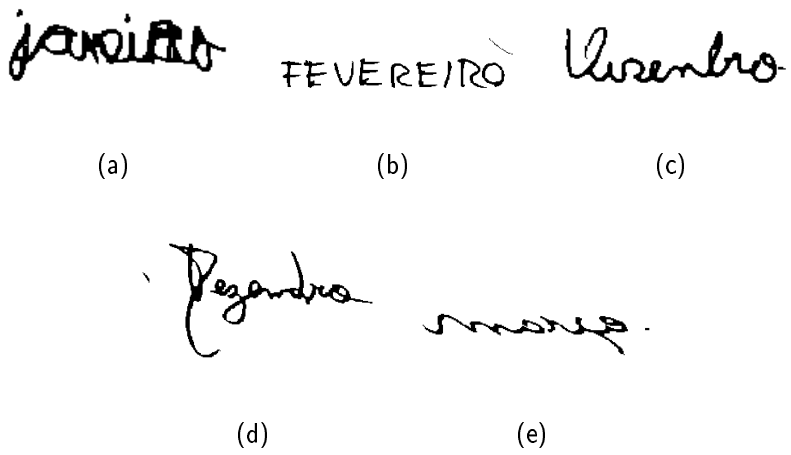
Buscando uma melhor compreensão dos erros obtidos, foi feita uma análise visual subjetiva das imagens correspondentes aos erros apontados na Tabela 5.10. Alguns exemplos de palavras reconhecidas de forma correta e incorreta são mostrados na Figura 5.5.

Algumas conclusões importantes foram retiradas desta análise:

1. 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;

2. Considerando os estilos de escrita, o estilo caixa alta apresentou maior número de erros de classificação, possivelmente pela falta de ascendentes/descendentes.

Estes problemas são decorrentes das limitações inerentes ao método global de reconhecimento de palavras.



(a) (b) (c)

(d) (e)

Figura 5.5: Exemplos do resultado da classificação: (a) palavra reconhecida como **julho**, (b) palavra reconhecida como **novembro**, (c) palavra reconhecida como **junho**, (d) e (e) palavras reconhecidas corretamente.

5.5 Análise da diversidade dos classificadores

A solução mais simples para determinar a melhor configuração de um sistema com múltiplos classificadores é avaliar todas as possíveis combinações, como realizado na seção anterior, mas isto têm um custo computacional associado que varia com o número de classificadores considerados.

É interessante desenvolver um método alternativo que não use os escores de saída dos classificadores para avaliar a configuração. A idéia apresentada neste trabalho é usar a informação contida nas matrizes de confusão de cada classificador individual e calcular distâncias que representem a discordância dos classificadores.

Deste modo, as distâncias provêm uma medida quantitativa da diversidade dos classificadores, cujo conceito e a importância foi introduzido no Capítulo 4. O cálculo das distâncias juntamente com uma hipótese de correlação suave, descrita posteriormente, definem um mecanismo de avaliação *a priori* da melhor combinação de classificadores.

5.5.1 Definição do método

Duin, Pekalska e Tax (2004) aplicou o conceito de discordância para medir a diferença entre dois classificadores C_1 e C_2 treinados sobre um problema de classificação $P_j(j = 1, \cdots, N)$ em que N é o número de classes do problema. Deste modo, a discordância $d_j(C_1, C_2)$ pode ser formulada pela Equação 5.9:

$$d_j(C_1, C_2) = P(C_1(x) \neq C_2(x) | x \in P_j) \tag{5.9}$$

em que $C_i(x)$ retorna o rótulo dado ao objeto x pelo classificador C_i. Deste modo M classificadores constituem uma matriz de discordância D_j^C de tamanho $M \times M$ para o problema P_j, cujos elementos $D_j^C(m,n) = d_j(C_m, C_n)$. Este conceito será utilizado na definição do critério de distância, porém sendo aplicado sobre matrizes de confusão.

Como apresentado anteriormente, sabe-se que a matriz de confusão é uma representação quantitativa do desempenho obtido por cada classificador em termos do reconhecimento por classe. A matriz de confusão pode ser denotada matematicamente (ZOUARI, 2004), como mostra a Equação 5.10:

$$A = \begin{bmatrix} TR_{1,1} & TR_{1,2} & \cdots & TR_{1,N} \\ \vdots & \vdots & & \vdots \\ TR_{i,1} & TR_{i,2} & \cdots & TR_{i,N} \\ \vdots & \vdots & & \vdots \\ TR_{N,1} & TR_{N,2} & \cdots & TR_{N,N} \end{bmatrix} \tag{5.10}$$

em que $TR_{i,j}$ corresponde ao número total de amostras da classe C_i cuja solução correta é colocada na posição j.

A partir disso, as distâncias são obtidas de acordo com a definição dada pela Equação 5.11:

$$D^{A,B} = \sum_{i=1}^{N} \sum_{j=1}^{N} |TR_{i,j}^A - TR_{i,j}^B| \tag{5.11}$$

em que A e B são matrizes de confusão de mesmo tamanho.

Uma vez determinada as distâncias entre os classificadores é necessário estabelecer uma regra que defina a melhor combinação. Para isso, segue-se a idéia proposta por Hadjitodorov, Kuncheva e Todorova (2005): *Comitês de múltiplos classificadores selecionados considerando a diversidade mediana têm desempenho superior àqueles selecionados aleatoriamente ou considerando a diversidade máxima*. Esta hipótese será denominada de **regra da correlação suave**. Ou seja, uma vez determinada a distância entre as matrizes de confusão, a melhor configuração está localizada em torno da distância mediana.

5.5.2 Análise dos resultados obtidos

Na Tabela 5.11 são comparados os resultados obtidos na combinação de pares de classificadores usando o critério de distância apresentado e a regra da soma, respectivamente. As linhas em negrito na tabela indicam as melhores combinações de acordo com a regra da correlação suave.

A análise da tabela evidencia dois pontos importantes:

1. O grau de discordância determinado para classificadores com estruturas de pseudo-segmentação similares é muito próximo. Ou seja, a arquitetura de análise multi-vistas que

define estruturas de pseudo-segmentação diferentes é um fator gerador de diversidade.

2. A regra da correlação suave foi confirmada, desde que a combinação de melhor desempenho foi aquela com grau mediano de diversidade, ao contrário do que se poderia imaginar a princípio como sendo a de diversidade máxima.

Deste modo, tem-se uma ferramenta simples que utiliza uma representação clássica (matrizes de confusão) para prover mais informação sobre os classificadores. Entretanto, este resultado não prova que a regra da correlação suave é válida em todos os casos, sendo necessário um estudo aprofundado e uma comparação com outras medidas de avaliação da diversidade.

Tabela 5.11: Comparação do critério da distância em relação à regra da soma na combinação de classificadores.

Classificadores	Distância	Regra da soma
PF-D e PF-T	1.30	91.7 %
PF-P e PF-T	1.33	89.4 %
PF-P e PF-D	1.34	92.9 %
PF-P e PV	1.44	93.2 %
PF-D e PV	**1.77**	**95.0 %**
PF-T e PV	**1.88**	**93.4 %**
PR e PV	2.07	89.9 %
PR e PF-P	2.28	90.3 %
PR e PF-T	2.28	88.9 %
PR e PF-D	2.57	93.9 %

5.6 Resultados descritos na literatura

Comparar o desempenho do sistema descrito com outros sistemas de reconhecimento de palavras manuscritas discutidos na literatura não é simples, devido ao uso de bases de dados e/ou léxicos distintos. Além do estudo de Kapp (2004), que foi utilizado no desenvolvimento deste trabalho, o único estudo para Língua Portuguesa que se têm conhecimento é o de Morita et al. (2004) que utilizou o mesmo léxico com uma base de dados diferente. São utilizados dois conjuntos de características, um baseado na análise de concavidades e outro utilizando características perceptivas, tendo obtido uma taxa de reconhecimento média de 91,5% usando um método analítico com estágio de verificação.

Outro estudo similar foi apresentado por Kim et al. (2000) que combinou classificadores MLP e HMM para um dicionário de 12 classes, correspondente aos meses do ano para a Língua Inglesa

(que apresenta dificuldades de reconhecimento similares ao que ocorre na Língua Portuguesa), extraído da base de dados do CENPARMI (*Centre for Pattern Recognition and Machine Intelligence*) da *University of Concordia* no Canadá. 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. Considerando o classificador MLP isolado foi obtida uma taxa de reconhecimento média de 79,4%. Para a combinação MEM-MLP esta taxa atinge 88,8%.

Esta comparação mostra que o sistema descrito neste trabalho obtém taxas melhores que outros sistemas similares descritos recentemente na literatura.

5.7 Conclusão

Neste capítulo foram apresentados os testes efetuados e os resultados obtidos neste trabalho. A análise mostrou que o classificador PF-D obteve melhor desempenho individual. Porém, ele nem sempre é suficiente e sua combinação com os outros classificadores apresentados resolve grande parte das suas confusões isoladamente. Isto só é possível pois os classificadores apresentados são complementares e a combinação avaliada ressalta essa característica, melhorando o desempenho geral do sistema.

Também foi introduzida uma medida de diversidade para avaliar os classificadores, mostrando que estratégias de pseudo-segmentação diferentes tem um maior grau de discordância. Outra conclusão obtida através dos experimentos é que diversidade máxima não é sinônimo de melhor desempenho, sendo necessário estabelecer uma solução de compromisso entre estes dois paradigmas.

Deste modo, a arquitetura de análise multi-vistas foi validada, embora alguns classificadores tenham um desempenho inferior, sua participação na combinação melhora o resultado geral do sistema. Isso mostra que as diferentes aproximações aumentam o potencial discriminatório do sistema.

Capítulo 6

Considerações Finais e contribuições

O reconhecimento de palavras manuscritas é um problema de difícil solução devido à grande variedade de formas apresentada pela escrita manual. Uma forma de superar essa dificuldade no desenvolvimento de sistemas práticos é limitar o dicionário de interesse, ou seja, o conjunto de palavras a ser discriminado pelo sistema. Neste caso, a tarefa principal é aumentar a discriminação entre as palavras, que são definidas usando representações que procuram abstrair de suas imagens informações únicas à cada classe. Deste modo, a linha de investigação principal deste trabalho foi definir estas representações a partir de modelos perceptivos do processo de leitura humano.

Neste intuito, foi apresentada uma nova arquitetura de reconhecimento, denominada de **análise multi-vistas**, que consiste na utilização de diferentes processos de pseudo-segmentação. O objetivo principal desta análise era refinar a representação de imagens confusas usando diferentes aproximações que ajudassem no processo de tomada de decisão posterior de modo a obter um sistema robusto. Esta arquitetura definiu um sistema de múltiplos classificadores que foi ajustado com base na hipótese que o melhor classificador individual têm desempenho inferior àquele obtido pela combinação de classificadores complementares. De modo que foram construídos sistemas de reconhecimento complementares, sendo otimizados individualmente de modo a obter na combinação um desempenho melhor do sistema.

Esta arquitetura utilizou classificadores neurais e modelos escondidos de Markov no desenvolvimento dos múltiplos sistemas. A principal conclusão obtida é que os sistemas definidos são complementares e a estratégia de combinação proposta ressalta essa complementariedade. Deste modo, o sistema de múltiplos classificadores obteve uma solução melhor para o problema em análise do que qualquer dos classificadores tomados individualmente. Este resultado indica que estratégias similares podem ser aplicadas para outros dicionários restritos.

Sendo assim, este trabalho definiu uma arquitetura original baseada numa metodologia de análise multi-vistas de imagens de palavras manuscritas para propor um sistema de reconhecimento novo e eficiente, que foi aplicado no reconhecimento das palavras manuscritas correspondentes aos nomes dos meses do ano. Este sistema extrai características globais da imagem da palavra, evitando assim a necessidade de um procedimento de segmentação explícita. Outra característica original, é que este método explora informações do contexto da palavra e incorpora aspectos extraídos de modelos perceptivos. Deste modo, o sistema foi construído de modo a obter uma aproximação computacional dos mecanismos perceptivos usados no processo de leitura humano, como sugerido por teorias cognitivas.

Podem ser citadas as seguintes contribuições originais deste trabalho:

Concepção e desenvolvimento de uma nova arquitetura de reconhecimento global de palavras manuscritas aplicada a dicionários restritos.

Desenvolvimento de uma modelagem computacional do processo de leitura humano que pode ser aplicada em diferentes problemas de dicionário restrito na área de reconhecimento de palavras manuscritas.

Aplicabilidade prática do sistema desenvolvido na leitura das palavras dos meses do ano, comprovada pela análise dos resultados obtidos, já que o mesmo apresentou desempenho superior a outros sistemas relatados na literatura.

Construção de uma base de dados *omni-escritor* composta de 10.200 imagens provenientes de 850 escritores de diferentes níveis sociais e educacionais, sendo assim bastante heterogênea.

Avaliação e discussão do potencial de diferentes conjuntos de características e classificadores considerando um problema em comum na área de reconhecimento de palavras manuscritas.

Desenvolvimento de estudo sobre o conceito de diversidade e sua relação com sistemas de múltiplos classificadores, sendo proposta uma medida alternativa de avaliação da diversidade.

Na sequência, lista-se outras linhas de trabalho que podem ser investigadas a fim de contribuir para melhoria do trabalho proposto.

Aprofundamento do estudo da diversidade em múltiplos classificadores - Neste trabalho foi apresentada uma regra de correlação suave e uma medida de avaliação da discordância dos classificadores. Porém, é necessário que se faça um aprofundamento do estudo, fazendo uma comparação com outras medidas existentes de modo a garantir a eficiência do método apresentado.

Implementação de uma busca ótima dos pesos - Os resultados indicaram que a regra de combinação por soma ponderada determinou o melhor resultado. Entretanto, o procedimento de busca exaustiva utilizado não pode ser considerado ótimo. De modo que a aplicação de métodos de otimização como, por exemplo, algoritmos genéticos podem levar a taxas de reconhecimento maiores.

Ampliação da base de dados e aplicação em outros léxicos - Para uma melhor validação do sistema é necessário que a base de dados seja ampliada de modo a identificar os pontos de confusão e extender sua aplicação para outros léxicos, como àquele definido pelo extenso manuscrito dos cheques bancários.

Bibliografia

AIRES, S. B. K. *Reconhecimento de Caracteres Manuscritos Baseado em Regiões Perceptivas*. Dissertação (Mestrado) — Pontifícia Universidade Católica do Paraná, 2005.

ALLI, A. Z. et. *SNNS - Stuttgart Neural Network Simulator, User Manual, Version 4.2.* , 1994.

ÂVILA, M. *Optimisation de Modèles Markoviens pour la Reconnaissance de L'Ecrit*. Tese (Doutorado) — Université de Rouen, 1996.

BEUCHER, S.; MEYER, F. *Mathematical Morfology in Image Processing*. Marcel Dekker Inc., 1992.

BREVE, F. A.; JR., M. P. P.; MASCARENHAS, N. D. A. Combining methods to stabilize and increase performance of neural network-based classifiers. In: *Simpósio Brasileiro de Computação Gráfica e Processamento de Imagens - SIBGRAPI'2005*. 2005. p. 105–111.

BROWN, G. et al. Diversity creation methods: A survey and categorisation. *Information Fusion*, v. 6, n. 1, p. 5–20, 2005.

CARLETTA, J. Assessing agreement on classification tasks: the kappa statistic. *Computational Linguistics*, v. 22, n. 2, p. 249–254, 1996.

COHEN, J. A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, v. 20, n. 1, p. 37–46, 1960.

CORREIA, S. E. N. *Reconhecimento de Caracteres Manuscritos Usando Wavelets*. Tese (Doutorado) — Universidade Federal de Campina Grande, 2005.

CÔTÉ, M. et al. Automatic reading of cursive scripts using a reading model and perceptual concepts. *International Journal on Document Analysis and Recognition*, v. 1, p. 3–17, 1998.

COVER, T. M.; THOMAS, J. A. *Elements of Information Theory*. Wiley Series in Telecommunications, 1991.

CÔTÉ, M. . *Utilisation d'un Modèle d'Accès et de Concepts Perceptifs pour la Reconnaissance d'Images de Mots Cursifs*. Tese (Doutorado) — École Nationale Supérieure des Télécommunications, 1997.

DEMUTH, H.; BEALE, M. *Neural Network Toolbox - For Use with MATLAB*. The MathWorks, Inc., 1998.

DUIN, R. The combining classifier: To train or not to train? In: *International Conference on Pattern Recognition - ICPR'2002*. 2002. p. 765–770.

DUIN, R. P. W.; PEKALSKA, E.; TAX, D. M. J. The characterization of classification problems by classifier disagreements. In: *International Conference on Pattern Recognition - ICPR'2004*, 2004. p. 140–143.

EL-YACOUBI, A.; SABOURIN, R.; SUEN, C. Y. An hmm-based approach for off-line unconstrained handwritten word modeling and recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, v. 21, n. 8, p. 752–760, 1999.

FACON, J. *Morfologia Matemática: Teoria e Exemplos*. PUC-PR, 1996.

FILHO, J. G. *Gestalt do objeto - sistema de leitura visual da forma*. Escrituras, 2000.

FORESEE, F. D.; HAGAN, M. T. Gauss-newton approximation to bayesian regularization. In: *International Joint Conference on Neural Networks - IJCNN'1997*. 1997. p. 1930–1935.

FREITAS, C. O. A. *Uso de Modelos Escondidos de Markov para Reconhecimento de Palavras Manuscritas*. Tese (Doutorado) — Pontifícia Universidade Católica do Paraná, 2001.

FREITAS, C. O. A. Percepção visual e reconhecimento de palavras manuscritas. Tese submetida ao concurso de promoção da carreira docente à classe de Professor Titular. 2002.

FREITAS, C. O. A.; BORTOLOZZI, F.; SABOURIN, R. Handwritten isolated word recognition: An approach based on mutual information for feature set validation. In: *International Conference on Document Analysis and Recognition - ICDAR'2001*. 2001. p. 665–669.

FREITAS, C. O. de A.; BORTOLOZZI, F.; SABOURIN, R. A strategy for selecting classes of symbols from classes of graphemes in hmm-based handwritten word recognition. *Revista Eletrônica de Sistemas de Informação - RESI*, v. 3, n. 1, 2004.

FREITAS, C. O. de A.; BORTOLOZZI, F.; SABOURIN, R. Study of perceptual similarities between different lexicons. *International Journal on Pattern Recognition and Artificial Intelligence*, v. 18, n. 7, p. 1321–1338, 2004.

FREITAS, C. O. de A. et al. Brazilian bank check handwritten legal amount recognition. In: *Simpósio Brasileiro de Computação Gráfica e Processamento de Imagens - SIBGRAPI'2000*. 2000. p. 97–104.

GILLOUX, M.; LEROUX, M.; BERTILLE, J. M. Strategies for cursive script recognition using hidden markov models. *Machine Vision and Applications*, v. 8, p. 197–205, 1995.

GONZALEZ, R. C.; WOODS, R. E. *Digital Image Processing*. Addison-Wesley, 1992.

GUILLEVIC, D. *Unconstrained Handwriting Recognition Applied to the Processing of Bank Cheques*. Tese (Doutorado) — University of Concordia, 1995.

HADJITODOROV, S. T.; KUNCHEVA, L. I.; TODOROVA, L. P. Moderate diversity for better cluster ensembles. *Available on-line at http://www.informatics.bangor.ac.uk/kuncheva/-recent_publications.htm*, 2005.

HAYKIN, S. *Neural Networks - A Comprehensive Foundation*. Prentice Hall, 1996.

HEUTTE, L. *Reconnaissance de Caractères Manuscrits: Application à la Lecture Automatique des Chèques et des Enveloppes Postales*. Tese (Doutorado) — Université de Rouen, 1994.

HO, T.; HULL, J.; SRIHARI, S. N. Decision combination in multiple classifier systems. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, v. 16, n. 1, p. 66–75, 1994.

HOLT, C. M. et al. An improved parallel thinning algorithm. *Communications of the ACM*, v. 30, n. 2, p. 156–160, 1987.

KAPP, M. N. *Reconhecimento de Palavras Manuscritas Utilizando Redes Neurais Artificiais*. Dissertação (Mestrado) — Pontifícia Universidade Católica do Paraná, 2004.

KAPP, M. N.; FREITAS, C. O. A.; SABOURIN, R. Evaluating the conventional and class-modular architectures feedforward neural networks for handwritten word recognition. In: *Simpósio Brasileiro de Computação Gráfica e Processamento de Imagens - SIBGRAPI'2003*. 2003. p. 315–319.

KIM, G. *Recognition of Off-Line Handwritten Words and Its Extension to Phrase Recognition*. Tese (Doutorado) — State University of New York at Buffalo, 1996.

KIM, J. H. et al. A methodology of combining hmm and mlp classifiers for cursive word recognition. In: *International Conference on Pattern Recognition - ICPR'2000*. 2000. p. 2319–2322.

KITTLER, J. et al. On combining classifiers. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, v. 20, n. 3, p. 226–239, 1998.

KUNCHEVA, L. I.; WHITAKER, C. J. Measures of diversity in classifer ensembles and their relationship with the ensemble accuracy. *Machine Learning*, v. 51, n. 2, p. 181–207, 2003.

LAM, L.; SUEN, C. Y. Optimal combinations of pattern classifiers. *Pattern Recognition Letters*, v. 16, n. 3, p. 945–954, 1995.

LECOLINET, E. *Segmentation d'Images de Mots Manuscrits: Application à La Lecture de Chaînes de Caractères Majuscules Alphanmériques et à La Lecture de L'écriture Cursive*. Tese (Doutorado) — Université Pierre et Marie Curie (Paris VI), 1990.

LII, J.; PALUMBO, P. W.; SRIHARI, S. N. Address block location using character recognition and address syntax. In: *International Conference on Document Analysis and Recognition - ICDAR'1993*. 1993. p. 330–334.

MACKAY, D. J. C. Bayesian interpolation. *Neural Computation*, v. 4, n. 3, p. 415–447, 1992.

MADHVANATH, S.; GOVINDARAJU, V. The role of holistic paradigms in handwritten word recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, v. 23, n. 2, p. 149–164, 2001.

MANGABEIRA, J. A. C.; AZEVEDO, E. C.; LAMPARELLI, R. A. C. Avaliação do levantamento do uso das terras por imagens de satélite de alta e média resolução espacial. *Comunicado Técnico - Embrapa*, v. 1, n. 11, p. 1–15, 2003.

MARINAI, S.; GORI, M.; SODA, G. Artificial neural networks for document analysis and recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, v. 27, n. 1, p. 23–35, 2005.

MATOS, L. N. *Utilização de Redes Bayesianas como Agrupador de Classificadores Locais e Global*. Tese (Doutorado) — Universidade Federal de Campina Grande, 2004.

MOHAMED, M. A.; GADER, P. Generalized hidden markov models - Part I: Theoretical frameworks. *IEEE Transactions on Fuzzy Systems*, v. 8, n. 1, p. 67–81, 2000.

MORITA, M. et al. Segmentation and recognition of handwritten dates: An hmm-mlp hybrid approach. *International Journal on Document Analysis Recognition*, v. 6, p. 248–262, 2004.

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

OLIVEIRA Jr., J. J. de. *Avaliação de Conjuntos de Características no Reconhecimento de Palavras Manuscritas*. Dissertação (Mestrado) — Universidade Federal de Campina Grande, 2002.

OLIVEIRA Jr., J. J. de et al. Evaluation of handwritting recognition by hybrid nn and hmm classifiers. In: *Simpósio Brasileiro de Computação Gráfica e Processamento de Imagens - SIBGRAPI'2002*. 2002. p. 210–217.

OTSU, N. A threshold selection method from gray level histograms. *IEEE Transactions on System, Man and Cybernetics*, v. 9, n. 1, p. 62–66, 1979.

PARKER, J. R. *Algorithms for Image Processing and Computer Vision*. Jonh Wiley & Sons, 1997.

PLAMONDON, R.; SRIHARI, S. N. On-line and off-line handwriting recognition: A comprehensive survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, v. 22, n. 1, p. 63–84, 2000.

RABINER, L. R. A tutorial on hidden markov models and selected applications in speech recognition. *Proceedings of the IEEE*, v. 77, n. 2, p. 257–286, 1989.

RICHARD, M. D.; LIPPMANN, R. P. Neural network classifiers estimate bayesian *a posteriori* probabilities. *Neural Computation*, v. 3, p. 461–483, 1991.

RUTA, D.; GABRYS, B. Classifier selection for majority voting. *Information Fusion*, v. 6, n. 1, p. 63–81, 2005.

RUZON, M. A. *Texture Segmentation: An Introduction Primer*. Disponível on-line: http://robotics.stanford.edu/ ruzon/texseg/, 1997.

SCHALKOFF, R. *Pattern Recognition - Statistical, Structural and Neural Approaches*. Jonh Wiley & Sons, 1992.

SCHOMAKER, L. Artificial intelligence and natural perception. In: *International Produktionstechnisches Kolloquium - PTK'2004*. Berlin, 2004. p. 373–374.

SCHOMAKER, L.; SEGERS, E. A method for the determination of features used in human reading of cursive handwriting. In: *International Workshop on Frontiers for Handwritten Recognition - IWFHR'1998*. The Netherlands, 1998. p. 157–168.

SEKULER, R.; BLAKE, R. *Perception*. McGraw-Hill Inc., 1994.

SERIES, H. S. 5200c. Http://www.pandi.hp.com/pandi-db/prodinfo.main?-product=scanjet5200c&Region=non_us.

SUSSMANN, H. J. Uniqueness of the weights for minimal feedforward nets with a given input-output map. *Neural Networks*, v. 5, n. 4, p. 589–593, 1992.

TAPPERT, C. C.; SUEN, C. y.; WAKAHARA, T. The state of art in on-line handwritting recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, v. 12, n. 8, p. 787–808, 1990.

TRIER, O. D.; JAIN, A. K.; TAXT, T. Feature extraction methods for character recognition - a survey. *Pattern Recognition*, v. 29, n. 4, p. 641–662, 1996.

TSAI, W.-H. Moment-preserving thresholding: A new approach. *Computer Vision, Graphics and Image Processing*, v. 29, p. 377–393, 1985.

TSYMBAL, A.; PECHENIZKIY, M.; CUNNINGHAM, P. Diversity in search strategies for ensemble feature selection. *Information Fusion*, v. 6, n. 1, p. 83–98, 2005.

VELOSO, L. R. Sistema de reconhecimento de palavras manuscritas dependente do usuário para a língua portuguesa. Proposta de Tese. 2001.

VERNON, M. D. *Percepção e Experiência*. Editora Perspectiva, 1974.

WEBB, A. *Statistical Pattern Recognition*. Jonh Wiley & Sons, 2002.

WINDEATT, T. Diversity measures for multiple classifier system analysis and design. *Information Fusion*, v. 6, n. 1, p. 21–36, 2005.

XU, L.; KRZYZAK, A.; SUEN, C. Y. Methods of combining multiple classifiers and their applications to handwriting recognition. *IEEE Transactions on Systems, Man and Cybernetics*, v. 22, n. 3, p. 418–435, 1992.

YAN, H. Unified formulation of a class of image thresholding techniques. *Pattern Recognition*, v. 29, n. 12, p. 2025–2031, 1996.

YONEKURA, E. A.; FACON, J. Postal envelope segmentation by 2d histogram clustering through watershed transform. In: *International Conference on Document Analysis and Recognition - ICDAR'2003*. 2003. p. 338–342.

YU, B.; JAIN, A. K.; MOHIUDDIN, M. Address block location on complex mail pieces. In: *International Conference on Document Analysis and Recognition - ICDAR'1997*. 1997. p. 897–901.

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

ZOUARI, H. K. *Contribution à L'évaluation des Méthodes de Combinaison Parallèle de Classifieurs par Simulation*. Tese (Doutorado) — Université de Rouen, 2004.