

STEFANO KUBIÇA

METODOLOGIA PARA
MELHORAMENTO DE CONTEÚDOS
IMPRESSOS DE IMAGENS DE
DOCUMENTOS COMPLEXOS

Dissertação apresentada à Pontifícia
Universidade Católica do Paraná para a
obtenção do título de Mestre em Informática
Aplicada.

Área de concentração:
Análise e Reconhecimento de Documentos

Orientador:
Jacques Facon

Curitiba
2004

STEFANO KUBIÇA

**METODOLOGIA PARA
MELHORAMENTO DE CONTEÚDOS
IMPRESSOS DE IMAGENS DE
DOCUMENTOS COMPLEXOS**

Dissertação apresentada à Pontifícia
Universidade Católica do Paraná para a
obtenção do título de Mestre em Informática
Aplicada.

Curitiba
2004

Dedico este trabalho aos meus pais, Miguel e Sofia, que em um período difícil de nossas vidas, não mediram esforços para que eu pudesse continuar meus estudos.

Agradecimentos

A Deus que está comigo em todos os momentos da minha vida e que sem Ele, nem esta simples frase poderia ter sido escrita.

Aos meus filhos Izabel e Davi pelos momentos de alegria e descontração durante o desenvolvimento deste trabalho.

A CELEPAR pela oportunidade e apoio com importantes recursos para o desenvolvimento deste trabalho.

Ao Luiz Carlos de Almeida Oliveira pelo decisivo apoio, sem o qual este projeto não teria sido iniciado.

Ao professor orientador e amigo Jacques Facon, pela condução do projeto, apoio, dedicação, incentivo e compreensão em momentos de dificuldade.

A Maria Cristina Rocha Pombo Lessi pelo inestimável apoio, compreensão e importante colaboração na revisão do texto.

A todos os professores que com os seus conhecimentos, contribuíram de forma efetiva para a realização deste trabalho.

Aos meus amigos e todos que direta ou indiretamente contribuíram para este trabalho.

Sumário

Agradecimentos	ii
Sumário	iii
Lista de Figuras	vii
Lista de Tabelas	xii
Resumo	xv
Abstract	xvi
Introdução	1
Capítulo 1	
Informações Gerais	3
1.1 Desafio.....	3
1.2 Motivação.....	4
1.3 Proposta.....	4
1.4 Contribuição.....	5
1.5 Organização.....	6
Capítulo 2	
Documentos Complexos	7
2.1 Introdução.....	7
2.2 Características.....	7
2.3 Captura de Imagens de Documentos.....	10
2.3.1 Introdução.....	10
2.3.2 Captura de Documentos Simples.....	11
2.3.3 Captura de Documentos Complexos.....	12
2.3.4 Considerações.....	14
2.4 Rotulação.....	14
2.5 Base de Dados.....	16

Capítulo 3	
Estado da Arte	17
3.1 Introdução	17
3.2 Abordagens para Documentos Complexos	18
3.3 Considerações	24
Capítulo 4	
Limiarização	25
4.1 Introdução	25
4.2 Limiarização Global	27
4.3 Métodos de Limiarização Global	28
4.3.1 Método de Otsu	28
4.3.2 Método de Kittler e Illingworth	29
4.3.3 Método por Entropia de Pun	30
4.3.4 Método por Entropia de Kapur	31
4.3.5 Método por Entropia de Johannsen	31
4.3.6 Método por Entropia de Li-Lee	32
4.3.7 Método por Entropia de WuLu	33
4.3.8 Método por Abordagem Fuzzy de Huang	34
4.3.9 Método por Abordagem Fuzzy de Yager	35
4.3.10 Método por Abordagem Fuzzy C Means	36
4.3.11 Método por Seleção Iterativa de Thrussel	37
4.3.12 Método por Seleção Iterativa de Lloyd	38
4.4 Limiarização Local Adaptativa	38
4.5 Métodos de Limiarização Local Adaptativa	39
4.5.1 Método de Bernsen	39
4.5.2 Método de Niblack	39
4.5.3 Método de Eikvil	40
4.6 Escolha de um Método de Limiarização	40
Capítulo 5	
Morfologia Matemática Binária	43
5.1 Introdução	43
5.2 Operadores Morfológicos Binários	45
5.2.1 Erosão Binária	45
5.2.2 Dilatação Binária	47
5.2.3 Abertura Binária	48
5.2.4 Fechamento Binário	51
5.2.5 Reconstrução Binária	53
5.3 Escolha de Operadores, Elementos Estruturantes e Iterações	54
5.3.1 Escolha de Operadores Morfológicos	54

5.3.2 Escolha de Elementos Estruturantes	55
5.3.3 Escolha do Número de Iterações.....	56
Capítulo 6	
Metodologia Proposta e Testes	57
6.1 Metodologia Proposta	57
6.1.1 Fase de Captura	61
6.1.2 Fase de Rotulação	61
6.1.3 Fase de Processamento	63
6.1.4 Fase de Reconhecimento	65
6.1.5 Fase de Digitação	66
6.1.6 Fase de Avaliação da Qualidade.....	67
6.2 Resultados dos Testes.....	70
6.2.1 Resultados do Rótulo AFE	71
6.2.2 Resultados do Rótulo AFC	73
6.2.3 Resultados do Rótulo NFE.....	75
6.2.4 Resultados do Rótulo NFC.....	77
6.2.5 Resultados do Rótulo PFC	79
Capítulo 7	
Validação e Análise dos Resultados	81
7.1 Resultados da Validação.....	81
7.1.1 Resultados do Rótulo AFE	83
7.1.2 Resultados do Rótulo AFC	85
7.1.3 Resultados do Rótulo NFE.....	87
7.1.4 Resultados do Rótulo NFC.....	89
7.1.5 Resultados do Rótulo PFC	91
7.2 Análise dos Resultados	93
7.2.1 Introdução	93
7.2.2 Análise dos Resultados Quantitativos.....	93
7.2.3 Análise dos Problemas Encontrados.....	103
7.2.4 Aplicação da Metodologia em Outros Tipos de Documento.....	106
Capítulo 8	
Conclusões, Contribuições e Perspectivas	108
8.1 Conclusões Sobre o Trabalho.....	108
8.1.1 Considerações Gerais	108
8.1.2 Conclusões Sobre os Resultados	109
8.2 Contribuições do Trabalho.....	111
8.2.1 Contribuições para a Pesquisa Científica.....	111
8.2.2 Contribuições para Aplicação nas Empresas.....	111

8.3 Perspectivas para o Futuro	112
8.3.1 Perspectivas para a Pesquisa	112
8.3.2 Perspectivas para as Empresas	113
Capítulo 9	
Referências Bibliográficas	114

Lista de Figuras

Figura 2.1 Tipo de documento simples: Texto impresso em fundo claro com bom contraste entre o objeto e o fundo.....	8
Figura 2.2 Tipo de Documento complexo: Capturado em nível cinza com baixa qualidade dos conteúdos impressos.....	8
Figura 2.3 Documento simples capturado em 256 níveis de cinza	12
Figura 2.4 Documento simples capturado em preto e branco (limiarização efetuada pelo equipamento de captura)	12
Figura 2.5 Documento complexo em seu formato e características originais.....	13
Figura 2.6 Documento complexo capturado em tonalidades de 256 níveis de cinza.....	13
Figura 2.7 Documento complexo capturado em preto e branco (limiarizado pelo equipamento de captura).....	13
Figura 2.8 Região com conteúdo de caracteres alfabéticos impressos em fundo escuro, denominada como: AFE	15
Figura 2.9 Região com conteúdo de caracteres alfabéticos impressos em fundo claro, denominada como: AFC	15
Figura 2.10 Região com conteúdo de caracteres numéricos impressos em fundo escuro, denominada como: NFE	15
Figura 2.11 Região com conteúdo de caracteres numéricos impressos em fundo claro, denominada como: NFC	15
Figura 2.12 Região com conteúdo de caracteres alfabéticos pré-impressos em fundo claro, denominada como: PFC	15
Figura 3.1 Resultado do processo de filtragem em níveis de cinza; (A) imagem original, (B) imagem recuperada.....	19
Figura 3.2 Resultado do processo de filtragem em imagens binárias; (A) imagem sem filtragem antes da limiarização, (B) imagem recuperada com filtro antes da limiarização.	19

Figura 3.3 Processamento de página da Bíblia; (A) processada, (B) reconhecida por OCR.....	22
Figura 3.4 Resultado do melhoramento de conteúdo manuscrito; parte superior, imagem original; parte inferior, imagem melhorada.....	23
Figura 3.5 Resultado do melhoramento de conteúdo manuscrito; parte superior, imagem original; parte inferior, imagem melhorada.....	23
Figura 4.1 Efeitos da variação do limiar em conteúdo impresso: (A) limiar muito alto; (B) limiar ideal; (C) limiar muito baixo.....	26
Figura 4.2 Histograma de imagem de documento com bom contraste entre fundo e objeto, típico de documento simples	27
Figura 4.3 Histograma de imagem com componentes de objetos esparsos, típico de documento complexo.....	28
Figura 5.1 Tipos de elementos estruturantes básicos utilizados na morfologia matemática binária	44
Figura 5.2 Exemplo de iteração do elemento estruturante em uma imagem binária.....	44
Figura 5.3 Efeitos da erosão binária com elemento estruturante quadrado duas iterações: (a) imagem original; (b) imagem com erosão.	46
Figura 5.4 Efeitos da dilatação binária com elemento estruturante quadrado duas iterações: (a) imagem original; (b) imagem com dilatação.....	48
Figura 5.5 Efeitos da abertura binária com elemento estruturante quadrado duas iterações: (a) imagem original; (b) imagem com abertura.	51
Figura 5.6 Efeitos do fechamento binário com elemento estruturante quadrado duas iterações: (a) imagem original; (b) imagem com fechamento.	53
Figura 5.7 Processo da reconstrução binária: (a) imagem original; (b) marcador obtido pela erosão da imagem original com elemento estruturante quadrado duas iterações; (c) imagem reconstruída a partir do marcador, utilizando a imagem original como máscara.	54
Figura 6.1 Taxas de reconhecimento em documentos simples versus taxas de reconhecimento em documentos complexos	58
Figura 6.2 Mesma abordagem para documentos simples e complexos em soluções de GED.	59
Figura 6.3 Imagem de documento complexo, digitalizado utilizando a abordagem tradicional.....	59

Figura 6.4 Visão geral das fases da metodologia proposta.....	61
Figura 6.5 Ferramenta desenvolvida em Visual C para auxiliar na rotulação	62
Figura 6.6 Ferramenta desenvolvida em Visual C para processamento dos rótulos.....	63
Figura 6.7 Configuração dos parâmetros da morfologia matemática binária utilizados para elementos estruturantes e número de iterações.....	64
Figura 6.8 Parâmetros do OCR OMNIPAGE10 utilizados em todos os processamentos.....	66
Figura 6.9 Visão geral da geração dos dados necessários para avaliação da qualidade	67
Figura 6.10 Ferramenta desenvolvida em Visual C para avaliação da qualidade..	67
Figura 6.11 Exemplo de avaliação da qualidade na etapa de geração de resultado parcial.	68
Figura 6.12 Análise comparativa dos resultados de teste para o rótulo AFE: Método de Otsu versus WuLu, Li-Lee e LLoyd; acrescidos dos melhores resultados da morfologia matemática binária	72
Figura 6.13 Análise comparativa entre a abordagem tradicional e o melhor resultado obtido nos testes da metodologia proposta para o rótulo AFE.....	72
Figura 6.14 Análise comparativa dos resultados de teste para o rótulo AFC: Método de Otsu versus WuLu, Li-Lee e LLoyd; acrescidos dos melhores resultados da morfologia matemática binária	74
Figura 6.15 Análise comparativa entre a abordagem tradicional e o melhor resultado obtido nos testes da metodologia proposta para o rótulo AFC.....	74
Figura 6.16 Análise comparativa dos resultados de teste para o rótulo NFE: Método de Otsu versus WuLu, Li-Lee e LLoyd; acrescidos dos melhores resultados da morfologia matemática binária	76
Figura 6.17 Análise comparativa entre a abordagem tradicional e o melhor resultado obtido nos testes da metodologia proposta para o rótulo NFE.....	76
Figura 6.18 Análise comparativa dos resultados de teste para o rótulo NFC: Método de Otsu versus WuLu, Li-Lee e LLoyd; acrescidos dos melhores resultados da morfologia matemática binária	78
Figura 6.19 Análise comparativa entre a abordagem tradicional e o melhor resultado obtido nos testes da metodologia proposta para o rótulo NFC.....	78

Figura 6.20 Análise comparativa dos resultados de teste para o rótulo PFC: Método de Otsu versus WuLu, Li-Lee e LLoyd; acrescidos dos melhores resultados da morfologia matemática binária 80

Figura 6.21 Análise comparativa entre a abordagem tradicional e o melhor resultado obtido nos testes da metodologia proposta para o rótulo PFC 80

Figura 7.1 Análise comparativa dos resultados da validação para o rótulo AFE: Método de Otsu versus WuLu, Li-Lee e LLoyd; acrescidos dos melhores resultados da morfologia matemática binária 84

Figura 7.2 Análise comparativa entre a abordagem tradicional e o melhor resultado obtido na validação da metodologia proposta para o rótulo AFE... 84

Figura 7.3 Análise comparativa dos resultados da validação para o rótulo AFC: Método de Otsu versus WuLu, Li-Lee e LLoyd; acrescidos dos melhores resultados da morfologia matemática binária 86

Figura 7.4 Análise comparativa entre a abordagem tradicional e o melhor resultado obtido na validação da metodologia proposta para o rótulo AFC... 86

Figura 7.5 Análise comparativa dos resultados da validação para o rótulo NFE: Método de Otsu versus WuLu, Li-Lee e LLoyd; acrescidos dos melhores resultados da morfologia matemática binária 88

Figura 7.6 Análise comparativa entre a abordagem tradicional e o melhor resultado obtido na validação da metodologia proposta para o rótulo NFE..... 88

Figura 7.7 Análise comparativa dos resultados da validação para o rótulo NFC: Método de Otsu versus WuLu, Li-Lee e LLoyd; acrescidos dos melhores resultados da morfologia matemática binária 90

Figura 7.8 Análise comparativa entre a abordagem tradicional e o melhor resultado obtido na validação da metodologia proposta para o rótulo NFC..... 90

Figura 7.9 Análise comparativa dos resultados da validação para o rótulo PFC: Método de Otsu versus WuLu, Li-Lee e LLoyd; acrescidos dos melhores resultados da morfologia matemática binária 92

Figura 7.10 Análise comparativa entre a abordagem tradicional e o melhor resultado obtido na validação da metodologia proposta para o rótulo PFC..... 92

Figura 7.11 Análise comparativa por rótulo da abordagem tradicional versus método de Otsu versus demais métodos de limiarização utilizados..... 96

Figura 7.12 Incremento do melhor resultado da morfologia binária sobre a limiarização de Otsu..... 97

Figura 7.13 Incremento do melhor resultado da morfologia binária sobre a limiarização de Li-Lee	97
Figura 7.14 Incremento do melhor resultado da morfologia binária sobre a limiarização de LLoyd	98
Figura 7.15 Incremento do melhor resultado da morfologia binária sobre a limiarização de WuLu.....	98
Figura 7.16 Análise comparativa dos resultados entre a abordagem tradicional, limiarização múltipla e morfologia matemática binária	100
Figura 7.17 Problema para o reconhecimento via OCR: (A) imagem original, (B) imagem segmentada, (C) segmentos onde não é possível obter reconhecimento via OCR.....	104
Figura 7.18 Parte superior esquerda de uma imagem da base, recortada manualmente em função da dificuldade de encontrar o ponto de corte na margem superior por excesso de ruído.....	105
Figura 7.19 Problema de contraste no conteúdo impresso: (A) original, (B) com ENWU+EROC01 deteriora a parte “fraca”, (C) com ENLI+ABEc01 deixa borrada a parte “forte”	106

Lista de Tabelas

Tabela 2.1 Quantidades de documentos da base de dados para teste e validação por rótulo (região)	16
Tabela 4.1 Resultados preliminares de testes nos métodos de limiarização global implementados no laboratório LUCI da PUC-Pr	42
Tabela 4.2 Resultados preliminares de testes nos métodos de limiarização local adaptativa implementados no laboratório LUCI da PUC-Pr	42
Tabela 5.1 Tempo e resultado visual das operações com Morfologia Matemática Binária (EE = Elementos Estruturantes)	55
Tabela 5.2 Resultado visual com elementos estruturantes (EE) não adequados para conteúdos impressos	55
Tabela 5.3 Tempo e Resultado visual com progressão do número de iterações.....	56
Tabela 6.1 Exemplo de resultados parciais do rótulo PFC para os métodos de processamento: ENWU+EROC01	69
Tabela 6.2 Exemplo de resultado da totalização de PFC para os métodos de processamento: ENWU+EROC01	69
Tabela 6.3 Resultados de teste da limiarização combinada com morfologia binária do rótulo AFE.....	71
Tabela 6.4 Resultados de teste da limiarização combinada com morfologia binária do rótulo AFC.....	73
Tabela 6.5 Resultados de teste da limiarização combinada com morfologia binária do rótulo NFE	75
Tabela 6.6 Resultados de teste da limiarização combinada com morfologia binária do rótulo NFC	77
Tabela 6.7 Resultados de teste da limiarização combinada com morfologia binária do rótulo PFC.....	79
Tabela 7.1 Resultados de validação da limiarização combinada com morfologia binária do rótulo AFE.....	83

Tabela 7.2 Resultados de validação da limiarização combinada com morfologia binária do rótulo AFC.....	85
Tabela 7.3 Resultados de validação da limiarização combinada com morfologia binária do rótulo NFE	87
Tabela 7.4 Resultados de validação da limiarização combinada com morfologia binária do rótulo NFC	89
Tabela 7.5 Resultados de validação da limiarização combinada com morfologia binária do rótulo PFC.....	91
Tabela 7.6 Exemplo de resultado visual conforme as técnicas aplicadas, onde existe uma relação direta crescente entre a qualidade visual e as taxas de reconhecidos corretos para o rótulo NFC.....	94
Tabela 7.7 Aspecto visual comparado com a taxa de reconhecidos corretos, onde da limiarização por Otsu para WuLu existe pouca diferença visual mas grande diferença na taxa de reconhecimento.....	94
Tabela 7.8 Análise comparativa por rótulo da abordagem tradicional versus método de Otsu versus demais métodos de limiarização utilizados	95
Tabela 7.9 Incrementos nas taxas de reconhecidos corretos com a utilização da morfologia matemática binária (melhores resultados) sobre os resultados dos respectivos métodos de limiarização	96
Tabela 7.10 Análise comparativa dos resultados obtidos entre a abordagem tradicional versus limiarização múltipla versus morfologia matemática binária.....	100
Tabela 7.11 Técnicas de processamento que obtiveram os melhores resultados para cada rótulo, com taxas de reconhecidos corretos e tempos de processamento acumulados.....	101
Tabela 7.12 Resultado com baixa taxa de reconhecimento e alta taxa de caracteres não reconhecidos.....	102
Tabela 7.13 Efeitos da limiarização WuLu com duas iterações da erosão binária no rótulo AFC	102
Tabela 7.14 Abordagens apresentando mesma taxa de reconhecidos corretos e taxas diferentes para não reconhecidos.....	102
Tabela 7.15 Acréscimos nas taxas de reconhecimento versus aumento de tempo de processamento, comparando a metodologia proposta com a abordagem tradicional.....	103

Tabela 7.16 Aplicação da metodologia proposta em outros tipos de documento.... 106

Resumo

Desde a invenção dos caracteres móveis em 1440 por Johann Gutemberg, o registro de informações impressas em papel na forma de documentos tem crescido em escala geométrica. Com a chegada dos computadores no século 20, o papel começa dar lugar aos bits, onde muitos dos bilhões de documentos gerados em papel nestes últimos séculos estão sendo convertidos para o meio eletrônico. A principal característica de um documento impresso é possuir um conteúdo que possa ser lido, seja pela visão humana quando se trata de documentos em papel, seja pela visão computacional quando se trata de documentos eletrônicos. Assim a qualidade do conteúdo de um documento impresso, está diretamente relacionada à possibilidade de leitura desse conteúdo.

Muitas pesquisas com êxito têm sido realizadas na área de reconhecimento de padrões, o que tem permitido a “leitura computacional” com altas taxas de reconhecimento em imagens de documentos simples. Porém para muitos documentos complexos, os resultados deixam a desejar, e na maioria desses casos, a dificuldade se dá pelas características e qualidade desse tipo de documento. Como em geral as aplicações de gerenciamento eletrônico de documentos manipulam muitos milhões de documentos, encontrar técnicas adequadas para melhorar a qualidade e ao mesmo tempo viabilizar grandes volumes é um enorme desafio.

Este trabalho, com o objetivo de apresentar uma metodologia para melhoramento da qualidade de conteúdos impressos de documentos complexos em grandes volumes, mostrou que com a combinação de técnicas adequadas de processamento de imagens, como a limiarização e a morfologia matemática binária, pode-se obter um significativo aumento na qualidade em relação a principal abordagem até então utilizada. Validada com 402 imagens de documentos complexos e tendo a qualidade avaliada por reconhecimento, uma nova abordagem mostra-se viável.

Palavras-chave: documentos complexos, qualidade conteúdos impressos, limiarização múltipla, morfologia matemática binária, reconhecimento.

Abstract

Since the invention of the mobile characters in 1440 by Johann Gutemberg, the register of printed information on paper as document has been grown on geometric scale. With the arrival of the computers at the 20th century, paper start to make way for the bits, where billions of documents created on paper last centuries are being converted to the electronic medium. The main characteristic of a printed document is to have a content that can be read by the human vision, when in document paper, as much as by computational vision, when in electronic documents. So, the quality of the content of the printed document is directly related to the possibility of the content reading.

Many successful researches have been accomplished at the patterns recognition area. This has allowed the “computational reading” with high rates of recognition in simple document images. But, for many complex documents, the results are unsatisfactories, and in many cases the difficulties are because of the characteristics and quality of this kind of documents. As the use of electronic documents management in general manipulate many millions of documents, to find appropriated techniques to improve the quality and at the same time make possible great volumes, is a huge challenge.

This dissertation, that objectives to present a methodology to improve the quality of the printed contents of the complex documents in large volumes, has showed that with the combination of appropriated techniques of the images processing as the a thresholding, and the binary mathematics morphology, it can be obtained a significant grow of quality in relation to the principal approach then used. Validated with 402 images of the complex documents and having the quality appraised by recognition, the new approach shows itself perfectly viable.

Keywords: complex documents, printed contents quality, multiple thresholding, binary mathematics morphology, recognition.

Introdução

A palavra documento, tem sua origem do Latim *documentu* e segundo o dicionário Aurélio [AURÉLIO, 1991] é qualquer base de conhecimento elaborada de maneira que se possa utilizar para consulta, estudo, prova, etc. Documentos surgiram quando a humanidade percebeu a necessidade de registrar informações. Hoje, documentos fazem parte tanto de nossas vidas que um dos fatos mais importantes quando nascemos é gerar um documento, a certidão de nascimento. Muito conhecimento gerado pela humanidade nestes últimos séculos está registrado em documentos.

Documentos podem variar muito em forma e meio pelo qual são apresentados. Podemos considerar como sendo um documento, desde uma anotação feita em um pequeno pedaço de papel, até um grande conjunto de dados armazenados em computadores.

O uso de documentos no mundo tem crescido em escala geométrica e a tecnologia da informação, através do Gerenciamento Eletrônico de Documentos (GED), tem permitido que documentos sejam gerados, armazenados e recuperados eletronicamente de forma segura e a custos razoáveis. Mas o grande desafio hoje é a garantia da qualidade dos conteúdos impressos das imagens destes documentos.

Escritórios particulares e departamentos de órgãos públicos, vivem abarrotados de papel.

Os principais problemas com documentos em papel são:

- Espaço físico para guarda dos documentos,
- Deterioração pelo manuseio dos documentos e
- Disponibilidade para uso.

Os documentos podem ser guardados, manuseados e disponibilizados de forma eletrônica. Qualquer documento em formato analógico como o papel ou microfilme, pode ser convertido para o formato digital através de um processo de captura conhecido como digitalização. Através da digitalização, gera-se uma imagem eletrônica do

documento. Uma imagem digital pode ser representada por uma matriz de pontos chamados pixels, onde os índices de linha e coluna referenciam o brilho médio no ponto correspondente do documento capturado.

O processamento eletrônico de documentos é uma área de estudo que vem crescendo no mundo inteiro. Com os atuais avanços dos sistemas de computação, principalmente em termos de armazenamento de dados e velocidade de processamento, existe hoje um maior interesse no estudo das técnicas de processamento de imagens de documentos [GORMAN, 1998].

Considerando-se que o conteúdo impresso é o que mais caracteriza um documento e que, aumentar o reconhecimento tanto visual como computacional desse conteúdo significa aumentar a sua qualidade, o objetivo maior deste trabalho é a busca de novas abordagens para geração de imagens de documentos com maior qualidade. Para viabilizar uma nova abordagem, este trabalho, tem como objetivo específico propor uma metodologia para melhorar conteúdos impressos de documentos complexos em grandes volumes, uma vez que, a abordagem tradicional em GED para documentos complexos não traz resultados satisfatórios. Os resultados da metodologia proposta, serão avaliados por reconhecimento automático das imagens processadas, calculando-se os incrementos nas taxas de reconhecimento obtidas por OCR [MORI, 1992] e dessa forma apurando-se a melhoria da qualidade dos conteúdos impressos.

Capítulo 1

Informações Gerais

1.1 Desafio

Os sistemas de Gerenciamento Eletrônico de Documentos (GED) têm como meta manipular milhões de imagens de documentos. Estas imagens devem ter uma qualidade aceitável e devem ocupar o menor espaço possível para não congestionar o tráfego de rede e viabilizar os custos de armazenamento. Tanto a melhoria da qualidade quanto a redução de espaço, podem ser obtidas aplicando-se técnicas de processamento de imagem. Existem muitas técnicas descritas mas a maioria não foi formulada para documentos em particular [GORMAN, 1998]. Atualmente a maioria dos estudos tem sido voltados para imagens médicas, mas com a iminente legalidade dos documentos eletrônicos, as imagens de documentos terão sua importância redobrada.

Os poucos estudos existentes com imagens de documentos não levam em consideração a velocidade de processamento. Existem técnicas onde o processamento de uma imagem capturada a partir de um documento com tamanho A4 pode demorar até 10 minutos. Naturalmente em aplicações onde os volumes são pequenos isto não consiste em problema, mas com imagens de documentos onde em muitos casos os volumes são contados em milhões, tais técnicas são inviáveis. Por exemplo: cinco milhões de imagens é um volume comum em sistemas de GED. Processar esta quantidade de imagens com técnicas hoje disponíveis que demoram 10 minutos por imagem, levaria 95 anos de processamento, o que seria absolutamente inviável.

Portanto o desenvolvimento de uma Metodologia para melhorar conteúdos impressos de imagens de documentos que tenha como requisito viabilizar o processamento de milhões de imagens em tempo razoável é um grande desafio.

1.2 Motivação

Atualmente, governos e empresas do mundo inteiro estão se mobilizando para aprovar leis que darão aos documentos eletrônicos o mesmo valor legal dos documentos em papel. Com isso haverá grande demanda para converter os acervos em papel e microfilme para imagens digitais.

Documentos em papel e microfilme nas Organizações, contam-se aos milhões e por aspectos legais muitas vezes devem ser preservados por dezenas de anos. Muitos destes documentos são antigos, encontram-se em estado precário de conservação e grande parte deles sofre processo acelerado de deterioração por manuseio.

O GED tem como principal objetivo solucionar problemas envolvendo gestão documental. Mas, soluções comerciais hoje existentes não conseguem gerar imagens em preto e branco com qualidade satisfatória quando os documentos são complexos.

Todos os aspectos acima citados, motivam estudos e pesquisas na área de processamento de imagens com o objetivo de melhorar a qualidade dos conteúdos impressos de imagens de documentos complexos.

1.3 Proposta

Uma das principais preocupações dos sistemas de Gerenciamento Eletrônico de Documentos é manipular imagens de documentos que tenham boa qualidade e que estejam produzidas em preto e branco para não congestionar o tráfego de rede e não inviabilizar os custos de armazenamento. Como sistemas de GED, na grande parte dos casos, trabalham com muitas imagens de documentos, é fator crítico de sucesso que a metodologia proposta considere o tempo de processamento.

A proposta deste trabalho é criar uma metodologia que permita viabilizar novas abordagens para gerar grandes quantidades de imagens em preto e branco com melhor qualidade em conteúdos impressos. O principal objetivo da busca da melhoria da qualidade é aumentar o reconhecimento visual e computacional de conteúdos impressos de documentos complexos.

A metodologia consiste em capturar os documentos em imagens com 256 níveis de cinza, regionalizar as imagens conforme características particulares de cada região, converter as regiões de 256 níveis de cinza para preto e branco utilizando o conceito de limiarização múltipla [GORMAN, 1998] e mostrar que é possível melhorar ainda mais a qualidade dos conteúdos impressos utilizando técnicas de Morfologia Matemática Binária.

Somente serão consideradas técnicas de Limiarização (*thresholding*) e Morfologia Matemática Binária cujo tempo de processamento viabilize o tratamento de grandes quantidades de imagens para atender aos requisitos de volumes dos sistemas de GED.

A avaliação da qualidade será por reconhecimento [GORMAN, 1998] através de OCR comercial [MORI, 1992] e os resultados serão quantificados pelo percentual de caracteres reconhecidos corretos, percentual de caracteres reconhecidos errados e percentual de caracteres não reconhecidos em cada uma das regiões consideradas do documento.

1.4 Contribuição

A metodologia proposta oferece como principais contribuições:

- Melhoria na qualidade de conteúdos impressos em imagens de documentos complexos com a combinação de técnicas de processamento de imagens;
- Aumento na taxa de reconhecimento automático de caracteres em imagens de documentos complexos, facilitando a conversão de imagem para texto e automatizando o processo de geração de índices em sistemas de GED;
- Incremento no nível de recuperação de imagens em aplicações de FTR (*Full Text Retrieval*) em função do aumento das taxas de reconhecimento;
- Redução de custos em estações de captura, diminuindo trabalhos de redigitalização no processo de controle de qualidade das imagens digitalizadas;
- Possibilidade de incluir novas técnicas, facilitando o desenvolvimento de trabalhos futuros na área de processamento de imagens de documentos.

1.5 Organização

Esta Dissertação está organizada em oito capítulos assim distribuídos:

- Este Capítulo apresentou a importância do trabalho, o desafio a ser vencido, os aspectos que motivaram o desenvolvimento desta pesquisa, a proposta para o desenvolvimento e as principais contribuições;
- O Capítulo II caracteriza um documento complexo, define a estrutura de um tipo de documento complexo utilizado neste trabalho e descreve a base utilizada para testar e validar a metodologia proposta;
- O Capítulo III apresenta a aplicabilidade do processamento de imagens em documentos e descreve de forma sucinta o estado da arte no processamento de imagens de documentos complexos com o objetivo de melhorar a qualidade dos conteúdos impressos;
- O Capítulo IV apresenta os principais métodos de Limiarização;
- O Capítulo V apresenta os principais operadores da Morfologia Matemática Binária;
- O Capítulo VI mostra as fases da metodologia proposta, a estratégia utilizada neste trabalho para melhorar a qualidade dos conteúdos impressos de documentos complexos e os resultados obtidos com o processamento da base de testes;
- O Capítulo VII apresenta os resultados obtidos com a aplicação da metodologia, mostrando uma análise quantitativa dos resultados e dos problemas encontrados;
- O Capítulo VIII apresenta as conclusões do trabalho, a contribuição, a importância e a possibilidade de trabalhos futuros.

Capítulo 2

Documentos Complexos

2.1 Introdução

A área de processamento de imagens, em um escopo maior, trata imagens de diversos tipos. Nesta pesquisa, tem-se especial interesse em imagens de documentos. Mesmo no âmbito dos documentos, pode-se ter imagens com características distintas. Uma classificação que pode ser feita, está relacionada a complexidade dos documentos, neste caso, tem-se documentos simples e documentos complexos. Neste capítulo, serão caracterizados os documentos simples e complexos, abordados aspectos relacionados a captura desses documentos, apresentada a necessidade de rotulação para documentos complexos e especificada a base de dados utilizada nesta pesquisa.

2.2 Características

Documentos simples (figura 2.1) caracterizam-se por apresentarem bom contraste entre conteúdos impressos e o fundo, facilitando sua captura e reconhecimento e em geral são compostos de documentos não estruturados tais como: cartas, memorandos, contratos, textos sobre legislação, etc. Os textos impressos em fundo claro são exemplos clássicos de documentos simples.

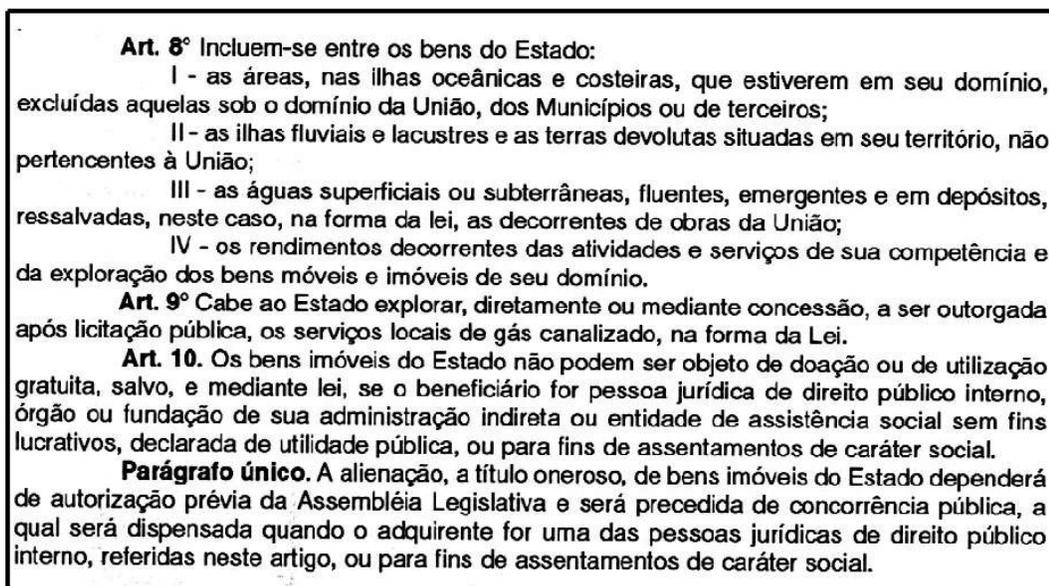


Figura 2.1 Tipo de documento simples: Texto impresso em fundo claro com bom contraste entre o objeto e o fundo



Figura 2.2 Tipo de Documento complexo: Capturado em nível cinza com baixa qualidade dos conteúdos impressos.

Documentos complexos (figura 2.2) são em sua maioria constituídos de documentos estruturados como por exemplo: documentos financeiros, contábeis, orçamentários, formulários para cadastros, etc., e caracterizam-se por apresentarem regiões com baixo contraste entre conteúdos impressos e o fundo, fundo complexo (como o artístico nos cheques bancários), fundo irregular, e também conteúdos impressos degradados, dificultando a captura e o reconhecimento. Em geral, a partir de documentos complexos, tem-se dificuldade em gerar imagens em preto e branco com

qualidade aceitável utilizando a abordagem tradicional (geração da imagem em preto e branco via scanner).

Documentos complexos podem estar degradados ou não. Documentos complexos não degradados podem gerar imagens degradadas, neste caso, a degradação acontece no processo de captura (geração da imagem). Por sua vez, documentos degradados sempre gerarão imagens degradadas, uma vez que o processo de captura é tão somente um processo de reprodução.

Segundo Baird [BAIRD, 2000], imagens degradadas são aquelas que possuem características como: baixa resolução de digitalização, manchas de tinta ou toner, deformações geométricas, etc. A degradação, segundo Baird [BAIRD, 2000], pode ser causada por uma série de motivos como por exemplo:

- Conteúdo desfocado pela digitalização
- Método inadequado de binarização (Limiarização)
- Impressão fraca
- Baixo contraste do conteúdo impresso
- Iluminação não uniforme
- Imperfeição de caracteres tipográficos
- Manchas e dobras na superfície do papel
- Respingos e borrões provenientes de tinta ou toner
- Adição de ruídos por componentes eletrônicos
- Papel desgastado, desbotado ou deformado

Em seu estudo sobre qualidade de imagens de documentos Baird [BAIRD, 1999] menciona o fato de que existe uma grande quantidade de documentos legados de baixa qualidade e que ainda não estamos providos de metodologias para avaliação da qualidade de imagens para que se possa de forma sistemática incrementar as taxas de reconhecimento de imagens degradadas. O método de Kanungo [KANUNGO, 1996] para estimar parâmetros de imagens degradadas utiliza técnica estatística, mas conforme Baird [BAIRD, 2000], apesar dos esforços, a modelagem da degradação de imagens de documentos ainda deixa muitas questões em aberto como, por exemplo, a falta de um referencial teórico para validação de modelos que possa prover uma base rigorosa para

critérios objetivos, empíricos e computáveis. Diante dessas dificuldades e com o objetivo de direcionar esta pesquisa para técnicas de processamento de imagens de documentos complexos para melhorar a qualidade e o reconhecimento de conteúdos impressos, a modelagem da degradação das imagens de documentos não é considerada. Propõe-se aqui, após a captura de uma imagem a partir de um documento complexo, proceder a verificação visual da degradação da imagem e suas regiões e realizar a rotulação para que técnicas mais adequadas possam ser utilizadas em cada uma das regiões.

2.3 Captura de Imagens de Documentos

2.3.1 Introdução

A aquisição de imagens, consiste em obter conteúdos digitais a partir de conteúdos analógicos como o papel e o microfilme. O bom resultado no processamento de imagens de documentos depende da qualidade da imagem gerada. Normalmente, tem-se uma estação de captura de documentos onde um dos principais componentes é um sensor. De forma geral, os sensores fornecem um sinal analógico correspondente a amostragem seqüencial da imagem. Como a maioria das técnicas de processamento de imagens é realizada de forma numérica em um computador, é necessária a discretização do sinal analógico. A amostragem da imagem contida no sinal analógico é obtida por transformação via conversor analógico-digital controlado por um sinal de relógio de um dispositivo digitalizador também conhecido como scanner. Cada transformação gera um ponto conhecido como pixel e o resultado final é uma matriz de pixels que corresponde a uma imagem eletrônica produzida a partir de um documento original.

Dependendo da necessidade da aplicação, as imagens podem ser adquiridas em preto e branco, em graduações de nível de cinza ou coloridas. Imagens em preto e branco são pobres em detalhes mas têm a vantagem de requerer pouco espaço de armazenamento e os procedimentos para processar imagens em preto e branco são geralmente simples e rápidos. Imagens em níveis de cinza têm uma escala que varia de 2 (preto e branco) até 256. Quanto maior o nível de cinza, maior e a riqueza de detalhes da imagem, porém, maior também a necessidade de espaço para o armazenamento e os

procedimentos para manipulação geralmente são complexos e mais lentos. Imagens coloridas de documentos ainda oferecem restrições em aplicações práticas principalmente devido a volumes de documentos, espaço de armazenamento e velocidade de processamento. Imagens coloridas podem ser muito úteis e viáveis como por exemplo em aplicações médicas. No processamento de documentos, grande parte das soluções são obtidas com imagens binárias ou em níveis de cinza.

Importantes cuidados devem ser observados na aquisição de imagens de documento para processamento eletrônico. Os equipamentos de captura e as imagens geradas, devem estar de acordo com os padrões de qualidade estabelecidos. A DLF (Digital Library Federation) congrega um grupo de pesquisadores [AMATO, 2000] que tem como objetivo, estabelecer padrões para equipamentos de captura e imagens eletrônicas. A Cornell University Library elaborou um conjunto de requisitos para conversão digital [KENNEY, 1999] onde define as principais questões para digitalização. Todos os grandes e importantes projetos de armazenamento e tratamento de imagens eletrônicas, reservam uma especial atenção a fase de captura, uma vez que qualquer falha introduzida, trará dificuldades e reflexos negativos em todas as fases posteriores. Quanto melhor for a qualidade das imagens adquiridas, mais fáceis e rápidas serão as fases de processamento e de reconhecimento.

2.3.2 Captura de Documentos Simples

Documentos simples tem características que facilitam a captura e apresentam poucas diferenças quando o mesmo documento é capturado em preto e branco e em níveis de cinza, como pode-se observar nas figuras 2.3 e 2.4. Na figura 2.3 tem-se um documento simples capturado em 256 níveis de cinza e na figura 2.4, o mesmo documento capturado em preto e branco. Como pode-se observar, as diferenças entre as duas imagens são quase imperceptíveis.

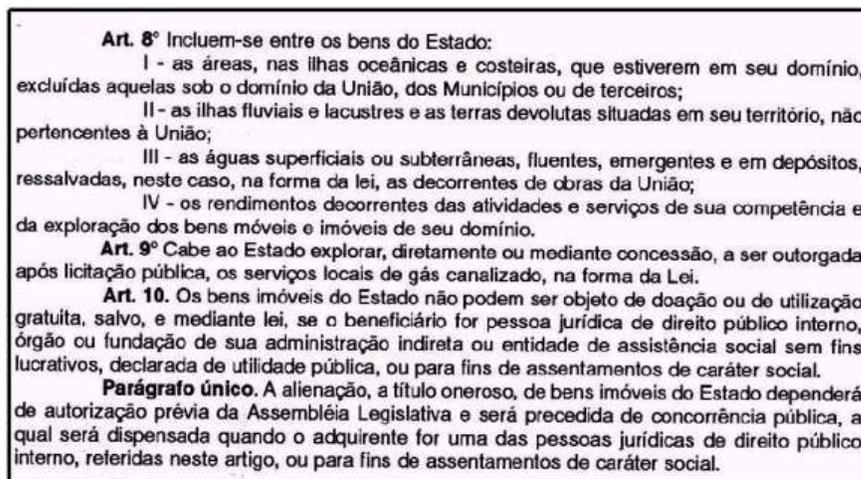


Figura 2.3 Documento simples capturado em 256 níveis de cinza

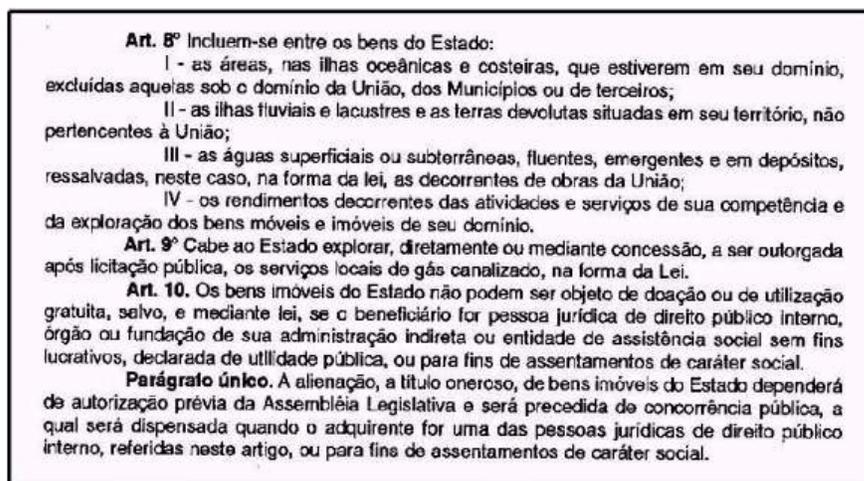


Figura 2.4 Documento simples capturado em preto e branco (limiarização efetuada pelo equipamento de captura)

2.3.3 Captura de Documentos Complexos

Na figura 2.5, tem-se um exemplo de documento complexo do ponto de vista de captura. O documento capturado em 256 níveis é mostrado na figura 2.6. O mesmo documento capturado em preto e branco (fig. 2.7) é muito diferente com significativa perda de conteúdo.

BANCO DO ESTADO DO PARANÁ S.A.
AVISO DE DÉBITO AO CLIENTE

Unidade Emissora: **BR/CURITIBA/GERENCO** Data: **31.01.94**

CURITIBA/COMENDADOR

CDF SEC A REG

138-4 003-5 **3545443 0** C-4 **267.714,54**

HISTÓRICO: Estamos contabilizando o valor acima, relativo a **3545443-0**

CONSIG JAN/94 AGE BOP 240 003 COD: 5 DR

CRÉD: ODOMED C/C 50386-4

Figura 2.5 Documento complexo em seu formato e características originais

BANCO DO ESTADO DO PARANÁ S.A.
AVISO DE DÉBITO AO CLIENTE

Unidade Emissora: **BR/CURITIBA/GERENCO** Data: **31.01.94**

CURITIBA/COMENDADOR

CDF SEC A REG

138-4 003-5 **3545443 0** C-4 **267.714,54**

HISTÓRICO: Estamos contabilizando o valor acima, relativo a **3545443-0**

CONSIG JAN/94 AGE BOP 240 003 COD: 5 DR

CRÉD: ODOMED C/C 50386-4

Figura 2.6 Documento complexo capturado em tonalidades de 256 níveis de cinza

BR/CURITIBA/GERENCO **31.01.94**

CURITIBA/COMENDADOR

CDF SEC A REG

138-4 003-5 3545443 0 267.714,54

CONSIG JAN/94 AGE BOP 240 003 COD: 5 DR

CRÉD: ODOMED C/C 50386-4

Figura 2.7 Documento complexo capturado em preto e branco (limiarizado pelo equipamento de captura)

2.3.4 Considerações

Como se pôde observar, alguns cuidados devem ser considerados na captura de documentos, principalmente se estes forem documentos complexos. Documentos simples na maioria das vezes podem ser capturados em preto e branco, uma vez que, um documento com único e bom contraste não apresenta diferenças entre uma captura em níveis de cinza e em preto e branco. Para documentos complexos, a captura em preto e branco, na maioria dos casos é inviável, como por exemplo o documento da figura 2.5, cujo resultado da captura observa-se na figura 2.7. Neste caso, captura-se o documento em 256 níveis de cinza (fig. 2.6) onde os detalhes e contrastes são preservados e a conversão para preto e branco pode ser realizada utilizando-se técnicas adequadas.

Para efeito desta pesquisa, o documento considerado é o da figura 2.6, a partir do qual se efetuará a rotulação e aplicação de técnicas de processamento de imagens com o objetivo de melhorar a qualidade e o reconhecimento dos conteúdos impressos.

2.4 Rotulação

A rotulação utilizada neste trabalho, consiste em identificar nas imagens de documentos complexos, regiões com características particulares onde serão aplicadas técnicas de processamento de imagens apropriadas para cada região.

O documento complexo (fig. 2.6), objeto desta pesquisa, possui diversas regiões com características particulares, sendo selecionadas para efeito deste estudo, 5 regiões abaixo identificadas:

- Região com baixo contraste (fig. 2.8) denominada neste trabalho como AFE, onde caracteres alfabéticos degradados estão impressos em fundo escuro.
- Região com alto contraste (fig. 2.9) denominada neste trabalho como AFC, onde caracteres alfabéticos degradados estão impressos em fundo claro.

- Região com baixo contraste (fig. 2.10) denominada neste trabalho como NFE, onde caracteres numéricos degradados estão impressos em fundo escuro.
- Região com alto contraste (fig. 2.11) denominada neste trabalho como NFC, onde caracteres numéricos degradados estão impressos em fundo claro.
- Região com baixo contraste (fig. 2.12) denominada neste trabalho como PFC, onde caracteres alfabéticos não degradados estão pré-impressos em fundo claro com ruídos.



Figura 2.8 Região com conteúdo de caracteres alfabéticos impressos em fundo escuro, denominada como: **AFE**.



Figura 2.9 Região com conteúdo de caracteres alfabéticos impressos em fundo claro, denominada como: **AFC**.

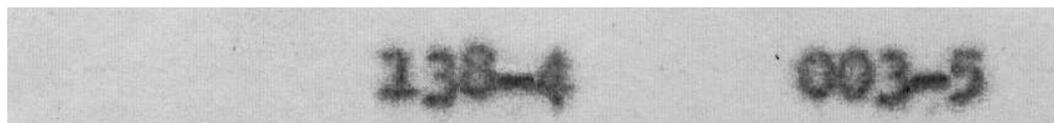


Figura 2.10 Região com conteúdo de caracteres numéricos impressos em fundo escuro, denominada como: **NFE**.



Figura 2.11 Região com conteúdo de caracteres numéricos impressos em fundo claro, denominada como: **NFC**.



Figura 2.12 Região com conteúdo de caracteres alfabéticos pré-impressos em fundo claro, denominada como: **PFC**.

2.5 Base de Dados

A base de dados para efeito de teste e validação da metodologia proposta nesta pesquisa é representada por um único tipo de documento complexo estruturado conforme mostra a figura 2.6, composta pelos seus respectivos rótulos conforme figuras 2.8 a 2.12.

Foram capturados 402 documentos sendo 134 escolhidos de forma aleatória para a base de testes e os restantes 268 considerados para a base de validação. Os documentos das bases foram rotulados, e as quantidades por rótulo estão detalhadas na tabela 2.1.

Rótulo	Teste (1/3)	Validação (2/3)	Total (3/3)
AFE	134	268	402
AFC	134	268	402
NFE	134	268	402
NFC	134	268	402
PFC	134	268	402
TOTAL GERAL	670	1.340	2.010

Tabela 2.1 Quantidades de documentos da base de dados para teste e validação por rótulo (região)

Capítulo 3

Estado da Arte

3.1 Introdução

O processamento de imagens de documentos apresenta como principal desafio o tratamento dos conteúdos escritos, uma vez que estes são os principais componentes de um documento. Um documento pode ter figuras, gráficos e outros componentes que não sejam caracteres, mas segundo a definição dos principais dicionários [AURÉLIO, 1991] [WEBSTER, 1998] [HOUAISS, 2001], o que de fato caracteriza um documento são os seus conteúdos escritos ou impressos. Assim sendo, um documento em formato analógico como o papel ou microfilme quando convertido para o formato digital, deve ter os seus conteúdos preservados com a melhor qualidade possível para garantir o reconhecimento tanto visual como automático.

Existem muitas técnicas formuladas para processamento de imagens eletrônicas mas a maior parte delas são utilizadas para processamento de imagens médicas, aéreas e industriais. Katz [KATZ, 2000] propõe o uso de técnicas de limiarização e erosão em imagens binárias para segmentar e classificar células de sangue, Pearlestine [PEARLESTINE, 2000] utiliza técnicas de limiarização para segmentar e identificar plantas invasoras em imagens aéreas da vegetação da costa leste da Flórida.

Segundo Lawrence O’Gorman [GORMAN, 1998] a maioria das técnicas formuladas para processamento de imagens de documentos, foi testada para documentos simples mas não em documentos com características particulares como os documentos

complexos. Os poucos estudos comparativos para processamento de imagens de documentos complexos como cheques bancários [MATTANA, 1999] não levam em consideração a necessidade de tratamento de grandes volumes, o que exige processamento rápido e na prática pode inviabilizar a sua utilização.

3.2 Abordagens para Documentos Complexos

A maior parte das abordagens propostas para melhorar o reconhecimento de conteúdos impressos de documentos, utiliza técnicas de processamento apenas para localização de componentes em imagens, como por exemplo: segmentação de páginas [MAO, 2000]; rotulação e localização de componentes baseada em regiões [YANIKOGLU, 1998] [RANDRIAMASY, 1994]; análise, rotulação e localização de textos [GORMAN, 1998] [BAYER, 1997]; localização de blocos em documentos impressos [TSUJIMOTO, 1992]; localização de blocos em documentos manuscritos [XUE, 1999]; localização de nomes e endereços em formulários de impostos [SRIHARI, 1996]; localização e separação de textos e não-textos [LE, 1995]; análise de documentos impressos em alto nível e baixo nível [DENGEL, 1990]. Estas abordagens não consideram o processamento dos conteúdos localizados com o objetivo de melhorar a qualidade e por consequência o reconhecimento. Porém encontram-se na literatura algumas propostas que têm esse objetivo como por exemplo:

- O projeto DEBORA [DEBORA, 2000] tem como finalidade, definir um conjunto de ferramentas que possibilitem de forma remota e colaborativa o acesso a conteúdos de livros antigos, de forma que não seja necessário o contato direto com os originais. Tem como grande preocupação a qualidade dos conteúdos impressos. Envolve um grande número de pesquisadores europeus com equipes da França, Portugal e Reino Unido. A base de dados é composta por aproximadamente 120 livros do século 16 das bibliotecas de Lyon na França, Roma na Itália e Coimbra em Portugal. Na etapa de recuperação de imagens, o projeto prevê a eliminação de degradação presente tanto no próprio documento como também aquela introduzida por equipamentos de captura. A captura é realizada em 256 níveis de cinza com no mínimo 300DPI, e filtros são utilizados para recuperar imagens degradadas. A figura 3.1(A) mostra

uma imagem original degradada e a figura 3.1(B) a mesma imagem recuperada por processo de filtragem. A figura 3.2(A) mostra uma imagem binária que não sofreu processo de filtragem antes da limiarização e a figura 3.2(B) a mesma imagem recuperada. Como os escritos são antigos e rebuscados, a avaliação por reconhecimento via OCR não pôde ser utilizada. Neste caso, o projeto prevê a segmentação, extração e classificação de caracteres aplicando técnicas para reconhecimento próprias.

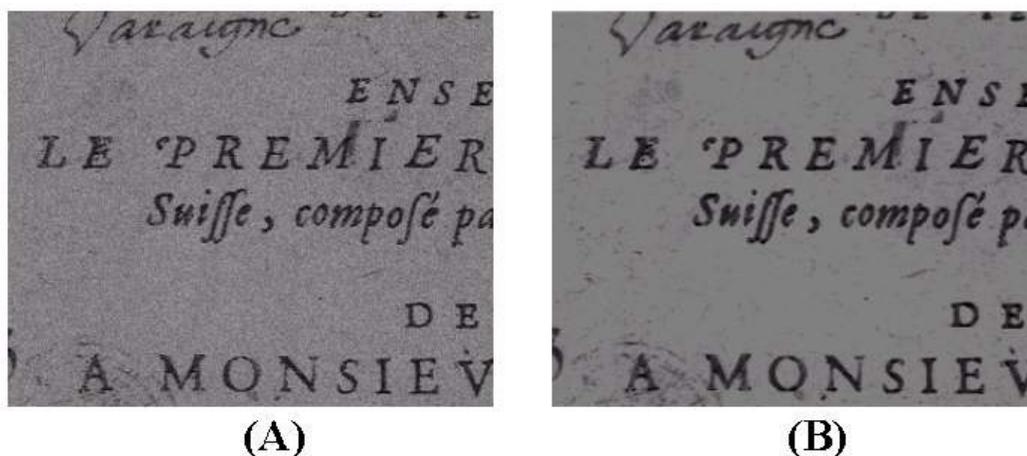


Figura 3.1 Resultado do processo de filtragem em níveis de cinza; (A) imagem original, (B) imagem recuperada.

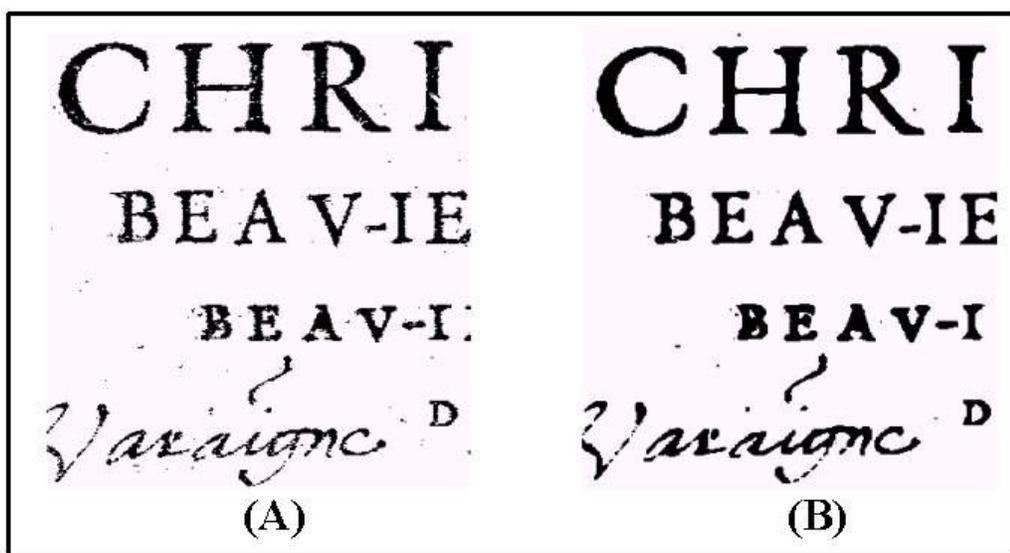


Figura 3.2 Resultado do processo de filtragem em imagens binárias; (A) imagem sem filtragem antes da limiarização, (B) imagem recuperada com filtro antes da limiarização.

- Cannon [CANNON, 1999] propôs um método para estimar a degradação em imagens de documentos impressos com caracteres tipográficos e selecionar automaticamente o melhor método de melhoria da qualidade de acordo com a degradação estimada. Foram utilizados 139 documentos do departamento de energia dos Estados Unidos, capturados em preto e branco com resolução de 300DPI. Os documentos foram separados em duas bases: 98 documentos com taxas médias de rejeição por OCR e 41 documentos com taxas médias de rejeição de 31,28%. Foram formuladas 5 medidas de degradação e uma base de treinamento com redes neurais foi utilizada para separar documentos. Técnicas de fechamento morfológico e eliminação de ruído com filtros foram utilizadas de acordo com a classificação dos documentos. A avaliação dos resultados foi realizada pela taxa de reconhecimento por OCR, comparando-se o índice de rejeição sem processamento com o índice de rejeição com o processamento dos conteúdos. A metodologia apresentou uma taxa de redução na rejeição do OCR em torno de 38%.
- Chang [CHANG, 2000] propôs uma técnica de limiarização híbrida para melhorar a taxa de reconhecimento por OCR de caracteres Chineses impressos. Foi utilizada uma base com 60 artigos, sendo 20 recentes, 20 de cinco anos e 20 de trinta anos. Cada artigo com aproximadamente 500 caracteres foi digitalizado em 256 níveis de cinza. A técnica consiste em aplicar a limiarização global nos caracteres maiores e uma limiarização baseada em janelas (3x3,5x5,9x9,etc.) em caracteres menores, uma vez que é muito comum na escrita chinesa a mistura de caracteres pequenos com grandes. Uma limitação é o uso de heurística na escolha da janela, uma vez que o seu tamanho depende do tamanho dos caracteres menores. A avaliação dos resultados foi feita comparando-se as taxas de reconhecimento do OCR da limiarização global com as taxas da limiarização híbrida. Nos artigos recentes o incremento foi de 0% a 10%, nos artigos de cinco anos de 5% a 23% e nos artigos de trinta anos de 20% a 37%.

- Mattana, [MATTANA, 1999] propôs um estudo sobre métodos de segmentação por binarização para cheques bancários brasileiros, visando à análise da qualidade final das imagens obtidas. Vários algoritmos de limiarização globais e locais adaptativos foram aplicados, e a qualidade dos resultados foi avaliada por reconhecimento via OCR das imagens limiarizadas. Nos métodos globais, os que apresentaram melhores resultados foram: Otsu com uma taxa de 86.12% para cheques não preenchidos e 66.1% para cheques preenchidos a máquina; o método fuzzy da abordagem de Yager com uma taxa de 86.60% e 59.55%, respectivamente; já nos métodos locais adaptativos, o método de Bernsen forneceu uma taxa de 92.79% e 84.03%, respectivamente.
- Hobby [HOBBY, 1996] apresentou um algoritmo para recuperação de caracteres a partir de imagens degradadas de documentos impressos. Utilizou um conjunto de 217 imagens binárias sobre as quais executou um algoritmo de 5 passos: 1) suaviza a imagem original “A” gerando uma imagem “B”; 2) reforça a imagem “B” adicionando níveis de cinza; 3) sobrepõe a imagem “A” com “B” e executa um processo de limiarização gerando uma imagem “C”; 4) suaviza a imagem “C”; 5) reforça os contornos da imagem “C”, gerando uma imagem com os caracteres degradados recuperados. A avaliação da qualidade foi por reconhecimento via OCR, analisando-se a redução de erro entre a imagem original e a imagem processada. Na base de dados utilizada, a taxa média de redução de erro foi de 20%.
- Kanungo [KANUNGO, 1999] propôs uma metodologia para comparar o reconhecimento de OCR em imagens degradadas das páginas da Bíblia. Utilizou como base de dados páginas do Novo Testamento, onde 198 imagens binárias e 198 em níveis de cinza foram digitalizadas a 600DPI. Os textos das respectivas páginas foram localizados manualmente e o fechamento morfológico com elemento estruturante quadrado foi utilizado para melhorar o conteúdo. A qualidade foi avaliada por reconhecimento via OCR, onde listagens das imagens processadas e de caracteres reconhecidos foram apresentadas conforme mostra a figura 3.3.

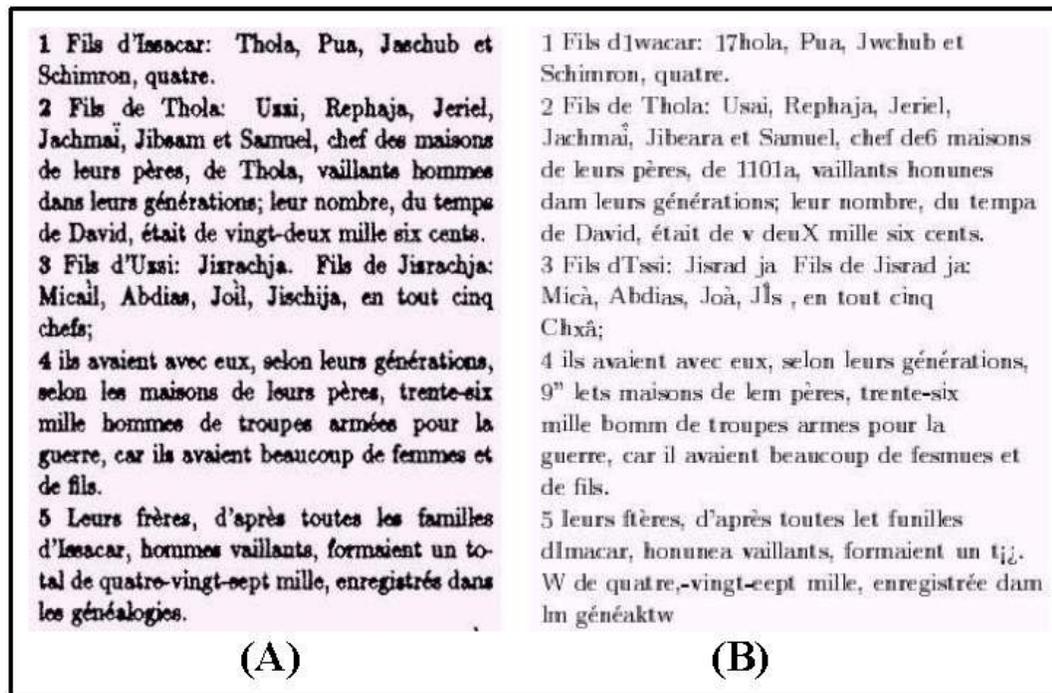


Figura 3.3 Processamento de página da Bíblia; (A) processada, (B) reconhecida por OCR

- Swain [SWAIN, 1997] propôs um estudo para recuperação, reconstrução e melhoramento de conteúdos manuscritos digitalizados. Utilizou uma biblioteca contendo livros manuscritos onde a recuperação é baseada em modelos de degradação para recuperar informações perdidas, a reconstrução é baseada na análise de contexto dos dados existentes na imagem e o melhoramento é baseado no uso de heurística para aprimorar o aspecto visual do conteúdo manuscrito. Em cada página do livro foram utilizados filtros não lineares para suavizar as imagens e o fechamento morfológico com elemento estruturante cruz foi utilizado para melhorar os conteúdos manuscritos. A avaliação foi feita de forma visual onde são comparadas as imagens antes do processamento com as respectivas imagens processadas conforme mostram as figuras 3.4 e 3.5. Segundo o autor, em muitos casos, a técnica não consegue reconstruir totalmente caracteres fragmentados, o que considera uma limitação do método.

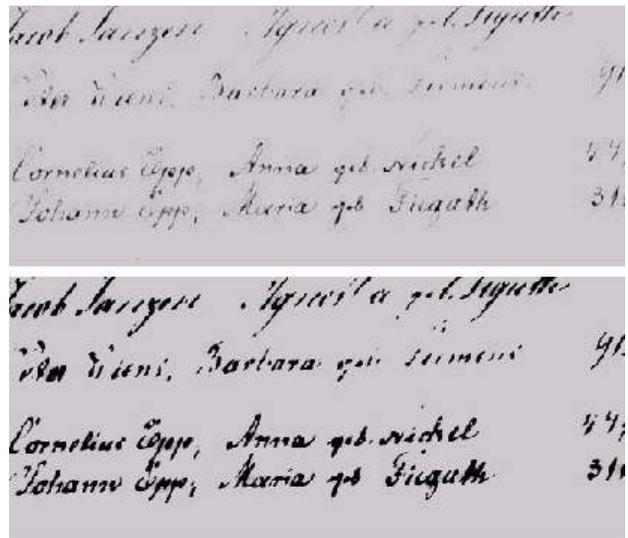


Figura 3.4 Resultado do melhoramento de conteúdo manuscrito; parte superior, imagem original; parte inferior, imagem melhorada.

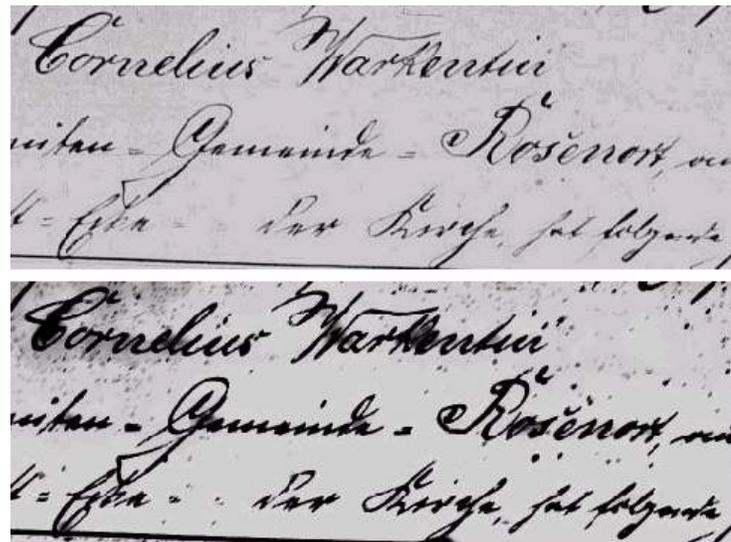


Figura 3.5 Resultado do melhoramento de conteúdo manuscrito; parte superior, imagem original; parte inferior, imagem melhorada.

3.3 Considerações

De acordo com o estudo bibliográfico apresentado, podem-se observar algumas peculiaridades nas abordagens analisadas:

- Um grande número de abordagens tem como propósito aprimorar o reconhecimento de conteúdos impressos pela eficiência da localização dos componentes, sem no entanto melhorar estes conteúdos.
- Muitas abordagens utilizam técnicas de processamento de imagens tão somente para facilitar a localização dos componentes.
- Das abordagens que utilizam a morfologia matemática binária, todas fazem uso apenas do operador de fechamento.
- Nenhuma das abordagens analisadas combina técnicas de limiarização com morfologia matemática binária para aumentar a qualidade dos conteúdos impressos.
- Nenhuma das abordagens estudadas, considera o tempo de processamento com o objetivo de viabilizar grandes volumes.

Capítulo 4

Limiarização

4.1 Introdução

A limiarização (*thresholding*) consiste em converter uma imagem com mais de 2 níveis de cinza para apenas 2 níveis 0 (preto) e 1 (branco). Uma imagem com n níveis de cinza conterá preto no nível 0 e branco no nível $n-1$ sendo os demais níveis (se existirem) tons intermediários entre o preto e o branco. Uma imagem de 256 níveis de cinza para ser limiarizada, deve ter definidas duas classes: uma classe de níveis que devem ser convertidos para preto (0) e outra classe de níveis que devem ser convertidos para branco (1). Em uma imagem com conteúdos impressos em tons mais escuros sobre um fundo mais claro, a classe de níveis convertidos para preto, fará parte do primeiro plano (objeto), enquanto que a classe de níveis convertidos para branco, fará parte do segundo plano (fundo) da imagem.

O valor que separa estas duas classes é chamado de limiar. Portanto a limiarização consiste em encontrar o valor ideal do limiar (aquele que melhor separa os componentes do objeto e fundo) e converter todos os pixels abaixo deste valor para preto e todos os pixels iguais ou acima deste valor para branco. A figura 4.1 mostra alguns efeitos da limiarização sobre um conteúdo impresso conforme o valor do limiar.

A limiarização tem como principais objetivos: reduzir a quantidade de dados a serem tratados em aplicações de processamento de imagens; reduzir o tamanho das imagens para diminuir custos de armazenamento em soluções de GED com grandes volumes; reduzir o tamanho das imagens para viabilizar o tráfego de rede; facilitar o reconhecimento de componentes existentes no documento e reduzir/eliminar ruídos que se caracterizam como dados estranhos a imagem.

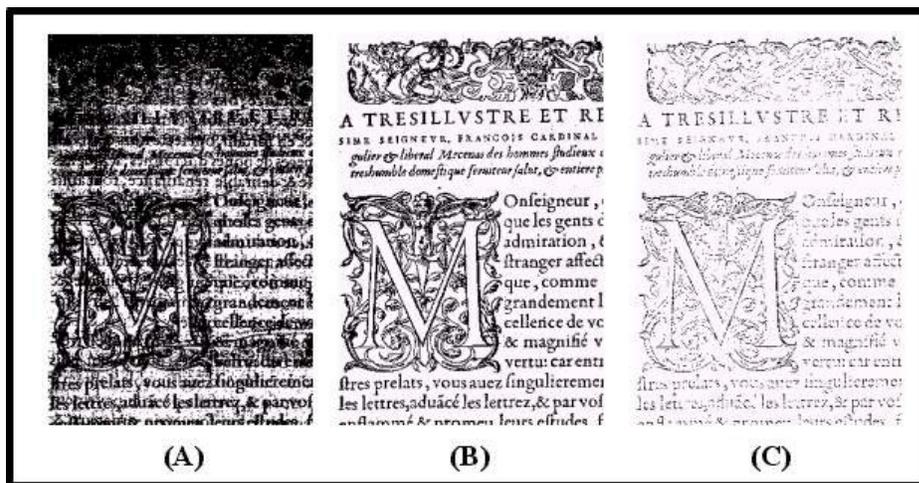


Figura 4.1 Efeitos da variação do limiar em conteúdo impresso: (A) limiar muito alto; (B) limiar ideal; (C) limiar muito baixo.

Para documentos simples ou com um bom contraste de componentes contra um fundo uniforme, os scanners binários estão disponíveis. Eles combinam a digitalização com a limiarização e geram imagens em preto e branco a partir de um limiar fixo. Entretanto, para documentos complexos que, em geral, possuem uma vasta faixa de fundo e intensidade de objetos, este nível de limiar fixo frequentemente não gera imagens com uma clara separação entre os componentes do objeto e os de fundo. Para estes casos, existem duas alternativas. Uma é determinar empiricamente a melhor configuração da limiarização no scanner (a maioria dos scanners binários possuem este ajuste), e fazer isto cada vez que a imagem for fraca “binariamente”. A outra alternativa é capturar as imagens em níveis de cinza a partir do estágio da digitalização e então usar métodos para obtenção automática do limiar que melhor execute a limiarização.

Enquanto a última alternativa requer mais dados de entrada e mais processamento, sua vantagem é que um bom nível de limiar pode ser encontrado automaticamente, assegurando boas e consistentes imagens e evitando a necessidade de consumo de tempo através do ajuste manual e digitalização repetida.

Se os valores dos pixels componentes do objeto e do fundo são razoavelmente consistentes em seus respectivos valores sobre a imagem inteira, então um único valor de limiar pode ser encontrado para esta imagem. Esta utilização de um único limiar para todos os pixels da imagem é chamada de **limiarização global**. Contudo, para muitos documentos complexos [GORMAN, 1998], um único valor de limiar global não pode

ser usado por causa das deformidades dentro das regiões de primeiro e segundo planos. Para este tipo de imagem, valores de limiar diferentes são requeridos para regiões/locais diferentes. Este tipo é chamado de **limiarização local adaptativa**.

4.2 Limiarização Global

O modo mais direto para selecionar automaticamente um valor global de limiar é através do uso de um histograma das intensidades dos pixels com valores de cada nível de intensidade.

Para uma imagem com intensidades de primeiro e segundo planos bem diferenciados, como no caso de documentos simples, o histograma terá dois picos distintos (fig. 4.2). O vale entre estes picos pode ser encontrado como o mínimo entre dois máximos, e o valor da intensidade aí é escolhido como limiar que melhor separa os dois picos.

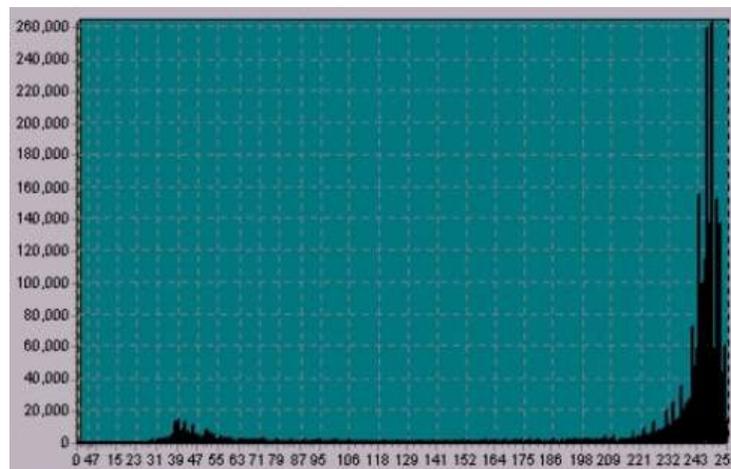


Figura 4.2 Histograma de imagem de documento com bom contraste entre fundo e objeto, típico de documento simples

Cabe destacar que, quando as imagens não contém intensidades de primeiro e segundo planos bem diferenciados ou contém componentes de primeiro plano esparsos, como a maioria das imagens de documentos complexos, o pico representando estas intensidades será muito menor do que o pico das intensidades de segundo plano (fundo) (fig. 4.3). Com estas informações, o histograma pode indicar que, nestes casos, a limiarização global não é apropriada.

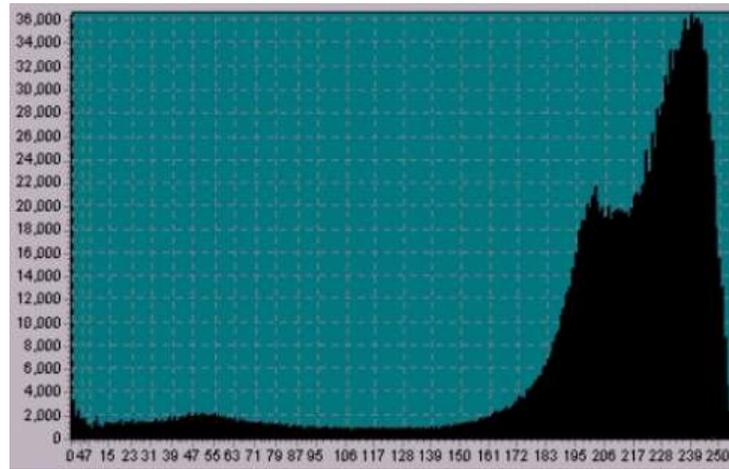


Figura 4.3 Histograma de imagem com componentes de objetos esparsos, típico de documento complexo

A seguir apresenta-se um conjunto de métodos de limiarização global, sendo que, muitos deles utilizam informações de histograma para determinar o valor do limiar.

4.3 Métodos de Limiarização Global

4.3.1 Método de Otsu

O método de Otsu [OTSU, 1979] baseia-se na análise discriminante. A operação de limiarização é obtida pelo particionamento dos pixels de uma imagem em duas classes:

$$C_0 = \{0, 1, 2, \dots, t\} \text{ e } C_1 = \{t+1, t+2, \dots, l-1\}, \text{ isto é, objeto e fundo.}$$

Otsu demonstrou que o valor ótimo do limiar t^* pode ser determinado pela maximização do critério de separabilidade da função $\eta = \sigma_B^2 / \sigma_T^2$, onde η é a variância entre classes dividida pela variância total: $t^* = \text{ArgMax } \sigma_B^2$,

$$\sigma_T^2 = \sum_{i=0}^{l-1} (i - \mu_T)^2 P_i \quad (4.1) \quad \mu_T = \sum_{i=0}^{l-1} i p_i \quad (4.2),$$

$$\sigma_B^2 = \omega_0 \omega_1 (\mu_1 - \mu_0)^2 \quad (4.3)$$

$$\omega_1 = 1 - \omega_0 \quad (4.4)$$

$$\omega_0 = \sum_{i=0}^t P_i \quad (4.5),$$

$$\mu_1 = \frac{\mu_T - \mu_i}{1 - \omega_0} \quad (4.6)$$

$$\mu_0 = \frac{\mu_T}{\omega_0} \quad (4.7)$$

$$\mu_i = \sum_{i=0}^t i p_i \quad (4.8),$$

sendo que ω_0 e ω_1 correspondem às variâncias nas classes C_0 e C_1 ; μ_0 e μ_1 correspondem às médias das classes, respectivamente. σ_B^2 , σ_T^2 as variâncias interclasses e total, respectivamente.

4.3.2 Método de Kittler e Illingworth

Com este método, Kittler e Illingworth [SAHOO, 1988] prevêm o histograma como uma estimativa da função densidade de probabilidade $p(g)$ da população, que compreende os níveis de cinza dos objetos e do fundo da seguinte maneira:

$$p(g) = \frac{1}{\sigma_1 \sqrt{2\pi}} * e^{-\left(\frac{(g-\mu_1)^2}{2\sigma_1^2}\right)} + \frac{1}{\sigma_2 \sqrt{2\pi}} * e^{-\left(\frac{(g-\mu_2)^2}{2\sigma_2^2}\right)} \quad (4.9)$$

Os parâmetros μ_i , σ_i^2 e as probabilidades a priori p_i com $i = 1, 2, \dots$ da densidade de probabilidade $p(g)$, associados a uma imagem a ser limiarizada, não são usualmente conhecidos. Kittler e Illingworth introduziram uma função critério $J(t)$:

$$J(t) = 1 + 2(p_1(t) \log(\sigma_1(t)) + p_2(t) \log(\sigma_2(t))) - 2(p_1(t) \log(p_1(t)) + p_2(t) \log(p_2(t))) \quad (4.10)$$

onde os parâmetros $\mu_i(t)$, $\sigma_i^2(t)$ e as probabilidades a priori $p_i(t)$ com $i = 1, 2, \dots$ são definidos como:

$$p_1(t) = \sum_{g=0}^t h(g) \quad (4.11) \quad \text{e} \quad p_2(t) = \sum_{g=t+1}^{255} h(g) \quad (4.12)$$

$$\mu_1(t) = \frac{\sum_{g=0}^t gh(g)}{p_1(t)} \quad (4.13) \quad \text{e} \quad \mu_2(t) = \frac{\sum_{g=t+1}^{255} gh(g)}{p_2(t)} \quad (4.14)$$

$$\sigma_1^2 = \frac{\sum_{g=0}^t h(g)(g - \mu_1(t))^2}{p_1(t)} \quad (4.15) \quad \text{e} \quad \sigma_2^2 = \frac{\sum_{g=t+1}^{255} h(g)(g - \mu_2(t))^2}{p_2(t)} \quad (4.16)$$

onde g é o nível de cinza da imagem e $h(g)$ o valor do histograma para cada nível de cinza. Assim, o limiar T é obtido a partir da minimização da função $J(t)$.

4.3.3 Método por Entropia de Pun

Considerando que a entropia é uma medida de conteúdo da informação com diversos possíveis símbolos, e que um determinado símbolo pode ocorrer de acordo com uma probabilidade, Pun [PUN, 1981] sugeriu um algoritmo que tenta achar o limiar t que maximiza a função $H = H_b + H_w$ onde a **Equação 4.17** é a entropia associada, para uma imagem com 256 níveis de cinza, com os pixels pretos limiarizados com um limiar t . De forma análoga, a **Equação 4.18** é a entropia dos pixels brancos. Pun demonstrou que basta maximizar a função $f(t)$ conforme **Equação 4.19**.

$$H_b(t) = -\sum_{i=0}^t p_i \log(p_i) \quad (4.17) \quad H_w(t) = -\sum_{i=t+1}^{255} p_i \log(p_i) \quad (4.18)$$

$$f(t) = \frac{H_t}{H_T} \frac{\log(P_t)}{\log(\text{Max}\{p_0, p_1, \dots, p_t\})} + \left[1 - \frac{H_t}{H_T}\right] \frac{\log(1 - P_t)}{\log(\text{Max}\{p_{t+1}, p_{t+2}, \dots, p_{255}\})} \quad (4.19)$$

com

$$H_t = -\sum_{i=0}^t p_i \log(p_i) \quad (4.20), \quad H_T = -\sum_{i=0}^{255} p_i \log(p_i) \quad (4.21) \quad \text{e} \quad P_t = \sum_{i=0}^t p_i \quad (4.22).$$

onde H_t , H_T e P_t representam, respectivamente, a entropia dos pixels pretos, a entropia total e a probabilidade cumulativa até o nível de cinza t .

4.3.4 Método por Entropia de Kapur

Kapur [KAPUR, 1985], utilizando como base o método de Pun[PUN, 1981], procura definir a probabilidade de distribuição de um objeto (**Equação 4.23**) contra a probabilidade de distribuição do fundo (**Equação 4.24**).

$$\frac{p_0}{P_t}, \frac{p_1}{P_t}, \dots, \frac{p_t}{P_t} \quad (4.23)$$

$$\frac{p_{t+1}}{(1-p_t)}, \frac{p_{t+2}}{(1-p_t)}, \dots, \frac{p_{255}}{(1-p_t)} \quad (4.24)$$

O valor de limiar ótimo é o valor de t que maximiza a expressão $H = H_b(T) + H_w(T)$, onde as entropias dos pixels pretos é obtida conforme **Equação 4.25** e dos pixels brancos é obtida conforme **Equação 4.26**.

$$H_b(t) = -\sum_{i=0}^t \frac{p_i}{P_t} \log\left(\frac{p_i}{P_t}\right) \quad (4.25) \quad H_w(t) = -\sum_{i=t+1}^{255} \frac{p_i}{1-P_t} \log\left(\frac{p_i}{1-P_t}\right) \quad (4.26)$$

4.3.5 Método por Entropia de Johannsen

Com base na divisão dos níveis de cinza da imagem em duas partes a fim de minimizar a interdependência entre elas, Johannsen [JOHANNSEN, 1982] sugere que para se determinar o limiar, devem ser calculados os valores de $S_b(t)$ e $S_u(t)$:

$$S_b(t) = \log\left(\sum_{i=0}^t p_i\right) - \frac{1}{\sum_{i=0}^t p_i} [p_t \log(p_t) + \left(\sum_{i=0}^{t-1} p_i\right) \log\left(\sum_{i=0}^{t-1} p_i\right)] \quad (4.27)$$

$$S_w(t) = \log\left(\sum_{i=t}^{255} p_i\right) - \frac{1}{\sum_{i=t}^{255} p_i} [p_t \log(p_t) + \left(\sum_{i=t+1}^{255} p_i\right) \log\left(\sum_{i=t+1}^{255} p_i\right)] \quad (4.28)$$

O valor do limiar ótimo é o valor de t que minimiza a expressão $S_b(t) + S_u(t)$. É importante observar que os valores calculados para $S_b(t)$ e $S_u(t)$ devem ser desprezados para valores de t quando $p(t) = 0$.

4.3.6 Método por Entropia de Li-Lee

Li-Lee [LI, 1993] propõem a entropia mínima cruzada para medir a distância teórica da informação entre duas distribuições:

$P = \{p_1, p_2, \dots, p_n\}$ e $Q = \{q_1, q_2, \dots, q_n\}$, ($0 \leq n \leq N$) por:

$$D(Q, P) = - \sum_{k=0}^N q_k \log_2 \left(\frac{q_k}{p_k} \right) \quad (4.29)$$

O método da entropia mínima cruzada pode ser visto como uma extensão do método da entropia máxima pela atribuição de estimativas iniciais para todas as probabilidades quando nenhuma informação anterior está disponível.

Trabalhos aplicando o método de entropia máxima consideram o processo de geração de pixels como tentativas independentes. Eles usam o histograma normalizado de níveis de cinza como as distribuições de probabilidade de níveis de cinza baseados em tentativas aleatórias de pixels individuais que têm um certo nível de cinza e medem a entropia da distribuição dos pixels.

No esquema proposto por Li-Lee [LI, 1993] a imagem limiarizada $g(x, y)$ será construída a partir de $f(x, y)$ da seguinte forma:

$$g(x, y) = \begin{cases} \mu_1 & \text{se } f(x, y) < t \\ \mu_2 & \text{se } f(x, y) \geq t \end{cases} \quad (4.30)$$

A imagem limiarizada $g(x, y)$ é unicamente determinada a partir de $f(x, y)$ pela especificação de três parâmetros desconhecidos: t , μ_1 , e μ_2 . Uma função critério n deve ser construída para possibilitar o encontro do $g(x, y)$ ótimo ou equivalentemente ao conjunto ótimo de parâmetros t , μ_1 , e μ_2 que assemelha-se a $f(x, y)$ o máximo possível. Ou seja: $n[g(x, y)] \equiv n(t, \mu_1, \mu_2)$.

A função critério n é geralmente algum tipo de medida de distorção. Neste método é utilizada a entropia cruzada, que é a medida preferida para distribuições positivas e aditivas. Li-Lee definiram a função critério n da seguinte maneira:

$$\eta(t) = \sum_{g=1}^{t-1} gh(g) \log\left(\frac{g}{\mu_1(t)}\right) + \sum_{g=t}^{255} gh(g) \log\left(\frac{g}{\mu_2(t)}\right) \quad (4.31)$$

onde μ_1 , e μ_2 são assim definidos:

$$\mu_1(t) = \frac{\sum_{g=0}^{t-1} gh(g)}{\sum_{g=0}^{t-1} h(g)} \quad e \quad \mu_2(t) = \frac{\sum_{g=t}^{255} gh(g)}{\sum_{g=t}^{255} h(g)} \quad (4.32) \quad (4.33)$$

com g o nível de cinza da imagem e $h(g)$ o valor do histograma para cada nível de cinza.

O valor do limiar ótimo é então selecionado para o valor de t que minimiza a função critério $n(t)$.

4.3.7 Método por Entropia de WuLu

No método de limiarização proposto por Kapur, Sahoo e Wong [KAPUR, 1985], a limiarização ótima, tem seu valor representado pelo nível de cinza que maximiza a expressão $H = H_b(T) + H_w(T)$, onde as entropias dos pixels pretos é obtida conforme **Equação 4.25** e dos pixels brancos é obtida conforme **Equação 4.26**.

Porém, o que o método de WuLu [WU, 1998] propõe está baseado em estudo mais aprofundado de histogramas comuns a imagens dos aparelhos de ultra-som. Este tipo de aparelho gera uma imagem com pouco contraste onde, normalmente, a informação encontra-se misturada ao fundo e com ruídos inerentes à imagem. Para a resolução de tal problema, a técnica de WuLu objetiva encontrar o valor de nível de cinza que minimize (em vez de maximizar) a diferença entre as entropias do objeto e do seu fundo.

4.3.8 Método por Abordagem Fuzzy de Huang

Os métodos por abordagem fuzzy baseiam-se na lógica fuzzy onde é possível se determinar a probabilidade de um pixel de pertencer a uma das duas classes; primeiro plano (objeto) ou segundo plano (fundo). Inicialmente deve-se calcular a função de "classificação" que representa a probabilidade de um pixel pertencer a cada uma das classes. Para isso assume-se que μ_0 corresponde à média dos níveis de cinza do fundo e μ_1 à média dos níveis de cinza do objeto. O pixel pertencerá à classe da qual seu nível mais se aproximar. Quanto menor a diferença entre o nível de cinza g de um pixel e o valor da média da classe a que pertence, maior será a probabilidade $\mu_g(g)$ de classificação.

A função de classificação é obtida pela **Equação 4.34**.

$$\mu_g(g) = \begin{cases} \frac{1}{1+|g-\mu_0|/C} & \text{se } g < t \\ \frac{1}{1+|g-\mu_1|/C} & \text{se } g \geq t \end{cases} \quad (4.34)$$

Um valor estimado para ambos μ_0 e μ_1 é necessário. Para o limiar t dado, obtém-se, para μ_0 (fundo) e μ_1 (objeto):

$$\mu_0(t) = \frac{\sum_{g=0}^t gh(g)}{\sum_{g=0}^t h(g)} \quad (4.35)$$

$$\mu_1(t) = \frac{\sum_{g=t+1}^{255} gh(g)}{\sum_{g=t+1}^{255} h(g)} \quad (4.36)$$

Ambos valores dependem do limiar. O valor de C é calculado como uma constante: a diferença entre o máximo e o mínimo nível de cinza na imagem.

Cada pixel, cujo nível de cinza é g , fará parte do fundo ou da classe de objetos que compõem os objetos dependendo da sua relação entre seu nível de cinza do pixel e o limiar t . Para um pixel de objeto (onde $g > t$), a sua probabilidade de pertencer ao objeto é dada por $\mu_g(g)$ e deve estar situada entre os valores 1/2 e 1.

Os métodos por abordagem fuzzy fazem uso do conceito de "fuzziness" que corresponde a distância entre a imagem original em níveis de cinza e a imagem limiarizada correspondente [HUANG, 1995]. O fuzziness tem como valor mínimo 0 e

como valor máximo 1. Esses métodos especificam que quanto menor o fuzziness mais preciso é o valor do limiar.

Pelo Método de Huang [HUANG, 1995], calculada a função de classificação, um primeiro meio de medir a fuzziness é se basear na entropia que é calculada usando a função de Shannon.

$$Hf(x) = -x \log(x) - (1-x) \log(1-x) \quad (4.37)$$

O valor $E(t)$ da entropia do conjunto fuzzy que constitui a imagem é portanto:

$$E(t) = \frac{1}{nm} \sum_g Hf(\mu_g(g))h(g) \quad (4.38)$$

onde g representa o nível de cinza do pixel; $h(g)$ identifica o nível do histograma para determinado nível da escala de cinza; n representa o número de linhas da imagem e m representa o número de colunas da imagem.

Esta função é baseada em t porque μ_g , também baseia-se em t . O valor de limiar t apropriado é obtido por minimização da fuzziness, na prática obtida pela minimização da função $E(t)$.

4.3.9 Método por Abordagem Fuzzy de Yager

Outra forma para medir o fuzziness é proposta por Yager [YAGER, 1979] e baseia-se na idéia de que para um plano A onde não existem elementos em comum entre A e seu complemento A^c . O nível em que A e A^c são indistintos é a medida de fuzziness em A pode ser calculada pela equação:

$$D_p(t) = \left[\sum_g |\mu_g(g) - \mu_{\bar{g}}(g)|^p \right]^{1/p} \quad (4.39)$$

O valor de p é um inteiro usado para definir a distância medida; sendo $p = 2$, igual a distância Euclidiana.

Novamente o valor de limiar t apropriado é obtido por minimização da fuzziness, na prática obtida pela minimização da função $D_p(t)$.

4.3.10 Método por Abordagem Fuzzy C Means

O método de Limiarização Fuzzy C Means [JAWAHAR, 1997] propõe encontrar valores médios de fundo e objeto e através deles obter o valor limiar.

O método baseia-se na construção de dois valores iniciais para cada nível de cinza μ_o e μ_b , que são a probabilidade de se encontrar um determinado nível de cinza no objeto ou no fundo respectivamente:

$$\mu_o(t) = \frac{\sum_{j=0}^t j.h(j)}{\sum_{j=0}^t h(j)}, \quad (4.40) \quad \mu_B(t) = \frac{\sum_{j=t+1}^{255} j.h(j)}{\sum_{j=t+1}^{255} h(j)} \quad (4.41)$$

onde: j é o nível de cinza; $h(j)$ o número de níveis de cinza que o tom j possui na imagem; μ_o e μ_B são as probabilidades de o tom de cinza pertencer ao objeto e ao fundo respectivamente.

Como os somatórios tanto de μ_o como de μ_B não são valores contidos entre 0 e 1, uma manipulação algébrica é efetuada, para que essa condição seja satisfeita:

$$\bar{\mu}_o(t) = \frac{\mu_o(t)}{\mu_o(t) + \mu_B(t)} \quad (4.42) \quad \bar{\mu}_B(t) = \frac{\mu_B(t)}{\mu_o(t) + \mu_B(t)} \quad (4.43)$$

Após a inicialização de μ_o e μ_B , são calculadas duas equações abaixo descritas, sendo a primeira para obtenção do valor médio do nível de cinza do objeto e do fundo, e a segunda para atualizar os valores μ_o e μ_B .

$$v_i = \frac{\sum_{j=0}^{255} j.h(j).\mu_i(j)^\tau}{\sum_{j=0}^{255} h(j).\mu_i(j)^\tau}, \quad i = 1,2 \quad (4.44)$$

$$\bar{\mu}_o(j) = \frac{1}{1 + [d(j, v_o) / d(j, v_B)]^{2/(\tau-1)}}, \quad (4.45)$$

$$\bar{\mu}_B(j) = 1 - \bar{\mu}_o(j) \quad (4.46)$$

Onde: v_i é o nível médio do tom de cinza do objeto ou fundo; τ é o fuzziness e deve ser maior que 1; $d(j, v_i)$ é a distância absoluta entre o nível de cinza e o nível médio correspondente.

O método deve executar as equações até que não se encontre diferenças significantes em μ_O e μ_B , quando então, calcula-se o valor de Tf que é o valor do limiar Fuzzy. Tf é o valor médio entre v_B e v_O , sendo obtido pela equação:

$$Tf = \frac{v_O + v_B}{2} \quad (4.47)$$

4.3.11 Método por Seleção Iterativa de Thrussel

Pelo método de Thrussel [THRUSSEL, 1979] ou seleção iterativa, um processo é executado, no qual um valor inicial arbitrário de limiar é refinado por consecutivas varreduras sobre a imagem. A idéia do processo é não utilizar o histograma da imagem, mas separar as classes objeto e fundo repetidamente, utilizando os níveis de cada classe para aprimorar o limiar. O valor inicial é calculado a partir da coleta de estatísticas sobre regiões claras e escuras. A médias dos níveis de cinza para os pontos abaixo do limiar (T_b) e para os pontos acima do limiar (T_0) são encontradas. Com estes dois parâmetros calculados, o novo limiar T_k é calculado pela seguinte equação:

$$T_k = (T_b + T_0)/2 \quad (4.48)$$

Tendo-se este novo limiar, o processo repete-se até que não haja alteração de T_k , entre um passo e outro. O limiar T_k é, então, obtido pela seguinte equação:

$$T_k = \frac{\sum_{g=0}^{T_{k-1}} gh(g)}{2 \sum_{g=0}^{T_{k-1}} h(g)} + \frac{\sum_{g=T_{k-1}+1}^{255} gh(g)}{\sum_{g=T_{k-1}+1}^{255} h(g)} \quad (4.49)$$

onde $h(g)$ é o vetor histograma de níveis de cinza da imagem.

Neste caso, o processo de determinação do limiar termina e converge quando: $T_k = T_{k+1}$, sendo T_k o valor ótimo.

4.3.12 Método por Seleção Iterativa de LLoyd

O Método de LLoyd [LLOYD, 1988] é baseado no método de seleção iterativa de Ridler e Calvard [RIDLER, 1978], onde o valor do limiar t é calculado pela equação:

$$t = \frac{1}{2}(\bar{z}_1 + \bar{z}_2) \quad (4.50)$$

Porém pelo método proposto por LLoyd a equação é alterada para:

$$t = \frac{1}{2}(\bar{z}_1 + \bar{z}_2) + \frac{\sigma^2}{(\bar{z}_1 - \bar{z}_2)} \log(n_2/n_1) \quad (4.51)$$

Onde \bar{z}_1 e \bar{z}_2 são as médias dos níveis de cinza; n_1 e n_2 o número de pixels; e σ a variância.

A alteração proposta por LLoyd faz o método iterativo de limiarização convergir para um valor de limiar mais satisfatório. Mesmo com esta alteração, o método de LLoyd, é um dos métodos mais rápidos em termos de processamento.

4.4 Limiarização Local Adaptativa

A grande vantagem da limiarização local adaptativa, é a conversão de imagens em níveis de cinza para preto e branco quando mais de um valor de limiar por imagem é necessário, como na maioria dos documentos complexos.

Uma forma de executar a limiarização local adaptativa é analisando as intensidades de níveis de cinza dentro de uma janela local sobre a imagem para determinar limiares locais.

O problema principal com qualquer técnica de limiarização adaptativa é a escolha do tamanho da janela. O tamanho escolhido deve ser grande o suficiente para garantir que o número de pixels de segundo plano (fundo) incluídos sejam também grandes o suficiente para obter uma boa estimativa do valor da média, mas não tão grande como a média sobre as intensidades de segundo plano não uniformes. Porém, as características na imagem frequentemente variam em tamanho, de forma que existem

problemas com o tamanho de janela fixo. Para remediar este problema, a informação dependente de domínio pode ser usada para verificar se os resultados da limiarização geram as características esperadas. Se o resultado é inesperado, então o tamanho da janela pode ser modificado e a limiarização é aplicada outra vez. A seguir apresenta-se um conjunto de métodos de limiarização local adaptativa.

4.5 Métodos de Limiarização Local Adaptativa

4.5.1 Método de Bernsen

Bernsen [BERSEN, 1986] propôs um método local ou adaptativo onde para cada pixel (x, y) , o limiar $T(x, y) = Z_{\min} + Z_{\max} / 2$ é utilizado, onde Z_{\min} e Z_{\max} são os valores dos níveis máximos e mínimos da escala de cinza em uma vizinhança quadrada de dimensões $r \times r$ centralizada em (x, y) .

4.5.2 Método de Niblack

A idéia do método de limiarização local adaptativa de Niblack [NIBLACK, 1986], é variar o limiar sobre a imagem, baseado em informações da região e na divergência padrão do local. O limiar T do ponto (x, y) é assim calculado:

$$T(x, y) = 'peso' * d(x, y) + v(x, y) \quad (4.52)$$

onde $d(x, y)$ é a divergência padrão calculada e $v(x, y)$ é o valor da vizinhança local de (x, y) de acordo com o tamanho da região $(n \times n)$. O valor *peso* é usado como ajuste dependendo o quanto das bordas do objeto de primeiro plano são consideradas como parte do objeto. Um valor recomendado para *peso* é -0.2 para imagens com objetos de primeiro plano pretos, e *peso* = +0.2 para imagens com objetos de primeiro plano brancos.

4.5.3 Método de Eikvil

Pelo método de Eikvil [EIKVIL, 1991], os pixels existentes em uma janela pequena P de tamanho $(n \times n)$ são limiarizados com base nas classes dos pixels de uma janela concêntrica maior G de tamanho $(m \times m)$. P e G deslizam sobre a imagem de acordo com o tamanho de P . Para todos os pixels de G , um limiar T é calculado com divisão dos pixels em duas classes $cl1$ e $cl2$. Se as duas classes calculadas são separadamente adicionais a um limite, $cl2 - cl1 > dif_limite$, os pixels de P são limiarizados pelo limiar T . Quando $cl2 - cl1 < dif_limite$, os pixels de P são todos designados à classe com o valor mais próximo atualizado, isto é, objeto ou fundo.

O método de limiarização global Otsu é usado para encontrar o valor de T . Um *peso* é usado para calcular valores de atualização $CL1$ e $CL2$ para cada classe. Para cada janela P , valores novos são calculados da seguinte forma:

$$CL1 = 'peso' * CL1 + (1.0 - 'peso') * cl1 \quad (4.53)$$

$$CL2 = 'peso' * CL2 + (1.0 - 'peso') * cl2 \quad (4.54)$$

peso = 1.0 é sempre usado para valores iniciais de $CL1$ e $CL2$, enquanto que *peso* = 0.0 é usado para valores $cl1$ e $cl2$ das duas classes de G e são usadas no passo seguinte.

4.6 Escolha de um Método de Limiarização

Diferentes métodos de limiarização foram aqui apresentados. Segundo O'Gorman [GORMAN, 1998], nenhum método é melhor para todos os tipos de imagens e aplicações. Para problemas mais simples, onde as características da imagem não variam muito, os métodos mais simples serão suficientes. Para problemas mais difíceis, como no caso de documentos complexos, métodos mais sofisticados geralmente são requeridos. Os produtos comerciais variam em suas capacidades de limiarização. Os scanners atuais executam a limiarização com um valor de limiar fixo. Sistemas de documentos mais sofisticados provêm técnicas manuais ou automáticas baseadas no popular método de Otsu para a limiarização global. O uso mais comum de limiarização adaptativa está em sistemas com objetivos específicos. De acordo com O'Gorman [GORMAN, 1998], a melhor maneira de escolher um método é primeiro

estreitar as escolhas através das descrições dos métodos e então ir experimentando-os com os outros e examinando seus resultados.

Um problema que requer mais trabalho é identificar métodos de limiarização ou enfoques que melhor trabalham com documentos com características particulares. Muitos dos métodos anteriormente descritos não foram formulados para documentos em particular, e seus desempenhos não são bem conhecidos. Para imagens de documentos com conteúdos impressos O’Gorman [GORMAN, 1998] sugere executar o reconhecimento óptico de caracteres – OCR [MORI, 1992] nos resultados limiarizados e medir a taxa de reconhecimento para limiares diferentes.

Porém em sistemas de GED com processamento de milhões de imagens, o principal fator na escolha de um método é o tempo de processamento. Assim, além de apresentar os melhores resultados em termos de reconhecimento, devem ser escolhidos métodos rápidos.

Dos métodos aqui apresentados, muitos estão implementados no laboratório LUCI (Laboratório Unificado de Ciência da Imagem) da PUC-Pr. Para estes métodos (tab. 4.1 e tab. 4.2) foi possível avaliar os tempos de processamento, e uma avaliação preliminar pôde ser feita em termos de resultados visuais conforme pode ser observado nas tabelas 4.1 e 4.2.

O tempo de processamento, nestes casos, foi normalizado para 1,00 considerando o tempo de Otsu, e os tempos de todos os demais métodos foram considerados em função de Otsu. O método de Otsu foi utilizado como referência por ser muito popular. A popularidade de Otsu se deve ao fato de ser muito rápido e recomendado para documentos em geral.

Como pode ser observado na tabela 4.1, encontram-se métodos de limiarização global que são mais rápidos que o método de Otsu e com resultados visuais tão bons como os de Otsu. Isto se verifica nos resultados dos métodos de Li-Lee, WuLu e LLoyd. Em especial na questão do tempo, estes métodos podem viabilizar o processamento de milhões de imagens, o que é comum em sistemas de GED.

Pode-se observar na tabela 4.2, que os métodos de limiarização local adaptativa apesar de na literatura serem recomendados para documentos complexos, apresentam tempos de processamento proibitivos para soluções de GED com grandes volumes.

Método	Tipo	Tempo	Resultado
Otsu	Global	01,00	BANCO DO ESTADO DO PARANA' S.A.
Kittler	Global	00,93	BANCO DO ESTADO DO PARANA' S.A.
Fuzzy C Means	Global	01,07	BANCO DO ESTADO DO PARANA' S.A.
Fuzzy Huang	Global	01,05	BANCO DO ESTADO DO PARANA' S.A.
Fuzzy Yager	Global	01,03	BANCO DO ESTADO DO PARANA' S.A.
Pun	Global	00,93	BANCO DO ESTADO DO PARANA' S.A.
Johannsen	Global	01,00	BANCO DO ESTADO DO PARANA' S.A.
Li-Lee	Global	00,96	BANCO DO ESTADO DO PARANA' S.A.
WuLu	Global	00,89	BANCO DO ESTADO DO PARANA' S.A.
Thrussel	Global	01,10	(conteúdo impresso totalmente eliminado)
LLoyd	Global	00,87	BANCO DO ESTADO DO PARANA' S.A.

Tabela 4.1 Resultados preliminares de testes nos métodos de limiarização global implementados no laboratório LUCI da PUC-Pr

Método	Tipo	Tempo	Resultado
Bernsen	Local	17,14	BANCO DO ESTADO DO PARANA' S.A.
Eikvil	Local	11,05	BANCO DO ESTADO DO PARANA' S.A.
Niblack	Local	41,14	BANCO DO ESTADO DO PARANA' S.A.

Tabela 4.2 Resultados preliminares de testes nos métodos de limiarização local adaptativa implementados no laboratório LUCI da PUC-Pr

Em resumo, os métodos de limiarização global, têm a grande vantagem de serem rápidos e podem ser utilizados em grandes volumes quando um único valor de limiar para toda a imagem é suficiente. Já os métodos de limiarização local adaptativa, têm grande vantagem quando mais de um limiar para uma mesma imagem é necessário, porém, como são muito lentos, não são adequados para processamento de grandes volumes e ainda têm a desvantagem da necessidade de escolha do tamanho da janela.

Capítulo 5

Morfologia Matemática Binária

5.1 Introdução

Morfologia é a ciência que trata das formas que a matéria pode tomar, qualquer que seja a área em que se aplique (biologia, sociologia, gramática, etc.).

Na área de processamento de imagens, a morfologia matemática é uma ferramenta muito versátil porque ela permite processar imagens com objetivos de realce, de segmentação, de detecção de bordas, de esqueletização, de afinamento, de análise de formas, entre outras.

Encontram-se na literatura [DEBORA, 2000] [JESUS, 1999] [MATRAKAS, 2000], muitos estudos que propõem o uso da morfologia matemática binária com o objetivo de facilitar a localização, segmentação e extração de componentes. Das propostas estudadas e apresentadas no Capítulo 3, poucas utilizam a morfologia matemática binária para melhorar conteúdos impressos [CANON, 1999] [KANUNGO, 1999] [SWAIN, 1997], e dessas, todas propõem um mesmo e único operador morfológico para buscar os resultados esperados.

O princípio básico da morfologia matemática, em termos de imagem, é extrair informações relacionadas à geometria e à topologia de um conjunto desconhecido X . Este conjunto X é comparado, a partir de uma transformação, com um conjunto B completamente definido e conhecido (forma, tamanho), chamado de elemento estruturante. O resultado dessa transformação permite avaliar o conjunto X e extrair as informações de geometria e topologia. A Figura 5.1 apresenta os tipos de elementos estruturantes binários básicos considerados neste trabalho que estão implementados no LUCI (Laboratório Unificado de Ciência da Imagem) da PUC-Pr.

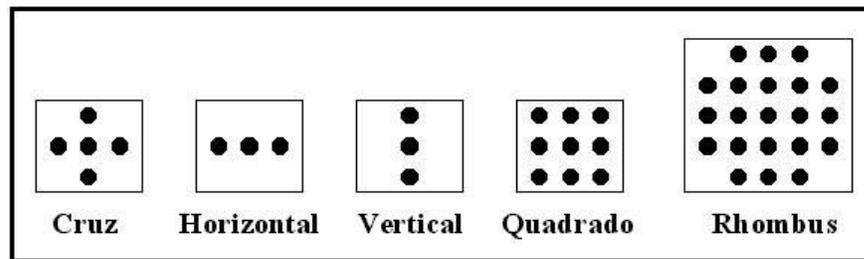


Figura 5.1 Tipos de elementos estruturantes básicos da morfologia matemática binária considerados neste trabalho

O formato e o tamanho do elemento estruturante possibilitam testar e quantificar de que maneira o elemento estruturante interage na imagem. A Figura 5.2 ilustra uma imagem binária, um elemento estruturante e exemplo de iteração.

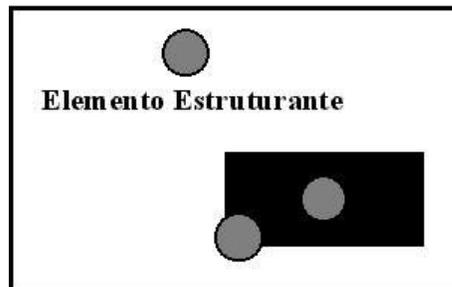


Figura 5.2 Exemplo de iteração do elemento estruturante em uma imagem binária

Com o exemplo da Figura 5.2, nota-se que, em algumas situações, o elemento estruturante encaixa-se à imagem e, em outras, não. Se o formato ou o tamanho desse elemento estruturante for mudado, outras respostas serão obtidas. Portanto, o tipo e a natureza da informação extraída depende necessariamente do tipo de elemento estruturante e do tipo de imagem estudada. Uma operação morfológica binária é completamente determinada através da vizinhança examinada ao redor do ponto central, da configuração dos pontos pretos (pixel ativo) e brancos (pixel inativo ou neutro) nessa vizinhança e do algoritmo. Este tipo de morfologia se aplica em imagens binárias.

A morfologia matemática, ao contrário dos métodos de limiarização apresentados no Capítulo 4, não requer cálculos matemáticos complexos. As operações

têm princípios de funcionamento simples o que garante tempos de processamento muito rápidos se comparados com outros métodos de processamento de imagens.

Nas próximas seções, serão apresentados alguns operadores morfológicos binários com seus respectivos parâmetros e tempos de processamento. Estes operadores são definidos, pelos autores, de maneiras diferentes [BANON, 1994]. Este trabalho adotou o formalismo de Minkowski [SERRA, 1982] e as definições foram extraídas de [FACON, 1996][SERRA, 1982].

5.2 Operadores Morfológicos Binários

5.2.1 Erosão Binária

A erosão de um conjunto X pelo elemento estruturante B é a seguinte:

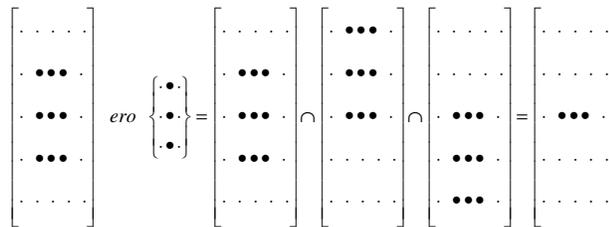
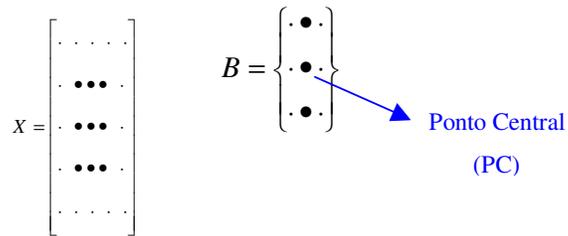
$$\mathcal{E}^B(X) = X \ominus \tilde{B} = \bigcap_{b \in \tilde{B}} X_b \quad (5.1)$$

onde \tilde{B} representa o transposto de B (o elemento transposto é obtido pela simetria do centro a origem dos eixos), \ominus representa a subtração de Minkowski.

Conforme a definição acima, o conjunto a ser erodido, X , é deslocado em função das posições permitidas pelo elemento estruturante B . Os deslocamentos são realizados em relação ao ponto central de \tilde{B} , conforme mostra o exemplo 5.1.

A partir de agora um pixel inativo ou neutro é representado por “.”, e um pixel ativo é representado por “•”.

Exemplo 5.1:



Cada movimento de X é realizado em relação ao ponto central (PC) de \tilde{B} . Neste caso, B é simétrico ($\tilde{B} = B$) e PC representa, portanto, a origem das translações. Verifica-se que a erosão diminuiu o conjunto X .

Os efeitos obtidos com a erosão binária são (Figura 5.3):

- Diminuição das partículas (pixels pretos);
- Eliminação de grãos (pixels pretos) de tamanho inferior ao tamanho do elemento estruturante;
- Aumento dos buracos (pixels brancos);
- Separação de caracteres próximos.

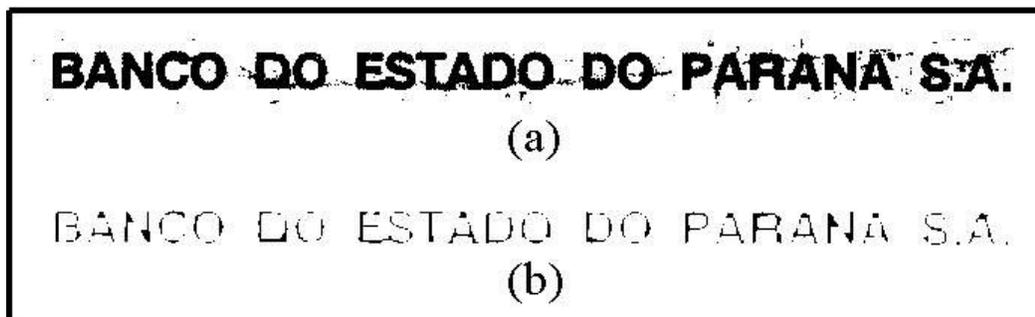


Figura 5.3 Efeitos da erosão binária com elemento estruturante quadrado duas iterações: (a) imagem original; (b) imagem com erosão.

5.2.2 Dilatação Binária

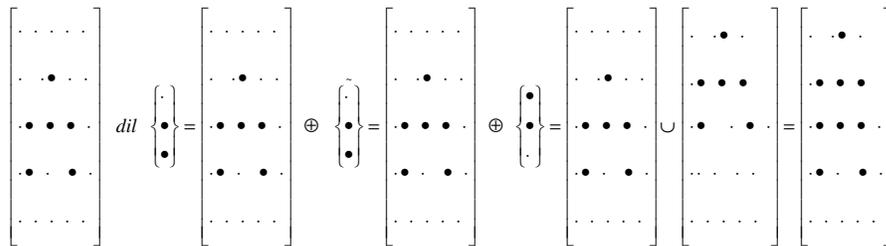
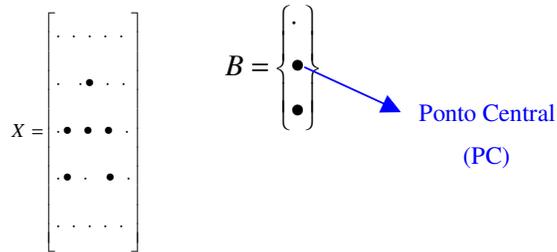
A dilatação de um conjunto X pelo elemento estruturante B é a seguinte:

$$\mathcal{D}^B(X) = X \oplus \tilde{B} = \bigcup_{b \in \tilde{B}} X_b \quad (5.2)$$

onde \tilde{B} representa o transposto de B, \oplus representa a adição de Minkowski.

Conforme a definição acima, o conjunto a ser dilatado, X, é deslocado em função das posições permitidas pelo elemento estruturante B. Os deslocamentos são realizados em relação ao ponto central de \tilde{B} , conforme mostra o exemplo 5.2.

Exemplo 5.2:



X é movimentado em relação ao ponto central PC de \tilde{B} que representa, portanto, a origem das translações. PC faz parte dos pontos relevantes de \tilde{B} . Neste

exemplo, o elemento estruturante não é simétrico, isto faz com que $B \neq \tilde{B}$. A dilatação é uma operação dual à erosão. Ela aumenta o conjunto X .

Os efeitos obtidos com a dilatação binária são (Figura 5.4):

- Expandir partículas;
- Preencher pequenos buracos;
- Conectar caracteres próximos.

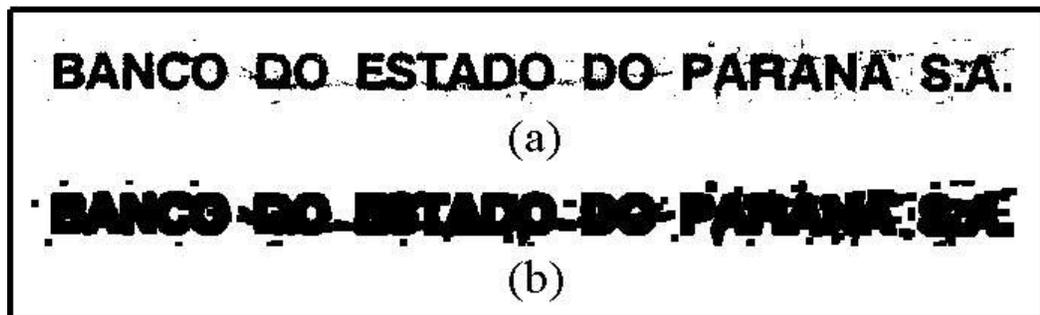


Figura 5.4 Efeitos da dilatação binária com elemento estruturante quadrado duas iterações: (a) imagem original; (b) imagem com dilatação.

5.2.3 Abertura Binária

Ao contrário da erosão e da dilatação, a abertura binária consiste em eliminar as partículas indesejáveis, sem modificar o tamanho das outras entidades. Intuitivamente, podemos prever que esta operação consiste em erodir e, depois, dilatar o resultado da erosão.

A abertura de um conjunto X pelo elemento estruturante B é:

$$\phi^B(X) = \delta^B(\varepsilon^{\tilde{B}}(X)) \quad (5.3)$$

O segundo passo consiste em processar o resultado anterior por adição de Minhowski com B. Essa adição resulta da união de X com X, deslocado segundo as posições permitidas por B.

$$\begin{bmatrix} \cdot & \cdot \\ \cdot & \cdot \\ \cdot & \cdot \\ \cdot & \cdot \\ \cdot & \cdot \\ \cdot & \cdot \\ \cdot & \cdot \\ \cdot & \cdot \\ \cdot & \cdot \\ \cdot & \cdot \\ \cdot & \cdot \end{bmatrix} \oplus \left\{ \begin{array}{c} \cdot \\ \cdot \\ \cdot \end{array} \right\} =$$

$$\begin{bmatrix} \cdot & \cdot \\ \cdot & \cdot \\ \cdot & \cdot \\ \cdot & \cdot \\ \cdot & \cdot \\ \cdot & \cdot \\ \cdot & \cdot \\ \cdot & \cdot \\ \cdot & \cdot \\ \cdot & \cdot \\ \cdot & \cdot \end{bmatrix} \cup \begin{bmatrix} \cdot & \cdot \\ \cdot & \cdot \\ \cdot & \cdot \\ \cdot & \cdot \\ \cdot & \cdot \\ \cdot & \cdot \\ \cdot & \cdot \\ \cdot & \cdot \\ \cdot & \cdot \\ \cdot & \cdot \\ \cdot & \cdot \end{bmatrix} \cup \begin{bmatrix} \cdot & \cdot \\ \cdot & \cdot \\ \cdot & \cdot \\ \cdot & \cdot \\ \cdot & \cdot \\ \cdot & \cdot \\ \cdot & \cdot \\ \cdot & \cdot \\ \cdot & \cdot \\ \cdot & \cdot \\ \cdot & \cdot \end{bmatrix} = \begin{bmatrix} \cdot & \cdot \\ \cdot & \cdot \\ \cdot & \cdot \\ \cdot & \cdot \\ \cdot & \cdot \\ \cdot & \cdot \\ \cdot & \cdot \\ \cdot & \cdot \\ \cdot & \cdot \\ \cdot & \cdot \\ \cdot & \cdot \end{bmatrix}$$

O resultado mostra os dois conjuntos de X desconectados e com tamanho e forma idênticos aos de X.

Os efeitos obtidos pela abertura binária são (Figura 5.5):

- Não retornar, de forma geral, o conjunto inicial;
- Nivelar os contornos pelo interior;
- Separar os caracteres;
- Eliminar as pequenas partículas inferiores em tamanho, com relação ao elemento estruturante;
- As entidades restantes após a abertura ficam quase idênticas às originais;
- O conjunto aberto é mais regular que o conjunto inicial;
- O conjunto aberto é menos rico em detalhes que o conjunto inicial.

- As entidades restantes após o fechamento ficam quase idênticas;
- O conjunto fechado é mais regular que o conjunto inicial;
- Conjunto fechado é menos rico em detalhes que o conjunto inicial.

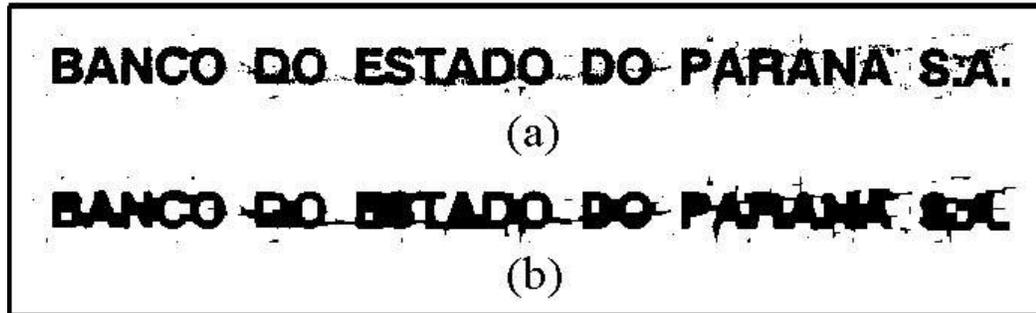


Figura 5.6 Efeitos do fechamento binário com elemento estruturante quadrado duas iterações: (a) imagem original; (b) imagem com fechamento.

5.2.5 Reconstrução Binária

O processo de reconstrução consiste em recuperar um conjunto S a partir de um conjunto Z . Tendo-se um subconjunto S de uma imagem binária X e uma fração Z de S , pode-se afirmar que S é marcado por Z . Assim define-se Z como **marcador** de S e define-se S como **máscara**.

A reconstrução permite eliminar os padrões indesejáveis sem deformar os desejáveis (fig. 5.7), porém diferentemente da abertura (fig. 5.5), a reconstrução não elimina ruídos que estejam conectados aos conteúdos impressos. Somente os ruídos desconectados dos conteúdos impressos e menores que o elemento estruturante são eliminados.

Por eliminar padrões indesejáveis sem deformar os desejáveis, a reconstrução é muito eficiente para muitas aplicações, como por exemplo em filtragens e separação de objetos com tamanhos diferentes [FACON, 1996]. Porém, no caso de conteúdos impressos, por não eliminar os ruídos conectados aos caracteres (fig. 5.7), o resultado da reconstrução, neste caso, contribui pouco para a melhoria da qualidade e o reconhecimento.

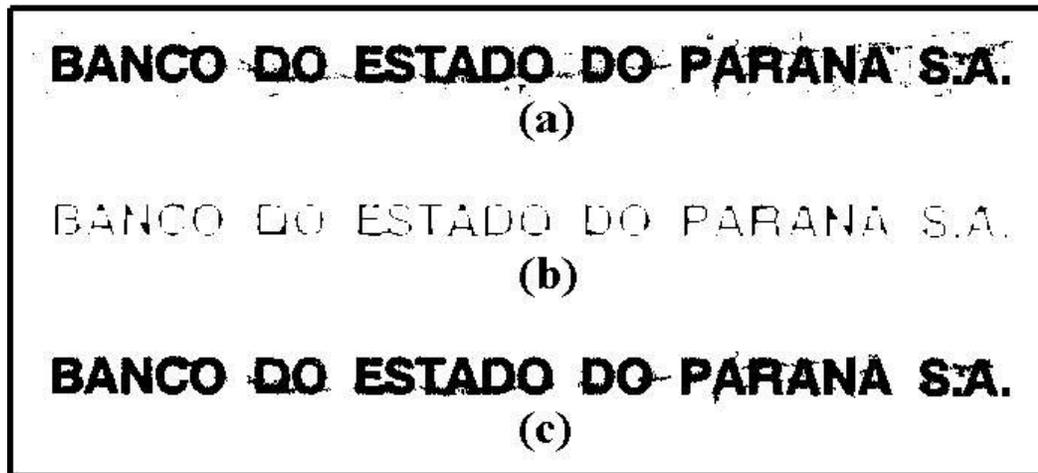


Figura 5.7 Processo da reconstrução binária: (a) imagem original; (b) marcador obtido pela erosão da imagem original com elemento estruturante quadrado duas iterações; (c) imagem reconstruída a partir do marcador, utilizando a imagem original como máscara.

5.3 Escolha de Operadores, Elementos Estruturantes e Iterações

A escolha de parâmetros para morfologia matemática binária (operadores, elementos estruturantes e número de iterações) sempre deve estar de acordo com os objetivos que se pretende atingir. Para o caso de se buscar melhorias em conteúdos impressos com grandes volumes, duas importantes condições devem ser observadas: primeiro o tempo de processamento deve ser rápido, segundo os parâmetros devem contribuir mais para a efetiva melhoria dos conteúdos.

5.3.1 Escolha de Operadores Morfológicos

De acordo com o que já foi apresentado e observando-se os resultados da tabela 5.1, pode-se estabelecer a hipótese (a ser comprovada nos testes e validação) de que os operadores erosão e abertura, neste caso, contribuem para a melhoria de conteúdos impressos.

Como o propósito deste trabalho é a melhoria dos conteúdos impressos em grandes volumes, combinando técnicas de limiarização com a morfologia matemática binária, busca-se um conjunto reduzido de parâmetros. Assim, neste trabalho, serão considerados apenas os operadores morfológicos da erosão e abertura binárias, uma vez que em testes preliminares realizados, mostraram-se muito eficientes em termos de tempo de processamento e apresentaram contribuição maior para a melhoria dos conteúdos impressos.

Com relação ao tempo de processamento, observa-se (tab. 5.1) que tanto a erosão como a abertura são muito rápidos, comparando-se com o tempo de Otsu, método pelo qual a imagem foi limiarizada antes de ser submetida a morfologia matemática binária.

Operação/EE/Iterações	Tempo	Resultado Visual
Imagem Original (OTSU)	1,00	BANCO DO ESTADO DO PARANA S.A.
Erosão Quadrado Uma	0,18	BANCO DO ESTADO DO PARANA S.A.
Dilatação Quadrado Uma	0,18	BANCO DO ESTADO DO PARANA S.A.
Abertura Quadrado Uma	0,36	BANCO DO ESTADO DO PARANA S.A.
Fechamento Quadrado Uma	0,36	BANCO DO ESTADO DO PARANA S.A.
Reconstrução Quadrado Uma	0,60	BANCO DO ESTADO DO PARANA S.A.

Tabela 5.1 Tempo e resultado visual das operações com Morfologia Matemática Binária (EE = Elementos Estruturantes)

5.3.2 Escolha de Elementos Estruturantes

A tabela 5.2 mostra que os elementos estruturantes horizontal, vertical e rhombus não se mostram adequados para melhorar conteúdos impressos.

Operação/EE/Iterações	Resultado Visual
Abertura Horizontal Três	BANCO DO ESTADO DO PARANA S.A.
Abertura Vertical Três	BANCO DO ESTADO DO PARANA S.A.
Abertura Rhombus Três	(conteúdo impresso totalmente eliminado)

Tabela 5.2 Resultado visual com elementos estruturantes (EE) não adequados para conteúdos impressos

Enquanto os elementos estruturantes horizontal e vertical deformam os conteúdos nas respectivas direções, o elemento estruturante rhombus é muito invasivo. Sendo assim, neste caso, os elementos estruturantes mais adequados são cruz e quadrado.

5.3.3 Escolha do Número de Iterações

Cada operador morfológico combinado com um elemento estruturante, pode ser aplicado uma ou mais vezes sobre a mesma imagem na mesma operação, o que determina o número de iterações.

A tabela 5.3 mostra que independente do elemento estruturante, quanto maior o número de iterações maior também é o tempo de processamento e mais forte é o efeito da operação em questão.

Operação/EE/Iterações	Tempo	Resultado Visual
Imagem Original (OTSU)	1,00	BANCO DO ESTADO DO PARANA S.A.
Abertura Cruz Uma	0,30	BANCO DO ESTADO DO PARANA S.A.
Abertura Cruz Duas	0,44	BANCO DO ESTADO DO PARANA S.A.
Abertura Cruz Três	0,60	BANCO DO ESTADO DO PARANA S.A.
Abertura Quadrado Uma	0,36	BANCO DO ESTADO DO PARANA S.A.
Abertura Quadrado Duas	0,60	BANCO DO ESTADO DO PARANA S.A.
Abertura Quadrado Três	0,80	'' '' ''

Tabela 5.3 Tempo e Resultado visual com progressão do número de iterações

Pode-se observar (tab. 5.3) que para o elemento estruturante cruz a partir da terceira iteração e para o elemento quadrado a partir da segunda iteração, os efeitos tendem a ficar muito invasivos e portanto não adequados para a melhoria de conteúdos impressos.

Capítulo 6

Metodologia Proposta e Testes

6.1 Metodologia Proposta

Soluções para Gerenciamento Eletrônico de Documentos têm funcionalidades que requerem imagens de boa qualidade. A funcionalidade de indexação requer altas taxas de reconhecimento para poder ser automatizada. A funcionalidade de conversão de imagens em textos requer taxas de reconhecimento boas o suficiente para evitar o processo manual via digitação. A funcionalidade de FTR (*Full Text Retrieval*) requer além de altas taxas de reconhecimento para aumentar a eficácia na recuperação das imagens, também imagens de boa qualidade uma vez que estas devem permanecer armazenadas para posterior acesso.

O principal problema surge quando uma solução de GED necessita viabilizar documentos complexos. Nestes casos, em geral, as taxas de reconhecimento via OCR [MORI, 1992] apresentam resultados insatisfatórios como se pode observar na figura 6.1. Embora os produtores de software de OCR anunciem altas taxas de reconhecimento para os seus produtos, isto só é verdadeiro para documentos simples [MELLO, 1999]. Para resolver este problema, deve-se buscar soluções que possam melhorar a qualidade dos conteúdos impressos de documentos complexos e por conseqüência as taxas de reconhecimento.



Figura 6.1 Taxas de reconhecimento em documentos simples versus taxas de reconhecimento em documentos complexos

Outro problema a ser considerado é a abordagem utilizada para geração de imagens de documentos em soluções de GED. Na abordagem tradicional para grandes volumes [CDIA, 2001], as imagens são digitalizadas em preto e branco, ou seja, a limiarização é executada com recursos do scanner onde na maioria dos casos somente a limiarização global está disponível, em geral, com método de ajuste manual do limiar, e em muitos casos com o método de Otsu. A abordagem tradicional consegue gerar imagens de boa qualidade a partir de documentos simples, mas como em geral, a mesma abordagem é utilizada para documentos complexos (fig. 6.2), não se consegue gerar imagens com qualidade satisfatória (fig. 6.3). Isto por consequência gera baixas taxas de reconhecimento, o que na maioria dos casos inviabiliza em soluções de GED funcionalidade como: indexação automática, conversão de conteúdos impressos de imagens para texto e FTR.

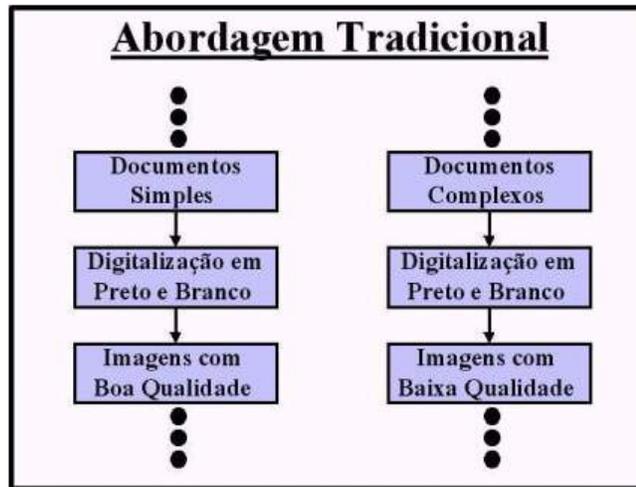


Figura 6.2 Mesma abordagem para documentos simples e complexos em soluções de GED

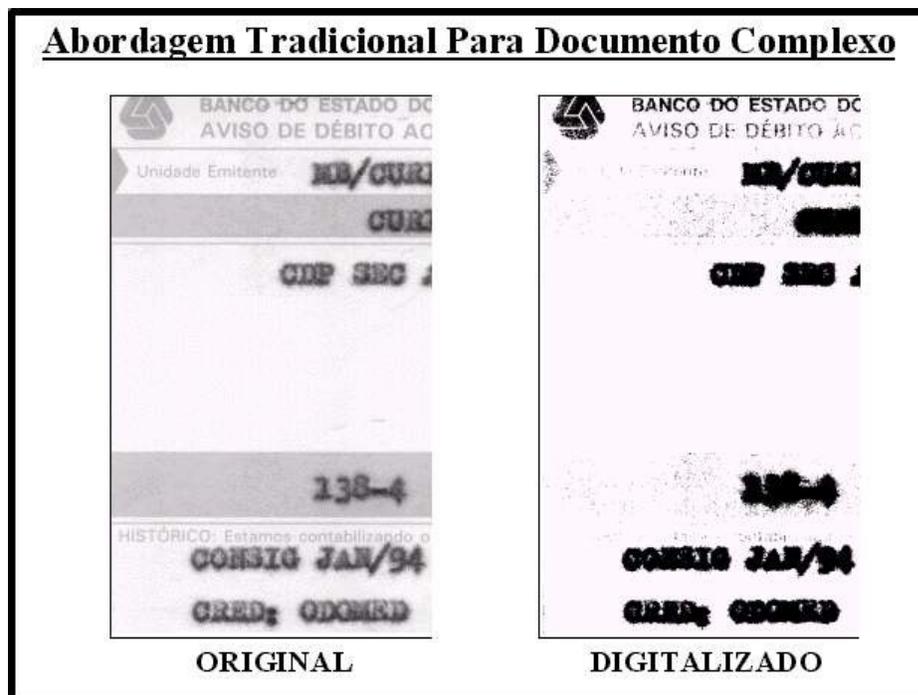


Figura 6.3 Imagem de documento complexo, digitalizado utilizando a abordagem tradicional

Considerando os problemas apresentados, com a conseqüente necessidade de maior qualidade e melhores taxas de reconhecimento a partir de documentos complexos, torna-se necessária uma nova abordagem. Esta nova abordagem, pode ser viabilizada com a aplicação combinada de técnicas de processamento de imagens, desde que as 3 seguintes hipóteses sejam verdadeiras:

- **Primeira:** Aplicando-se técnicas adequadas de limiarização global em regiões de documentos complexos com características particulares (rotulação), obtêm-se incremento de qualidade dos conteúdos impressos em relação a abordagem tradicional, evitando-se o uso de limiarização local adaptativa que por ser muito lenta não é apropriada para grandes volumes.
- **Segunda:** Encontram-se métodos de limiarização global para melhorar conteúdos impressos de documentos complexos que são melhores que Otsu, tanto em termos de resultados como em termos de tempo de processamento o que pode ser comprovado pela limiarização múltipla.
- **Terceira:** Aplicando-se parâmetros adequados, é possível através da morfologia matemática binária aprimorar a qualidade de conteúdos impressos de documentos complexos limiarizados.

O objetivo principal deste trabalho é comprovar estas 3 hipóteses, utilizando as bases de teste e validação definidas no capítulo 2, com a aplicação dos métodos de limiarização mais adequados estudados no capítulo 4, combinados com os parâmetros mais indicados da morfologia matemática binária estudada no capítulo 5. Para isso, propõe-se uma metodologia, a qual tem sua visão geral representada pela figura 6.4 e cujas fases estão a seguir detalhadas.



Figura 6.4 Visão geral das fases da metodologia proposta

6.1.1 Fase de Captura

Os documentos que compõem as bases (134 de teste e 268 para validação), foram capturados via scanner com 256 níveis de cinza e 300DPI. A captura foi efetuada em níveis de cinza para que métodos adequados de limiarização sejam utilizados para converter as imagens para preto e branco, melhorando os conteúdos impressos, reduzindo o seu tamanho e possibilitando a aplicação da morfologia matemática binária para incrementar a qualidade. Os documentos foram capturados com resolução de 300DPI porque esta é hoje a resolução mais utilizada em soluções de GED quando imagens são geradas com objetivo de reconhecimento via OCR.

Para esta fase, o único recurso de scanner utilizado foi a correção da inclinação, para evitar dificuldades na fase de rotulação.

6.1.2 Fase de Rotulação

As imagens geradas na fase de captura, foram rotuladas identificando-se 5 regiões com características particulares conforme foi apresentado no capítulo 2. A rotulação foi adotada nesta metodologia com dois objetivos: 1) possibilitar o uso de métodos de limiarização global em uma região como se fosse uma imagem inteira, evitando o uso de métodos locais adaptativos que por serem lentos, não são apropriados para grandes volumes. 2) avaliar o comportamento dos métodos de limiarização e morfologia matemática binária em mais de um tipo de conteúdo impresso.

Considerando que a modelagem automática das degradações em regiões é ainda uma questão em aberto [BAIRD, 2000], os pontos para segmentação foram encontrados manualmente, codificados como parâmetros, a partir dos quais com uma ferramenta desenvolvida especialmente para este fim (fig. 6.5) foi possível gerar todos os rótulos de forma automática. A opção “Remove bordas das sub-imagens” permite que eventuais bordas remanescentes que envolvem as regiões sejam eliminadas. A opção “Remove cantos das sub-imagens” não foi utilizada porque nos documentos das bases, as eventuais bordas que envolvem as regiões não têm cantos arredondados. A opção “Remove bordas das imagens originais” foi utilizada com objetivo de facilitar a parametrização das regiões para a rotulação.

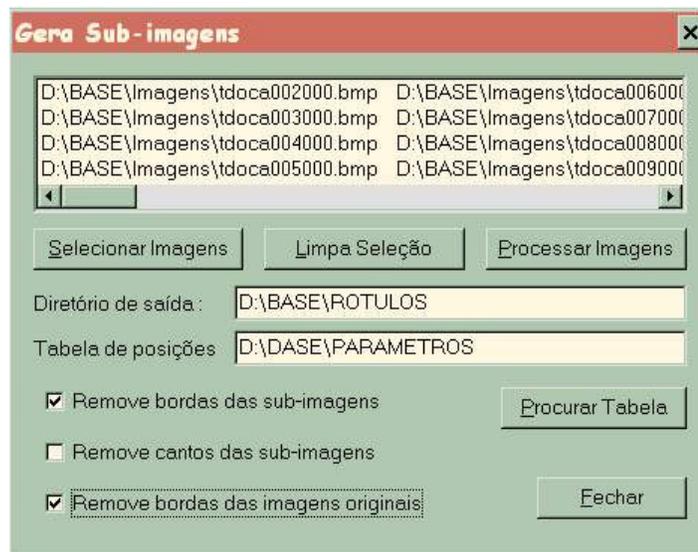


Figura 6.5 Ferramenta desenvolvida em Visual C para auxiliar na rotulação

Todos os 2.010 rótulos foram nomeados conforme exemplo: **tdoca001AFC.bmp**, onde: **t** indica a base (t = teste, v = validação), **doca** neste caso é uma constante fixa (indica o nome do documento), **001** indica o número do documento na base (para t = 001 a 134, para v = 001 a 268), **AFC** indica a região com características particulares conforme definido no capítulo 2, e **.bmp** indica o formato de gravação da imagem.

6.1.3 Fase de Processamento

Esta é a principal fase da metodologia. Nesta fase todos os rótulos foram processados combinando-se a técnica de limiarização múltipla com a morfologia matemática binária. A técnica de limiarização múltipla consiste em processar uma mesma imagem com métodos de limiarização diferentes com o objetivo de encontrar o método mais adequado para as características da imagem em questão. Para a limiarização múltipla foi escolhido o popular método de Otsu e os três melhores e mais rápidos que Otsu de acordo com testes preliminares mostrados no capítulo 4: WuLu, Li-Lee e LLoyd.

Para a morfologia matemática binária, foram considerados os parâmetros mais adequados de acordo com testes preliminares mostrados no capítulo 5, ou seja, as operações de abertura e erosão, os elementos estruturantes cruz e quadrado, sendo para cruz uma e duas iterações e para o quadrado uma iteração.

Para realizar o processamento de forma automática, foi construída uma ferramenta (fig. 6.6 e 6.7) através da qual todos os rótulos foram processados.

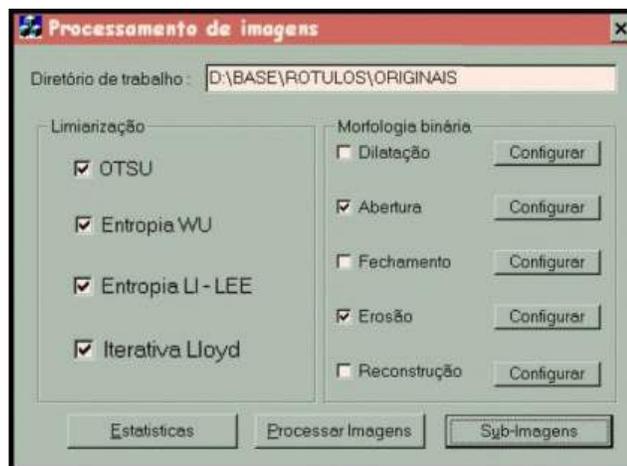


Figura 6.6 Ferramenta desenvolvida em Visual C para processamento dos rótulos



Figura 6.7 Configuração dos parâmetros da morfologia matemática binária utilizados para elementos estruturantes e número de iterações

O processamento dos rótulos da base de teste e validação, que gerou um total de 56.280 imagens processadas foi realizado da seguinte forma:

- Cada um dos 2.010 rótulos foi processado com a técnica de limiarização múltipla, gerando 4 imagens processadas para cada rótulo conforme os 4 métodos de limiarização escolhidos (Otsu, WuLu, Li-Lee e LLoyd). Tendo-se produzido nesta etapa 8.040 imagens limiarizadas, cada uma foi nomeada de forma a preservar a sua identificação original, acrescentando-se na identificação, o método ao qual foi submetida conforme exemplo: **tdoca001AFCOTSU.bmp** onde **OTSU** é o método de limiarização, ainda podendo ser: ENWU para entropia de WuLu, ENLI para entropia de Li-Lee e ITLL para iterativa de LLoyd.
- Cada uma das 8.040 imagens limiarizadas foi automaticamente submetida a morfologia matemática binária de acordo com os parâmetros escolhidos (abertura, erosão, cruz-1, cruz-2 e quadrado-1), gerando-se assim 48.240 imagens limiarizadas, combinadas com morfologia matemática binária. Cada imagem gerada foi nomeada de forma a preservar as características anteriores e identificar os parâmetros da morfologia matemática binária a qual foi submetida conforme exemplo: **tdoca001AFCOTSU+ABEc02.bmp** onde: **ABE** significa a operação morfológica (ABE para abertura e ERO para erosão), **c** significa o elemento estruturante (**c** para cruz e **q** para quadrado) e **02** significa o número de iterações.

- Todas as imagens processadas foram reunidas (totalizando 56.280) para serem submetidas a próxima fase da metodologia que é o reconhecimento.

Dessa forma, uma imagem nomeada como **vdoca131NFEENWU+EROq01.bmp** deve ser assim interpretada: Da base para validação, a região NFE do documento 131, foi primeiramente limiarizada pela entropia de WuLu e em seguida o resultado foi submetido a erosão binária com elemento estruturante quadrado uma iteração.

6.1.4 Fase de Reconhecimento

Nesta fase todas as 56.280 imagens geradas na fase anterior foram reconhecidas de forma automática utilizando-se o OCR comercial OMNIPAGE10 [CAERE, 2001], onde para cada imagem reconhecida foi gerado um texto com o conteúdo dos caracteres supostamente reconhecidos. O objetivo aqui foi criar condições para poder avaliar o efeito de cada processamento sobre cada rótulo através de reconhecimento.

Optou-se por utilizar o software OMNIPAGE10, porque em uma análise comparativa entre diversos softwares de OCR, Mello e Lins [MELLO, 1999] concluíram que o Omnipage alcançou as melhores taxas de reconhecimento em quase todos os casos analisados. Porém estes softwares possuem funcionalidades de análise de contexto que podem gerar como reconhecidos caracteres que de forma isolada não seriam reconhecidos e assim invalidar os resultados desta metodologia. O software OMNIPAGE10 tem 3 dessas funcionalidades:

- Com a opção “Most Accurate”, um conjunto de caracteres como por exemplo **319A385** pode ser reconhecido como **3194385** se a letra **A** estiver um pouco deteriorada. Neste caso na dúvida entre **A** e **4**, o software pode atribuir **4** porque no contexto o conjunto poderia se supor ser somente de dígitos.
- Com a opção “User Training File” o software é “forçado” a considerar um conjunto de pixels como um caracter, o que de forma normal não aconteceria.

- Com a opção “User Dictionary” a análise de contexto é inevitável. Um conjunto de caracteres como, por exemplo **foinalha** pode ser reconhecido como **fornalha** se o caracter **i** estiver acrescido de ruído.

Por estas razões, todos os cuidados foram tomados para que a análise de contexto não fosse considerada. Para isso, o OMNIPAGE10 foi parametrizado conforme mostra a figura 6.8 e assim todas as imagens foram reconhecidas.

Cada imagem reconhecida, teve o seu conteúdo gravado como texto e nomeado de forma a se poder identificar a imagem a partir da qual o texto foi gerado, conforme exemplo:

Imagem reconhecida => **vdoca131NFEENWU+EROq01.bmp**

Texto gerado como => **vdoca131NFEENWU+EROq01.txt**

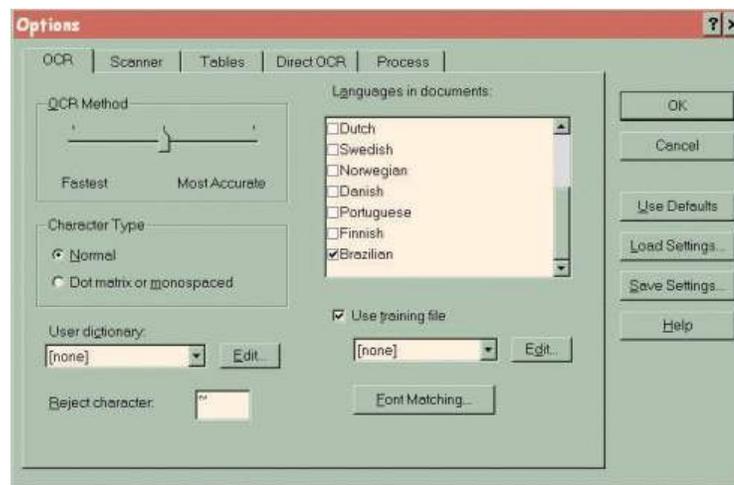


Figura 6.8 Parâmetros do OCR OMNIPAGE10 utilizados em todos os processamentos

6.1.5 Fase de Digitação

Esta fase tem como importante propósito, possibilitar que a avaliação da qualidade (na próxima fase) seja feita de forma automática e objetiva, evitando-se assim avaliações visuais, que na maioria dos casos desse tipo, carregam alto grau de subjetividade com a consequente dificuldade de julgamento.

Cada rótulo gerado na fase de rotulação foi digitado e conferido, criando-se para cada caracter existente na imagem o correspondente conteúdo em formato texto.

Cada conteúdo correspondente ao seu respectivo rótulo, foi nomeado de forma a permitir a sua perfeita identificação e relação com o rótulo que lhe deu origem, como por exemplo: Rótulo = tdoca001AFC.bmp digitado em tdoca001AFC.txt

Concluída a fase de digitação, têm-se todos os dados necessários para efetuar a avaliação da qualidade. A partir daqui, as imagens não são mais necessárias, uma vez que o conteúdo digitado e o conteúdo reconhecido (fig. 6.9) são suficientes para executar a próxima e última fase desta metodologia.

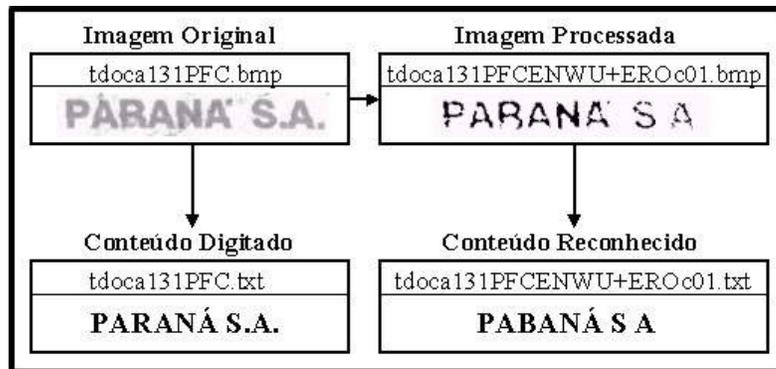


Figura 6.9 Visão geral da geração dos dados necessários para avaliação da qualidade

6.1.6 Fase de Avaliação da Qualidade

Para a avaliação da qualidade foi construída uma ferramenta específica conforme mostra a figura 6.10.



Figura 6.10 Ferramenta desenvolvida em Visual C para avaliação da qualidade

Para executar a avaliação, informa-se o diretório onde estão os conteúdos reconhecidos e o nível de consolidação dos resultados. Aqui foi utilizada a opção “Tipo/Recorte” onde “Tipo” é a base (teste ou validação) e “Recorte” é o rótulo.

A avaliação é totalmente automática e foi executada em duas etapas:

Na primeira etapa: Cada conteúdo reconhecido foi comparado caracter a caracter com o respectivo conteúdo digitado e um resultado parcial foi gerado conforme mostra a figura 6.11, onde: o número **10** é a quantidade de caracteres existentes no conteúdo digitado (neste caso PARANÁ S.A.), o número **7** é a quantidade de caracteres reconhecidos corretos (neste caso PABANÁ S A), o número **1** é a quantidade de caracteres reconhecidos errados (neste caso a letra R foi reconhecida como B) e o número **2** é a quantidade de caracteres não reconhecidos (neste caso ..).

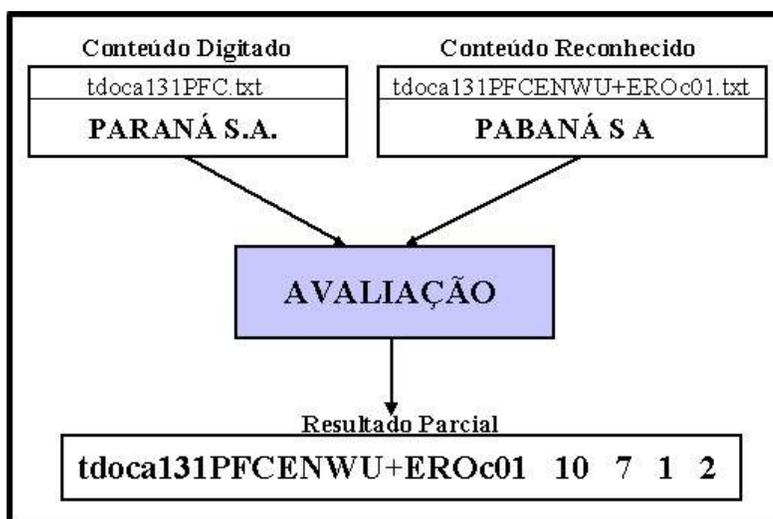


Figura 6.11 Exemplo de avaliação da qualidade na etapa de geração de resultado parcial

Na segunda etapa: Cada conjunto de resultados parciais (tab. 6.1) foi totalizado por rótulo/métodos e um resultado foi gerado conforme mostra a tabela 6.2.

tdoca001PFCENWU+EROc01	10	9	1	0
tdoca002PFCENWU+EROc01	10	8	1	1
tdoca003PFCENWU+EROc01	10	9	0	1
tdoca004PFCENWU+EROc01	10	8	0	2
tdoca005PFCENWU+EROc01	10	7	2	1
tdoca006PFCENWU+EROc01	10	9	1	0
.....				
.....				
.....				
tdoca131PFCENWU+EROc01	10	7	1	2
tdoca132PFCENWU+EROc01	10	9	0	1
tdoca133PFCENWU+EROc01	10	8	1	1
tdoca134PFCENWU+EROc01	10	9	1	0

Tabela 6.1 Exemplo de resultados parciais do rótulo PFC para os métodos de processamento: ENWU+EROc01

Base	Rótulo	Métodos	Total-Car	%Rec-Corr	%Rec-Err	%Não-Rec
t	PFC	ENWU+EROc01	1340	60	10	30

Tabela 6.2 Exemplo de resultado da totalização de PFC para os métodos de processamento: ENWU+EROc01

Os resultados da tabela 6.2 devem ser assim interpretados: Da base de teste, todos os rótulos PFC, foram primeiramente limiarizados por entropia de WuLu, em seguida o resultado foi submetido a erosão binária com elemento estruturante cruz uma iteração. A avaliação da qualidade totalizada no rótulo PFC para ENWU+EROc01 encontrou um total de 1340 caracteres, dos quais 60% foram reconhecidos corretos, 10% foram reconhecidos errados e 30% não foram reconhecidos.

6.2 Resultados dos Testes

Os resultados dos testes estão apresentados nas páginas seguintes, onde para cada rótulo é mostrada uma tabela a partir da qual foram gerados 2 gráficos, conforme:

Seção	Rótulo	Tabela	Gráficos
6.2.1	AFE	6.3	6.12 e 6.13
6.2.2	AFC	6.4	6.14 e 6.15
6.2.3	NFE	6.5	6.16 e 6.17
6.2.4	NFC	6.6	6.18 e 6.19
6.2.5	PFC	6.7	6.20 e 6.21

Optou-se mostrar nas tabelas todos os resultados obtidos para o rótulo, por conterem importantes informações para a análise apresentada na seção 7.2.

A primeira linha de dados de cada tabela corresponde aos resultados obtidos com a base de testes utilizando-se a abordagem tradicional em soluções de GED. Para isso, os documentos foram capturados em preto e branco, rotulados, reconhecidos e avaliada a qualidade de cada rótulo, sem utilizar técnicas de processamento de imagem.

A partir da segunda linha, para facilitar a análise, os dados estão classificados em ordem decrescente de reconhecidos corretos (%Rec-Corr).

Os dados de ENLI, ENWU, ITLL e OTSU correspondem aos resultados das respectivas limiarizações, sem aplicação de morfologia matemática binária.

Os gráficos, gerados a partir dos reconhecidos corretos, têm como principal objetivo facilitar a verificação de prova das 3 hipóteses anteriormente formuladas:

- **Primeira:** verifica-se que a abordagem tradicional mostrada no gráfico “**Tradicional X Metodologia**” tem resultados significativamente menores que cada um dos resultados dos métodos de limiarização global mostrados no gráfico “**Otsu X Demais métodos + Morfologia Binária**”.
- **Segunda:** verifica-se que no gráfico “**Otsu X Demais métodos + Morfologia Binária**” os resultados de WuLu, Li-Lee e LLOYD são melhores que os de Otsu.
- **Terceira:** verifica-se no gráfico “**Otsu X Demais métodos + Morfologia Binária**” que os resultados da morfologia matemática binária, na maioria dos casos, apresentam significativo acréscimo na melhoria da qualidade.

6.2.1 Resultados do Rótulo AFE

Base	Rótulo	Métodos	Total-Car	%Rec-Corr	%Rec-Err	%Não-Rec
t	AFE	TRADICIONAL	1870	1,18	7,91	90,91
t	AFE	ITLL+EROq01	1870	47,06	36,04	16,9
t	AFE	ENLI+EROq01	1870	43,21	35,99	20,8
t	AFE	ITLL+EROC01	1870	42,35	38,82	18,83
t	AFE	ENLI+EROC02	1870	42,09	40,32	17,59
t	AFE	ITLL+EROC02	1870	41,87	41,55	16,58
t	AFE	OTSU+EROC02	1870	40,53	38,02	21,45
t	AFE	ITLL+ABEq01	1870	39,84	40,05	20,11
t	AFE	OTSU+EROq01	1870	38,56	39,14	22,3
t	AFE	ENWU+ABEc01	1870	37,54	44,81	17,65
t	AFE	ENLI+EROC01	1870	36,74	36,63	26,63
t	AFE	ENWU	1870	34,55	34,71	30,74
t	AFE	ITLL+ABEc02	1870	34,44	35,83	29,73
t	AFE	ENLI+ABEq01	1870	33,16	40,59	26,25
t	AFE	ITLL+ABEc01	1870	32,09	40,53	27,38
t	AFE	ENWU+ABEq01	1870	30,96	42,3	26,74
t	AFE	ENWU+EROC01	1870	30,16	42,57	27,27
t	AFE	ENLI+ABEc02	1870	28,82	41,87	29,31
t	AFE	ENWU+ABEc02	1870	27,75	51,76	20,49
t	AFE	OTSU+EROC01	1870	27,43	38,88	33,69
t	AFE	ENLI+ABEc01	1870	23,64	42,94	33,42
t	AFE	ENWU+EROq01	1870	23,58	41,6	34,82
t	AFE	OTSU+ABEq01	1870	23,16	41,39	35,45
t	AFE	ITLL	1870	19,52	39,3	41,18
t	AFE	OTSU+ABEc02	1870	17,97	42,3	39,73
t	AFE	OTSU+ABEc01	1870	15,67	42,83	41,5
t	AFE	ENWU+EROC02	1870	14,44	42,46	43,1
t	AFE	ENLI	1870	12,67	37,91	49,42
t	AFE	OTSU	1870	7,38	35,78	56,84

Tabela 6.3 Resultados de teste da limiarização combinada com morfologia binária do rótulo AFE

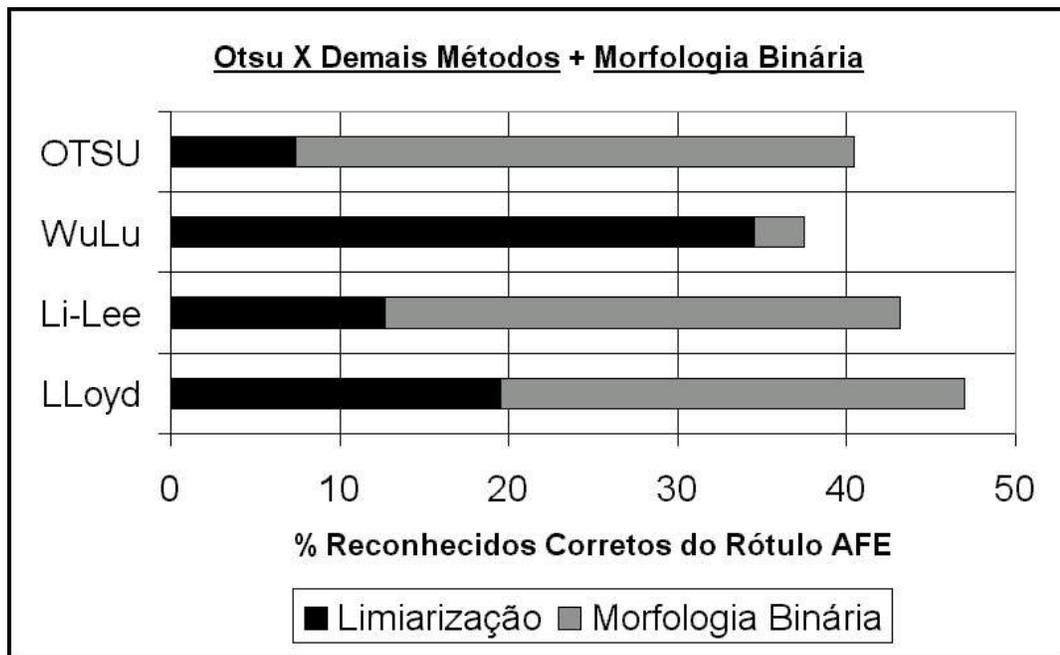


Figura 6.12 Análise comparativa dos resultados de teste para o rótulo AFE: Método de Otsu versus WuLu, Li-Lee e LLoyd; acrescidos dos melhores resultados da morfologia matemática binária

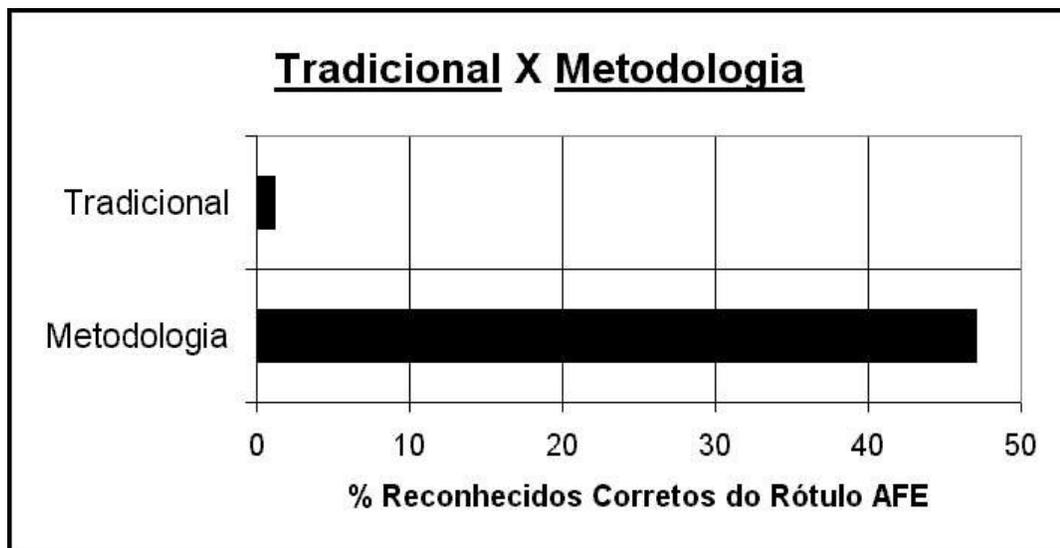


Figura 6.13 Análise comparativa entre a abordagem tradicional e o melhor resultado obtido nos testes da metodologia proposta para o rótulo AFE

6.2.2 Resultados do Rótulo AFC

Base	Rótulo	Métodos	Total-Car	%Rec-Corr	%Rec-Err	%Não-Rec
t	AFC	TRADICIONAL	1418	12,2	48,73	39,07
t	AFC	ITLL+EROq01	1418	49,37	41,75	8,88
t	AFC	ENLI+EROc02	1418	48,59	43,51	7,9
t	AFC	ITLL+EROc01	1418	47,81	41,33	10,86
t	AFC	ENLI+EROq01	1418	47,46	39,92	12,62
t	AFC	ITLL+EROc02	1418	47,04	46,61	6,35
t	AFC	OTSU+EROc02	1418	42,6	47,39	10,01
t	AFC	ITLL+ABEq01	1418	40,76	45,63	13,61
t	AFC	OTSU+EROq01	1418	39,28	44,5	16,22
t	AFC	ENLI+EROc01	1418	38,36	45,35	16,29
t	AFC	ITLL+ABEc01	1418	37,87	45,42	16,71
t	AFC	ITLL+ABEc02	1418	37,45	45,84	16,71
t	AFC	OTSU+EROc01	1418	30,61	48,17	21,22
t	AFC	ENLI+ABEq01	1418	29,76	49,58	20,66
t	AFC	ENLI+ABEc02	1418	29,27	51,41	19,32
t	AFC	ITLL	1418	27,08	49,08	23,84
t	AFC	ENLI+ABEc01	1418	26,52	53,46	20,02
t	AFC	OTSU+ABEc02	1418	23,84	55,15	21,01
t	AFC	OTSU+ABEq01	1418	23,41	51,62	24,97
t	AFC	ENLI	1418	19,39	55,43	25,18
t	AFC	OTSU+ABEc01	1418	18,34	55,99	25,67
t	AFC	ENWU	1418	15,16	51,83	33,01
t	AFC	OTSU	1418	13,05	53,67	33,28
t	AFC	ENWU+ABEc01	1418	8,11	39,56	52,33
t	AFC	ENWU+ABEc02	1418	4,23	37,17	58,6
t	AFC	ENWU+ABEq01	1418	4,23	28,42	67,35
t	AFC	ENWU+EROc01	1418	3,67	20,87	75,46
t	AFC	ENWU+EROq01	1418	2,47	29,48	68,05
t	AFC	ENWU+EROc02	1418	0,78	14,46	84,76

Tabela 6.4 Resultados de teste da limiarização combinada com morfologia binária do rótulo AFC

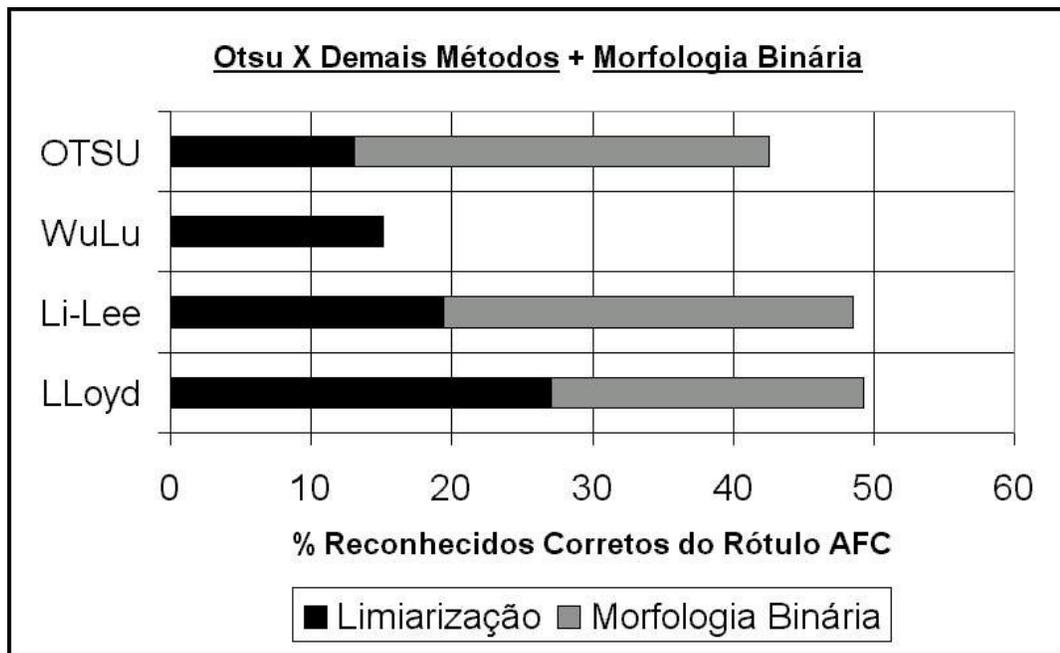


Figura 6.14 Análise comparativa dos resultados de teste para o rótulo AFC: Método de Otsu versus WuLu, Li-Lee e LLoyd; acrescidos dos melhores resultados da morfologia matemática binária

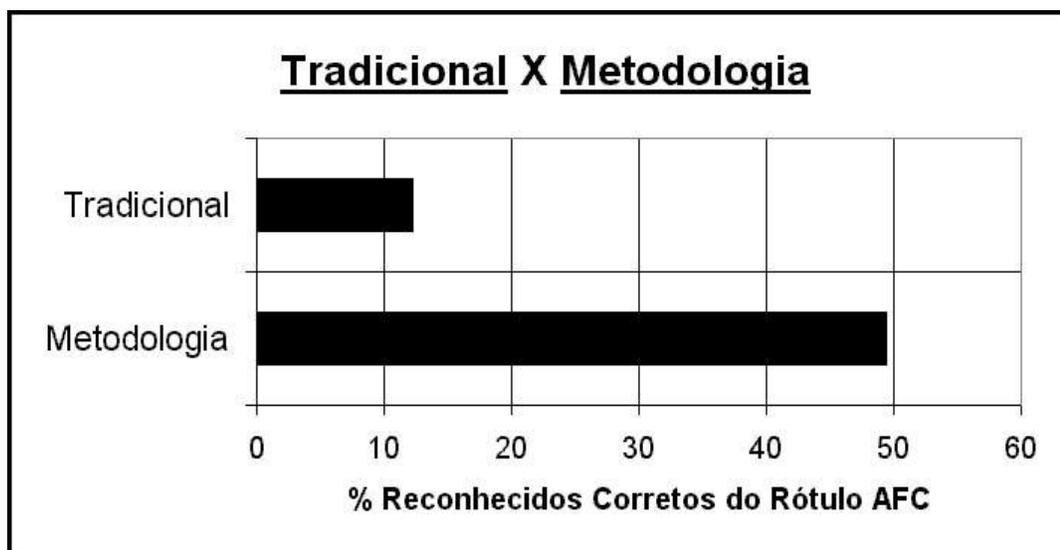


Figura 6.15 Análise comparativa entre a abordagem tradicional e o melhor resultado obtido nos testes da metodologia proposta para o rótulo AFC

6.2.3 Resultados do Rótulo NFE

Base	Rótulo	Métodos	Total-Car	%Rec-Corr	%Rec-Err	%Não-Rec
t	NFE	TRADICIONAL	1345	1,64	11,82	86,54
t	NFE	ITLL+EROc02	1345	80,15	19,85	0
t	NFE	ITLL+EROq01	1345	79,7	15,24	5,06
t	NFE	ENWU+ABEc01	1345	76,8	21,41	1,79
t	NFE	ENLI+EROc02	1345	75,39	17,7	6,91
t	NFE	ENLI+EROq01	1345	74,13	16,58	9,29
t	NFE	ENWU+ABEq01	1345	74,05	20,74	5,21
t	NFE	ENWU+EROc01	1345	71,9	21,78	6,32
t	NFE	ITLL+EROc01	1345	71,3	19,18	9,52
t	NFE	ENWU	1345	70,04	18,81	11,15
t	NFE	ENWU+ABEc02	1345	67,43	25,28	7,29
t	NFE	ITLL+ABEq01	1345	67,14	15,91	16,95
t	NFE	ENLI+EROc01	1345	66,39	20,97	12,64
t	NFE	ITLL+ABEc02	1345	65,35	23,35	11,3
t	NFE	ENWU+EROq01	1345	65,28	24,61	10,11
t	NFE	OTSU+EROc02	1345	64,01	22,53	13,46
t	NFE	OTSU+EROq01	1345	60,67	16,95	22,38
t	NFE	ITLL+ABEc01	1345	59,93	27,73	12,34
t	NFE	ENLI+ABEq01	1345	55,24	21,04	23,72
t	NFE	OTSU+EROc01	1345	54,42	39,48	6,1
t	NFE	ENLI+ABEc02	1345	54,35	29,52	16,13
t	NFE	ENWU+EROc02	1345	51,6	30,19	18,21
t	NFE	ENLI+ABEc01	1345	47,29	45,35	7,36
t	NFE	ITLL	1345	45,87	32,79	21,34
t	NFE	OTSU+ABEq01	1345	44,01	27,58	28,41
t	NFE	OTSU+ABEc02	1345	40,45	33,9	25,65
t	NFE	ENLI	1345	31,82	37,55	30,63
t	NFE	OTSU+ABEc01	1345	30,78	45,43	23,79
t	NFE	OTSU	1345	21,19	41,78	37,03

Tabela 6.5 Resultados de teste da limiarização combinada com morfologia binária do rótulo NFE

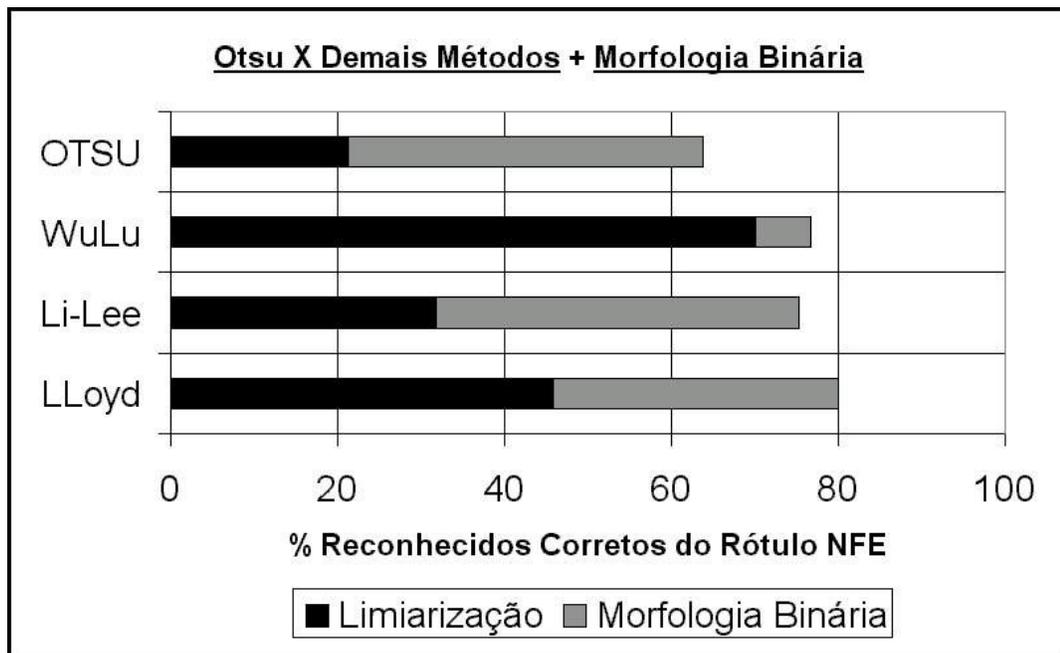


Figura 6.16 Análise comparativa dos resultados de teste para o rótulo NFE: Método de Otsu versus WuLu, Li-Lee e LLoyd; acrescidos dos melhores resultados da morfologia matemática binária

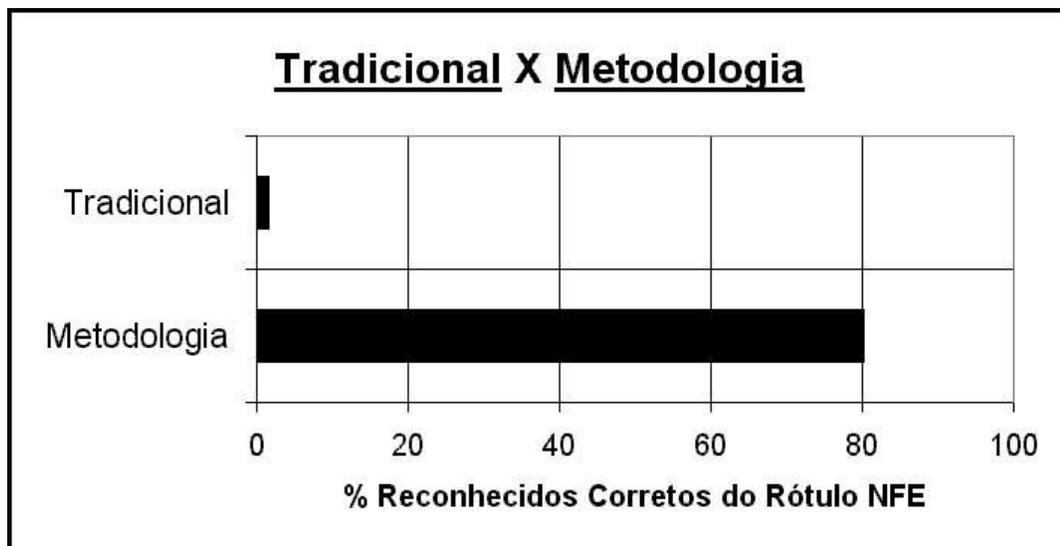


Figura 6.17 Análise comparativa entre a abordagem tradicional e o melhor resultado obtido nos testes da metodologia proposta para o rótulo NFE

6.2.4 Resultados do Rótulo NFC

Base	Rótulo	Métodos	Total-Car	%Rec-Corr	%Rec-Err	%Não-Rec
t	NFC	TRADICIONAL	1479	14,6	29,61	55,79
t	NFC	ITLL+EROc02	1479	78,57	20,01	1,42
t	NFC	ITLL+EROq01	1479	73,77	24,14	2,09
t	NFC	ENLI+EROc02	1479	73,63	23,73	2,64
t	NFC	ENWU	1479	71,26	25,02	3,72
t	NFC	ENLI+EROq01	1479	70,52	27,11	2,37
t	NFC	ENWU+ABEc01	1479	70,45	29,55	0
t	NFC	ITLL+EROc01	1479	67,75	27,92	4,33
t	NFC	ITLL+ABEq01	1479	67	29,68	3,32
t	NFC	ENWU+ABEq01	1479	64,71	30,22	5,07
t	NFC	ENWU+EROc01	1479	63,49	26,03	10,48
t	NFC	ENLI+EROc01	1479	60,45	30,09	9,46
t	NFC	ITLL+ABEc02	1479	59,5	34,08	6,42
t	NFC	OTSU+EROc02	1479	59,23	25,49	15,28
t	NFC	OTSU+EROq01	1479	57,67	29,82	12,51
t	NFC	ENWU+ABEc02	1479	56,52	34,35	9,13
t	NFC	ENLI+ABEq01	1479	55,65	33,87	10,48
t	NFC	ITLL+ABEc01	1479	54,16	37,93	7,91
t	NFC	ENWU+EROq01	1479	53,68	30,16	16,16
t	NFC	ENLI+ABEc02	1479	53,21	33,47	13,32
t	NFC	ENLI+ABEc01	1479	49,63	35,97	14,4
t	NFC	OTSU+EROc01	1479	44,9	32,66	22,44
t	NFC	OTSU+ABEq01	1479	43,95	35,5	20,55
t	NFC	ITLL	1479	42,46	33,4	24,14
t	NFC	OTSU+ABEc02	1479	42,39	37,12	20,49
t	NFC	ENWU+EROc02	1479	37,66	29,14	33,2
t	NFC	ENLI	1479	36,71	38,54	24,75
t	NFC	OTSU+ABEc01	1479	31,51	33,81	34,68
t	NFC	OTSU	1479	23,26	34,42	42,32

Tabela 6.6 Resultados de teste da limiarização combinada com morfologia binária do rótulo NFC

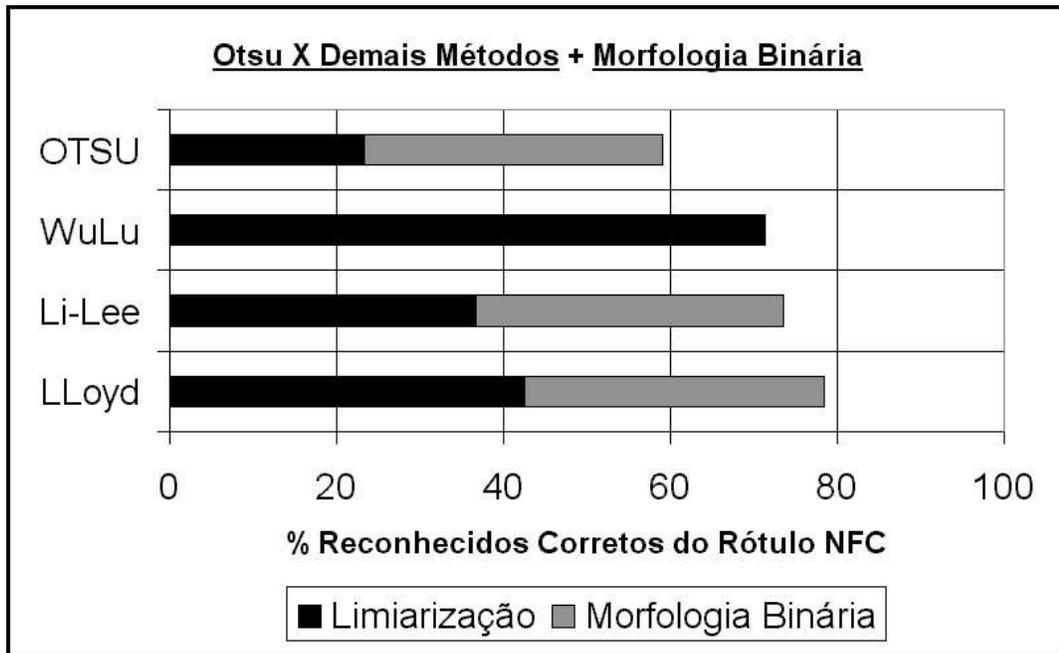


Figura 6.18 Análise comparativa dos resultados de teste para o rótulo NFC: Método de Otsu versus WuLu, Li-Lee e LLoyd; acrescidos dos melhores resultados da morfologia matemática binária

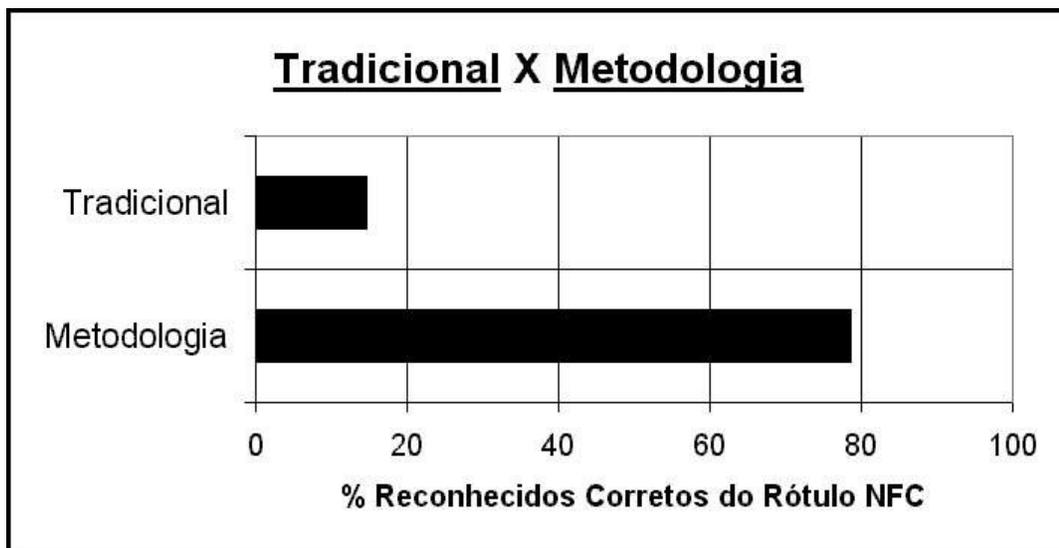


Figura 6.19 Análise comparativa entre a abordagem tradicional e o melhor resultado obtido nos testes da metodologia proposta para o rótulo NFC

6.2.5 Resultados do Rótulo PFC

Base	Rótulo	Métodos	Total-Car	%Rec-Corr	%Rec-Err	%Não-Rec
t	PFC	TRADICIONAL	3350	34,84	11,25	53,91
t	PFC	ENWU+ABEq01	3350	92,63	7,37	0
t	PFC	ENWU+EROq01	3350	92,63	7,37	0
t	PFC	ITLL+ABEc02	3350	92,48	7,52	0
t	PFC	ITLL+ABEq01	3350	92,27	7,73	0
t	PFC	ENWU+EROc01	3350	92,12	7,76	0,12
t	PFC	ENLI+ABEc02	3350	91,88	8,12	0
t	PFC	ENWU+ABEc02	3350	91,7	8,3	0
t	PFC	ITLL+EROq01	3350	91,7	7,76	0,54
t	PFC	ENLI+ABEq01	3350	91,61	8,39	0
t	PFC	ITLL+EROc01	3350	91,55	8,45	0
t	PFC	OTSU+ABEc02	3350	91,46	8,54	0
t	PFC	OTSU+ABEq01	3350	91,43	8,57	0
t	PFC	ENLI+EROq01	3350	91,31	7,79	0,9
t	PFC	ENLI+EROc01	3350	90,84	9,16	0
t	PFC	OTSU+EROq01	3350	90,72	8,54	0,74
t	PFC	OTSU+EROc01	3350	90,54	9,46	0
t	PFC	ENWU+ABEc01	3350	90,21	9,79	0
t	PFC	ITLL+ABEc01	3350	89,73	10,27	0
t	PFC	OTSU+EROc02	3350	88,84	10,75	0,41
t	PFC	ENLI+EROc02	3350	88,21	11,4	0,39
t	PFC	ENLI+ABEc01	3350	87,73	12,27	0
t	PFC	ENWU	3350	87,55	12,45	0
t	PFC	OTSU+ABEc01	3350	87,19	12,81	0
t	PFC	ITLL+EROc02	3350	87,1	12,3	0,6
t	PFC	ITLL	3350	84,54	15,46	0
t	PFC	ENLI	3350	81,94	18	0,06
t	PFC	ENWU+EROc02	3350	81,58	18,33	0,09
t	PFC	OTSU	3350	80,15	19,79	0,06

Tabela 6.7 Resultados de teste da limiarização combinada com morfologia binária do rótulo PFC

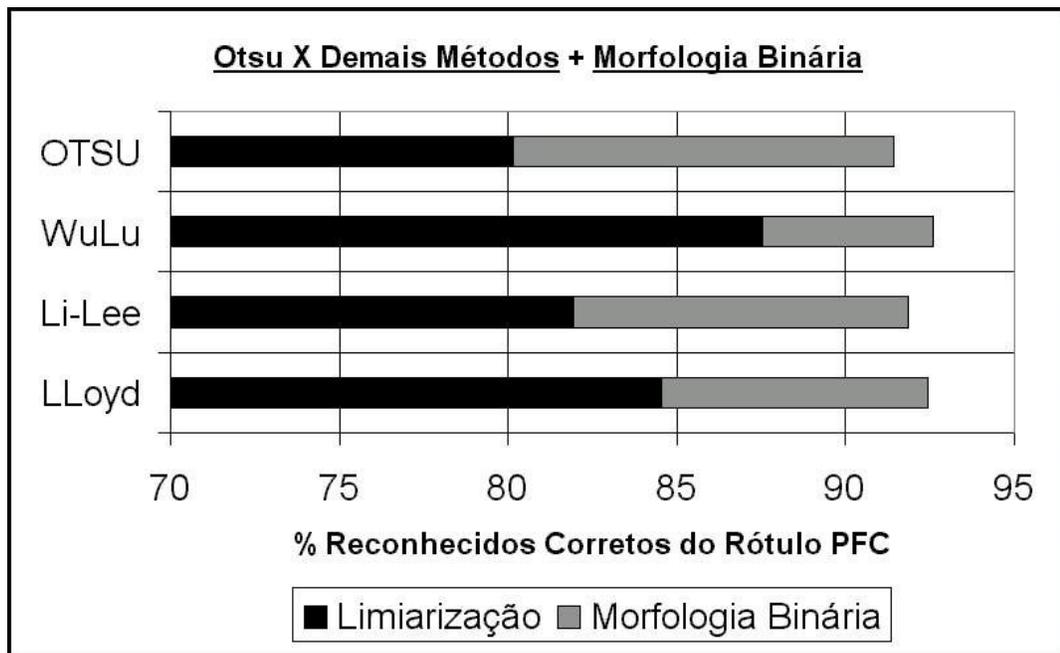


Figura 6.20 Análise comparativa dos resultados de teste para o rótulo PFC: Método de Otsu versus WuLu, Li-Lee e LLoyd; acrescidos dos melhores resultados da morfologia matemática binária

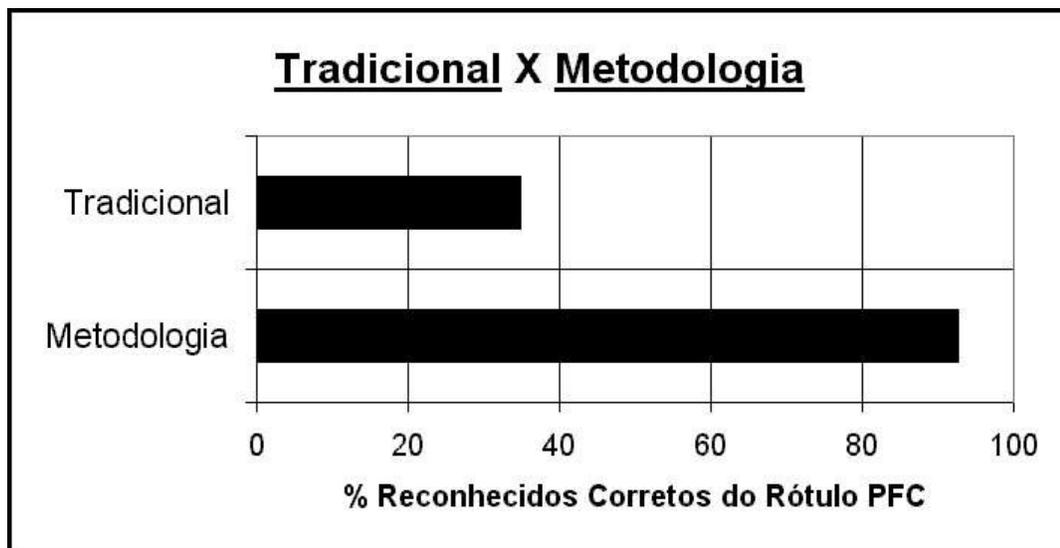


Figura 6.21 Análise comparativa entre a abordagem tradicional e o melhor resultado obtido nos testes da metodologia proposta para o rótulo PFC

Capítulo 7

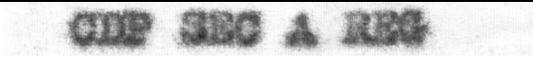
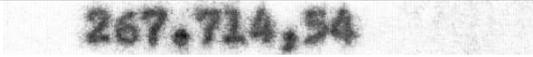
Validação e Análise dos Resultados

7.1 Resultados da Validação

Os resultados da validação estão apresentados nas páginas seguintes, onde para cada rótulo é mostrada uma tabela a partir da qual foram gerados 2 gráficos, conforme:

Seção	Rótulo	Tabela	Gráficos
7.1.1	AFE	7.1	7.1 e 7.2
7.1.2	AFC	7.2	7.3 e 7.4
7.1.3	NFE	7.3	7.5 e 7.6
7.1.4	NFC	7.4	7.7 e 7.8
7.1.5	PFC	7.5	7.9 e 7.10

Conforme foi detalhado no capítulo 2, os rótulos têm as seguintes principais características:

Rótulo	Imagem com conteúdo impresso	Características da região
AFE		Impresso alfabético fundo escuro
AFC		Impresso alfabético fundo claro
NFE		Impresso numérico fundo escuro
NFC		Impresso numérico fundo claro
PFC		Pré-impresso alfabét. fundo claro

Optou-se mostrar nas tabelas todos os resultados obtidos para o rótulo, por conterem importantes informações para a análise apresentada na seção 7.2.

A primeira linha de dados de cada tabela corresponde aos resultados obtidos com a base de validação utilizando-se a abordagem tradicional em soluções de GED. Para isso, os documentos foram capturados em preto e branco, rotulados, reconhecidos e avaliada a qualidade de cada rótulo, sem se utilizar técnicas de processamento de imagem.

A partir da segunda linha, para facilitar a análise, os dados estão classificados em ordem decrescente de reconhecidos corretos (%Rec-Corr).

Os dados de ENLI, ENWU, ITLL e OTSU correspondem aos resultados das respectivas limiarizações, sem aplicação de morfologia matemática binária.

Os gráficos, gerados a partir dos reconhecidos corretos, têm como principal objetivo facilitar a verificação de prova das 3 hipóteses anteriormente formuladas:

- **Primeira:** verifica-se que a abordagem tradicional mostrada no gráfico “**Tradicional X Metodologia**” tem resultados significativamente menores que cada um dos resultados dos métodos de limiarização global mostrados no gráfico “**Otsu X Demais métodos + Morfologia Binária**”.
- **Segunda:** verifica-se que no gráfico “**Otsu X Demais métodos + Morfologia Binária**” os resultados de WuLu, Li-Lee e LLoyd são muito melhores que os de Otsu.
- **Terceira:** verifica-se no gráfico “**Otsu X Demais métodos + Morfologia Binária**” que os resultados da morfologia matemática binária, na maioria dos casos, apresentam significativo acréscimo na melhoria da qualidade.

7.1.1 Resultados do Rótulo AFE

Base	Rótulo	Métodos	Total-Car	%Rec-Corr	%Rec-Err	%Não-Rec
v	AFE	TRADICIONAL	3591	1,03	12,53	86,44
v	AFE	ITLL+EROq01	3591	43,8	37,43	18,77
v	AFE	ENLI+EROc02	3591	42,47	39,38	18,15
v	AFE	ENLI+EROq01	3591	41,99	37,87	20,14
v	AFE	ITLL+EROc02	3591	41,91	45,59	12,5
v	AFE	ITLL+EROc01	3591	40,6	38,04	21,36
v	AFE	OTSU+EROc02	3591	37,87	40,88	21,25
v	AFE	ITLL+ABEq01	3591	37,34	39,54	23,12
v	AFE	ENLI+EROc01	3591	36,2	36,98	26,82
v	AFE	OTSU+EROq01	3591	36,17	39,96	23,87
v	AFE	ENWU+ABEc01	3591	35,23	43,19	21,58
v	AFE	ENWU	3591	33,72	35,56	30,72
v	AFE	ITLL+ABEc02	3591	32,41	42,24	25,35
v	AFE	ENWU+ABEq01	3591	31,61	46,23	22,16
v	AFE	ENWU+EROc01	3591	31,47	43,78	24,75
v	AFE	ENLI+ABEq01	3591	30,27	40,85	28,88
v	AFE	ITLL+ABEc01	3591	29,96	40,07	29,97
v	AFE	OTSU+EROc01	3591	29,04	39,29	31,67
v	AFE	ENWU+ABEc02	3591	26,23	53,49	20,28
v	AFE	ENLI+ABEc02	3591	26,15	43,33	30,52
v	AFE	ENWU+EROq01	3591	23,67	45,45	30,88
v	AFE	OTSU+ABEq01	3591	22,61	43,47	33,92
v	AFE	ENLI+ABEc01	3591	21,55	40,16	38,29
v	AFE	ITLL	3591	20,25	35,48	44,27
v	AFE	OTSU+ABEc02	3591	19,24	44,5	36,26
v	AFE	OTSU+ABEc01	3591	17,24	42,72	40,04
v	AFE	ENWU+EROc02	3591	15,32	35,59	49,09
v	AFE	ENLI	3591	12,84	35,31	51,85
v	AFE	OTSU	3591	8,08	33,5	58,42

Tabela 7.1 Resultados de validação da limiarização combinada com morfologia binária do rótulo AFE

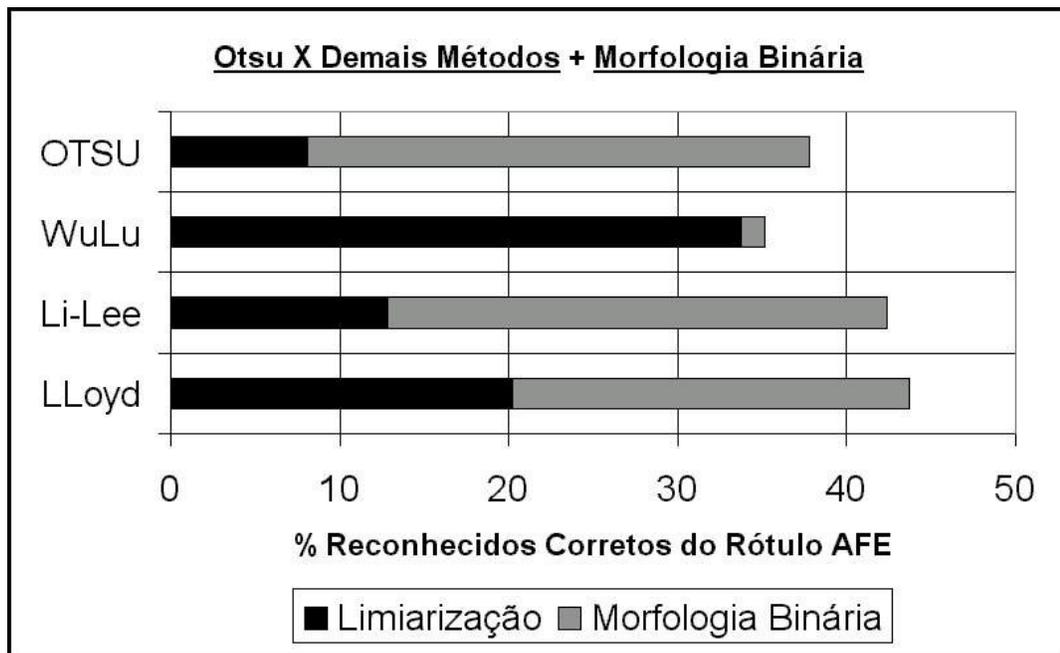


Figura 7.1 Análise comparativa dos resultados da validação para o rótulo AFE: Método de Otsu versus WuLu, Li-Lee e LLoyd; acrescidos dos melhores resultados da morfologia matemática binária

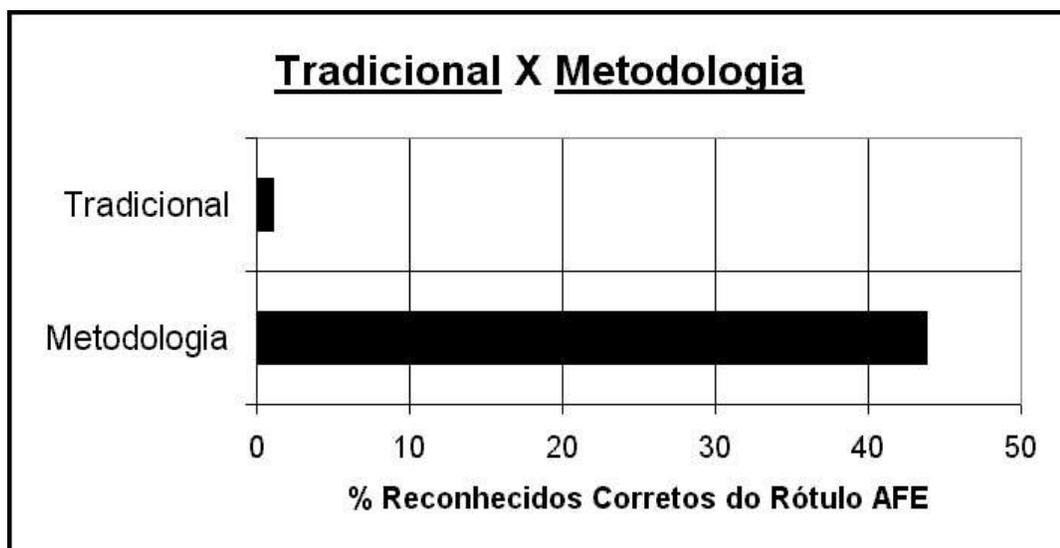


Figura 7.2 Análise comparativa entre a abordagem tradicional e o melhor resultado obtido na validação da metodologia proposta para o rótulo AFE

7.1.2 Resultados do Rótulo AFC

Base	Rótulo	Métodos	Total-Car	%Rec-Corr	%Rec-Err	%Não-Rec
v	AFC	TRADICIONAL	2823	11,73	48,53	39,74
v	AFC	ITLL+EROq01	2823	50,66	44,7	4,64
v	AFC	ENLI+EROc02	2823	48,81	42,97	8,22
v	AFC	ENLI+EROq01	2823	48,46	41,37	10,17
v	AFC	ITLL+EROc02	2823	47,33	46,72	5,95
v	AFC	ITLL+EROc01	2823	46,3	43,43	10,27
v	AFC	OTSU+EROc02	2823	41,94	47,08	10,98
v	AFC	ITLL+ABEq01	2823	41,87	42,86	15,27
v	AFC	OTSU+EROq01	2823	38,33	45,13	16,54
v	AFC	ENLI+EROc01	2823	37,87	44,56	17,57
v	AFC	ITLL+ABEc02	2823	37,27	47,18	15,55
v	AFC	ITLL+ABEc01	2823	35,32	47,54	17,14
v	AFC	ENLI+ABEq01	2823	31,56	48,32	20,12
v	AFC	OTSU+EROc01	2823	29,97	47,72	22,31
v	AFC	ENLI+ABEc02	2823	28,94	53,17	17,89
v	AFC	ITLL	2823	26,67	52,07	21,26
v	AFC	ENLI+ABEc01	2823	24,83	54,37	20,8
v	AFC	OTSU+ABEq01	2823	24,34	53,38	22,28
v	AFC	ENLI	2823	19,27	54,66	26,07
v	AFC	OTSU+ABEc02	2823	18,81	56,43	24,76
v	AFC	OTSU+ABEc01	2823	17,75	55,58	26,67
v	AFC	ENWU	2823	16,37	46,9	36,73
v	AFC	OTSU	2823	11,73	54,2	34,07
v	AFC	ENWU+ABEc01	2823	10,66	51,79	37,55
v	AFC	ENWU+ABEq01	2823	5,92	27,91	66,17
v	AFC	ENWU+EROc01	2823	4,57	27,13	68,3
v	AFC	ENWU+ABEc02	2823	4,07	40,95	54,98
v	AFC	ENWU+EROq01	2823	2,87	23,17	73,96
v	AFC	ENWU+EROc02	2823	0,96	13,89	85,15

Tabela 7.2 Resultados de validação da limiarização combinada com morfologia binária do rótulo AFC

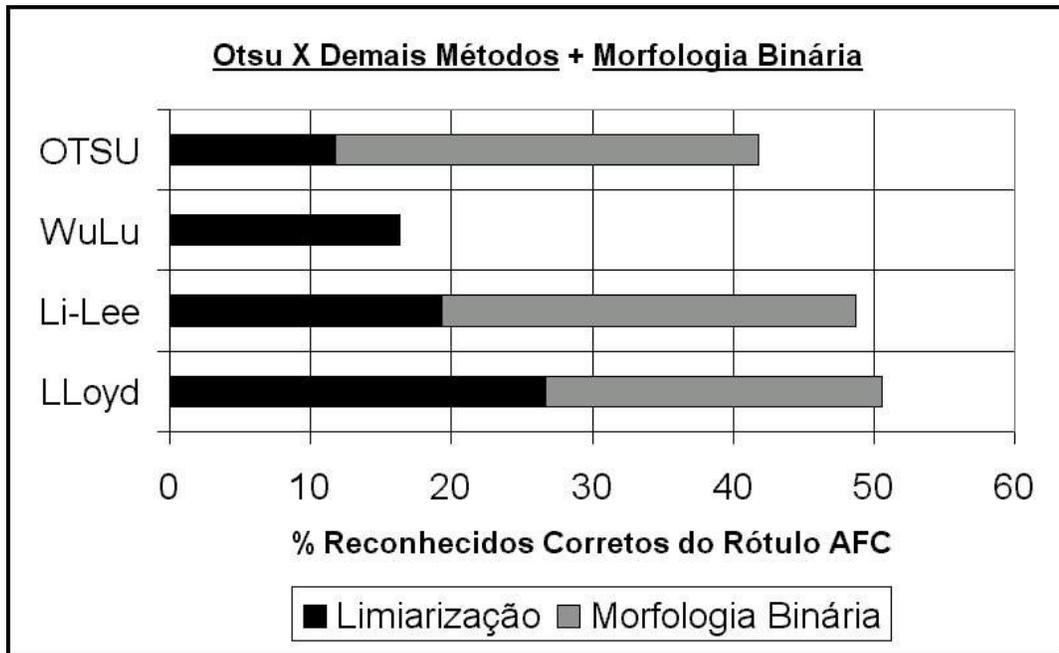


Figura 7.3 Análise comparativa dos resultados da validação para o rótulo AFC: Método de Otsu versus WuLu, Li-Lee e LLoyd; acrescidos dos melhores resultados da morfologia matemática binária

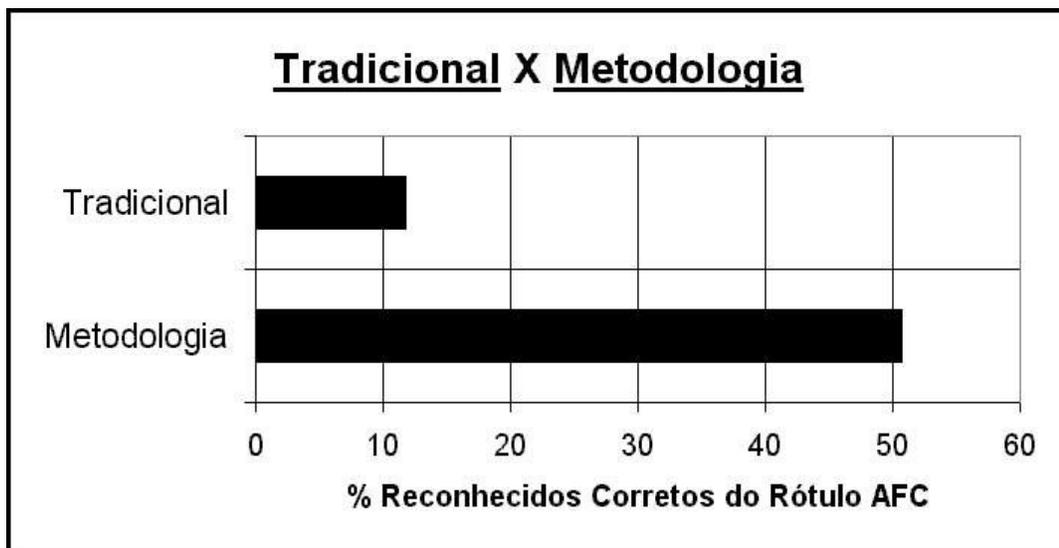


Figura 7.4 Análise comparativa entre a abordagem tradicional e o melhor resultado obtido na validação da metodologia proposta para o rótulo AFC

7.1.3 Resultados do Rótulo NFE

Base	Rótulo	Métodos	Total-Car	%Rec-Corr	%Rec-Err	%Não-Rec
v	NFE	TRADICIONAL	2682	3,13	18,61	78,26
v	NFE	ITLL+EROq01	2682	77,29	16,41	6,3
v	NFE	ENLI+EROc02	2682	75,8	17,34	6,86
v	NFE	ITLL+EROc02	2682	75,32	18,08	6,6
v	NFE	ITLL+EROc01	2682	72,74	19,95	7,31
v	NFE	ENWU+ABEc01	2682	72,52	20,1	7,38
v	NFE	ENWU+ABEq01	2682	72,26	22,63	5,11
v	NFE	ENLI+EROq01	2682	71,36	18,49	10,15
v	NFE	ENWU	2682	70,99	21,33	7,68
v	NFE	ENWU+EROc01	2682	68,16	25,39	6,45
v	NFE	ITLL+ABEq01	2682	66,37	18,61	15,02
v	NFE	ENWU+EROq01	2682	65,96	24,98	9,06
v	NFE	ENWU+ABEc02	2682	65,88	25,17	8,95
v	NFE	OTSU+EROc02	2682	65,81	22,74	11,45
v	NFE	ENLI+EROc01	2682	63,5	27,52	8,98
v	NFE	ITLL+ABEc02	2682	61,63	24,87	13,5
v	NFE	OTSU+EROq01	2682	61,26	19,39	19,35
v	NFE	ITLL+ABEc01	2682	58,17	33,52	8,31
v	NFE	ENLI+ABEq01	2682	55,52	22,48	22
v	NFE	ENLI+ABEc02	2682	54,1	26,06	19,84
v	NFE	OTSU+EROc01	2682	50,3	33,82	15,88
v	NFE	ENWU+EROc02	2682	47,95	28,82	23,23
v	NFE	ITLL	2682	47,58	35,05	17,37
v	NFE	ENLI+ABEc01	2682	46,76	39,3	13,94
v	NFE	OTSU+ABEq01	2682	44,89	23,68	31,43
v	NFE	OTSU+ABEc02	2682	39,45	32,14	28,41
v	NFE	ENLI	2682	33,22	36,99	29,79
v	NFE	OTSU+ABEc01	2682	33,11	44,71	22,18
v	NFE	OTSU	2682	22,04	38,96	39

Tabela 7.3 Resultados de validação da limiarização combinada com morfologia binária do rótulo NFE

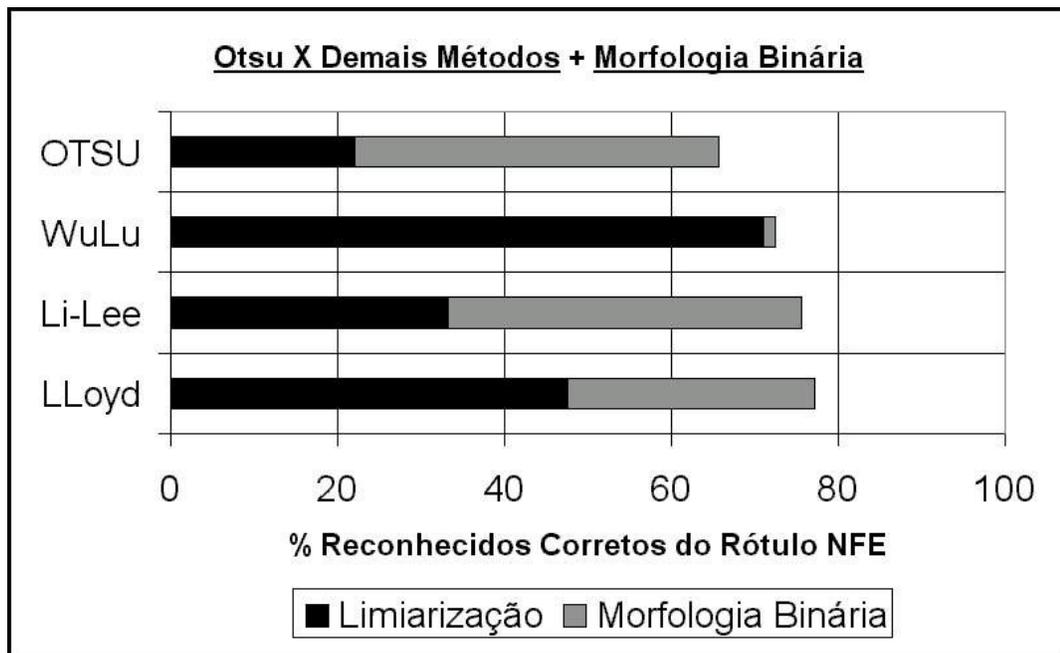


Figura 7.5 Análise comparativa dos resultados da validação para o rótulo NFE: Método de Otsu versus WuLu, Li-Lee e LLoyd; acrescidos dos melhores resultados da morfologia matemática binária

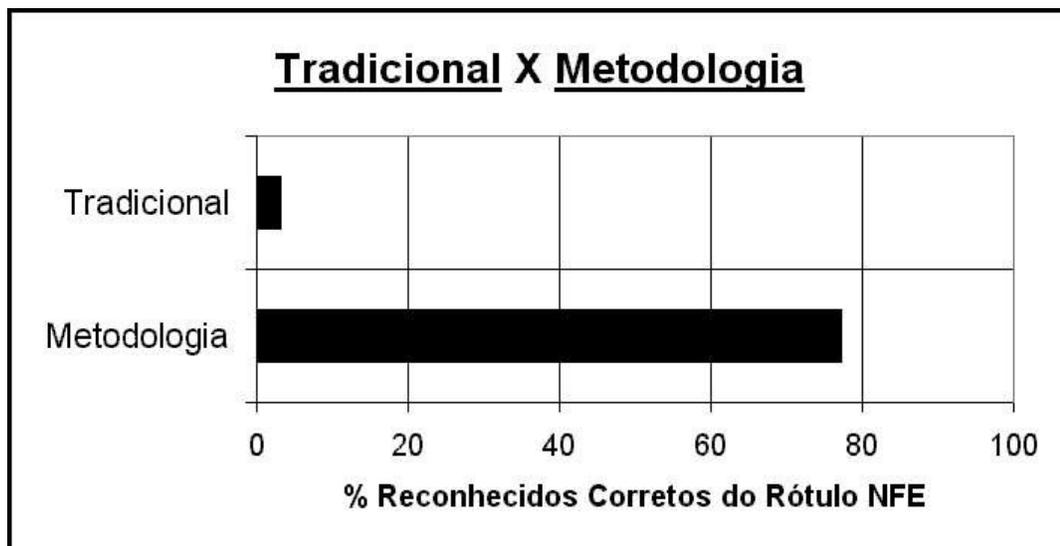


Figura 7.6 Análise comparativa entre a abordagem tradicional e o melhor resultado obtido na validação da metodologia proposta para o rótulo NFE

7.1.4 Resultados do Rótulo NFC

Base	Rótulo	Métodos	Total-Car	%Rec-Corr	%Rec-Err	%Não-Rec
v	NFC	TRADICIONAL	2964	13,39	26,79	59,82
v	NFC	ITLL+EROc02	2964	77,9	21,39	0,71
v	NFC	ENLI+EROc02	2964	76,25	21,86	1,89
v	NFC	ITLL+EROq01	2964	75,3	23,21	1,49
v	NFC	ENWU	2964	73,55	26,01	0,44
v	NFC	ENLI+EROq01	2964	72,1	23,18	4,72
v	NFC	ENWU+ABEc01	2964	70,24	27,56	2,2
v	NFC	ITLL+EROc01	2964	67,98	27,46	4,56
v	NFC	ITLL+ABEq01	2964	67	28,98	4,02
v	NFC	ENWU+ABEq01	2964	66,33	28,44	5,23
v	NFC	OTSU+EROc02	2964	64,54	27,63	7,83
v	NFC	ENLI+EROc01	2964	63,87	30,43	5,7
v	NFC	ITLL+ABEc02	2964	61,4	32,59	6,01
v	NFC	ENWU+EROc01	2964	61,2	26,52	12,28
v	NFC	ENLI+ABEq01	2964	61,1	32,12	6,78
v	NFC	OTSU+EROq01	2964	59,85	30,33	9,82
v	NFC	ENWU+ABEc02	2964	57,52	32,93	9,55
v	NFC	ENLI+ABEc02	2964	55,67	34,58	9,75
v	NFC	ITLL+ABEc01	2964	55,47	33,94	10,59
v	NFC	ENWU+EROq01	2964	51,62	28,68	19,7
v	NFC	OTSU+EROc01	2964	49,7	34,62	15,68
v	NFC	ENLI+ABEc01	2964	47,33	33,64	19,03
v	NFC	OTSU+ABEq01	2964	45,92	37,31	16,77
v	NFC	ITLL	2964	42,38	37,04	20,58
v	NFC	OTSU+ABEc02	2964	41,16	37,55	21,29
v	NFC	ENWU+EROc02	2964	39,57	30,13	30,3
v	NFC	ENLI	2964	34,41	35,63	29,96
v	NFC	OTSU+ABEc01	2964	33,87	37,62	28,51
v	NFC	OTSU	2964	22,91	34,31	42,78

Tabela 7.4 Resultados de validação da limiarização combinada com morfologia binária do rótulo NFC

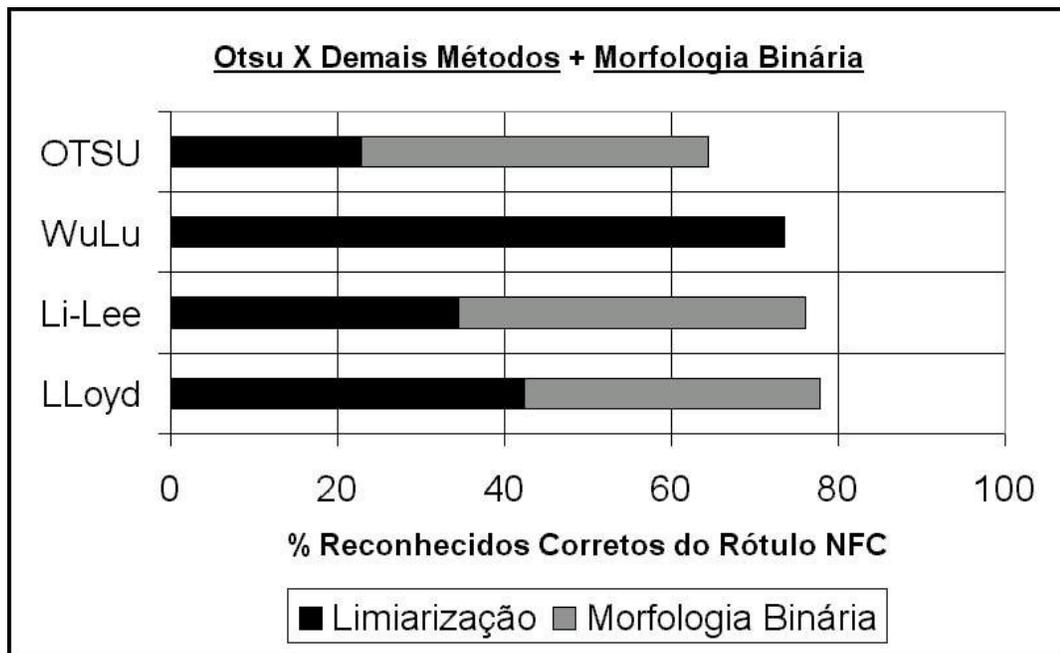


Figura 7.7 Análise comparativa dos resultados da validação para o rótulo NFC: Método de Otsu versus WuLu, Li-Lee e LLoyd; acrescidos dos melhores resultados da morfologia matemática binária

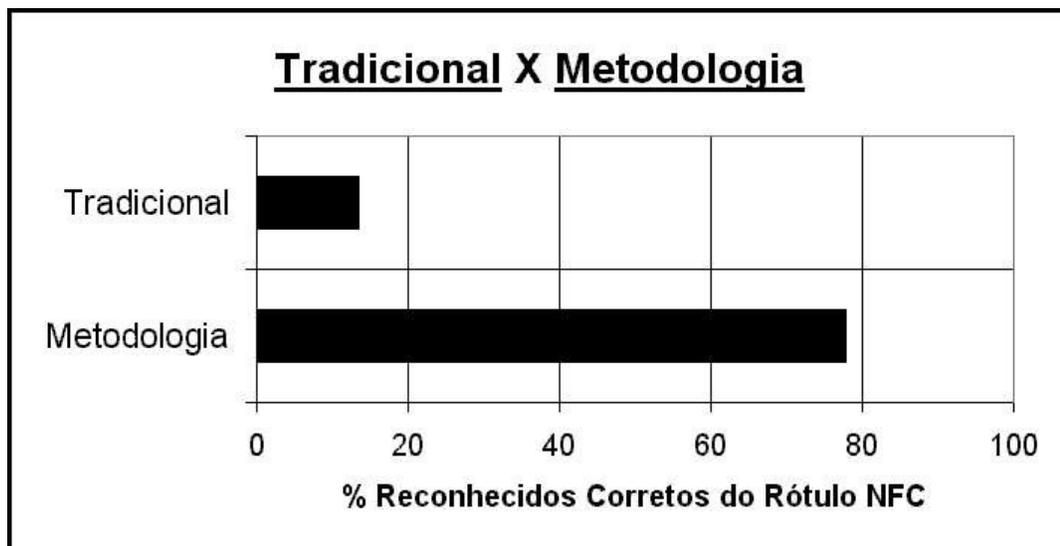


Figura 7.8 Análise comparativa entre a abordagem tradicional e o melhor resultado obtido na validação da metodologia proposta para o rótulo NFC

7.1.5 Resultados do Rótulo PFC

Base	Rótulo	Métodos	Total-Car	%Rec-Corr	%Rec-Err	%Não-Rec
v	PFC	TRADICIONAL	6700	30,99	14,13	54,88
v	PFC	ENWU+ABEq01	6700	91,84	8,16	0
v	PFC	ENWU+EROq01	6700	91,39	7,93	0,68
v	PFC	ITLL+ABEq01	6700	91,07	8,93	0
v	PFC	ITLL+ABEc02	6700	91,04	8,96	0
v	PFC	ITLL+EROq01	6700	90,94	8,28	0,78
v	PFC	ENWU+ABEc02	6700	90,66	9,25	0,09
v	PFC	ENWU+EROC01	6700	90,64	9,33	0,03
v	PFC	ENLI+ABEq01	6700	90,33	9,67	0
v	PFC	ENLI+ABEc02	6700	90,07	9,93	0
v	PFC	ENLI+EROq01	6700	90,04	9,16	0,8
v	PFC	ITLL+EROC01	6700	89,76	9,82	0,42
v	PFC	OTSU+ABEq01	6700	89,69	10,31	0
v	PFC	OTSU+EROq01	6700	89,57	9,46	0,97
v	PFC	ENLI+EROC01	6700	89,28	10,72	0
v	PFC	ENWU+ABEc01	6700	89,27	10,73	0
v	PFC	OTSU+ABEc02	6700	88,97	11,03	0
v	PFC	OTSU+EROC01	6700	88,55	11,45	0
v	PFC	ITLL+ABEc01	6700	87,66	12,34	0
v	PFC	OTSU+EROC02	6700	87,57	11,67	0,76
v	PFC	ENLI+EROC02	6700	87,45	12,1	0,45
v	PFC	ENLI+ABEc01	6700	86,1	13,9	0
v	PFC	ITLL+EROC02	6700	86,1	13,04	0,86
v	PFC	OTSU+ABEc01	6700	85,19	14,81	0
v	PFC	ENWU	6700	83,87	16,13	0
v	PFC	ITLL	6700	82	18	0
v	PFC	ENWU+EROC02	6700	81,31	17,85	0,84
v	PFC	ENLI	6700	79,99	20,01	0
v	PFC	OTSU	6700	78,13	21,79	0,08

Tabela 7.5 Resultados de validação da limiarização combinada com morfologia binária do rótulo PFC

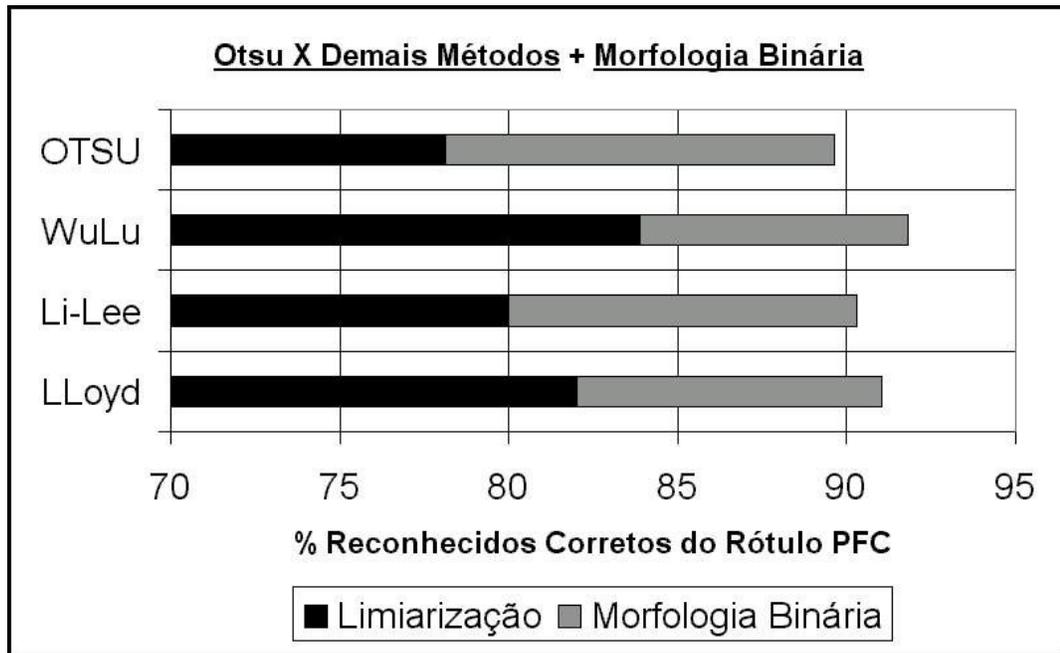


Figura 7.9 Análise comparativa dos resultados da validação para o rótulo PFC: Método de Otsu versus WuLu, Li-Lee e LLoyd; acrescidos dos melhores resultados da morfologia matemática binária

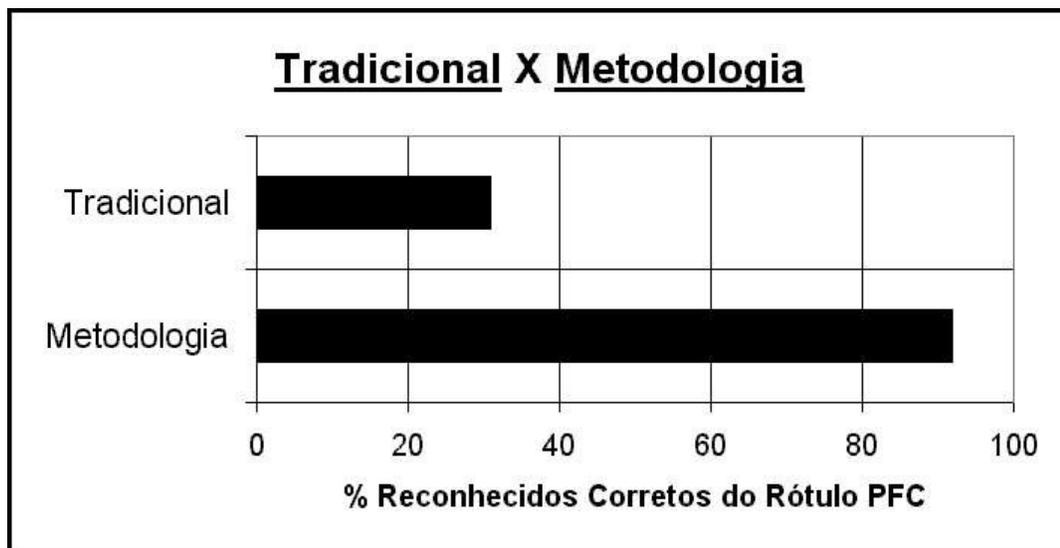


Figura 7.10 Análise comparativa entre a abordagem tradicional e o melhor resultado obtido na validação da metodologia proposta para o rótulo PFC

7.2 Análise dos Resultados

7.2.1 Introdução

A qualidade dos conteúdos impressos de imagens de documentos, é fator crítico de sucesso em soluções de GED para funcionalidades como indexação, conversão para texto e FTR. Com documentos complexos, esta qualidade deixa a desejar quando se utiliza a abordagem tradicional. Para viabilizar uma abordagem adequada, três hipóteses foram formuladas e uma metodologia foi proposta para comprovar estas hipóteses.

A metodologia proposta foi testada conforme resultados apresentados na seção 6.2 e validada conforme resultados apresentados na seção 7.1. Diferentes tipos de região de um documento complexo, conforme características apresentadas no capítulo 2, foram testadas e validadas.

A seguir apresenta-se a análise dos resultados quantitativos, a análise dos problemas encontrados e a aplicação da metodologia proposta para outros tipos de documento.

7.2.2 Análise dos Resultados Quantitativos

Esta análise inicia-se com a constatação de que os resultados quantitativos, obtidos por reconhecimento computacional, tem relação em termos de qualidade com uma avaliação visual. Como se pode perceber nas tabelas 7.6 e 7.7, o aspecto visual das imagens evolui em termos de qualidade de acordo com o percentual de caracteres reconhecidos corretos.

Às vezes, visualmente a diferença pode parecer mínima, mas em termos de taxa de reconhecimento automático ela pode ser grande. Na tabela 7.7, pode-se observar que entre o método de Otsu e o método de WuLu, o aspecto visual da imagem mudou pouco, porém, neste caso, a taxa de reconhecimento teve um acréscimo de 166%.

Processamento	Resultado Visual	Reconhecido	Corretos	Errados	N-Rec
Tradicional	431.368,08	A52.360900	40%	60%	00%
Limiarização Otsu	431.368,08	451.368908	80%	20%	00%
Limiarizaç. WuLu	431.368,08	431.368908	90%	10%	00%
LLoyd+EROc02	431.368,08	431.368,08	100%	00%	00%

Tabela 7.6 Exemplo de resultado visual conforme as técnicas aplicadas, onde existe uma relação direta crescente entre a qualidade visual e as taxas de reconhecidos corretos para o rótulo NFC

Processamento	Aspecto visual das imagens processadas	RecCorr
Tradicional	CURITIBA/AC. JOÃO NEGRÃO PS IPE	00,00%
Limiarização Otsu	CURITIBA/AC. JOÃO NEGRÃO PS IPE	11,11%
Limiarizaç. WuLu	CURITIBA/AC. JOÃO NEGRÃO PS IPE	29,63%
WuLu+EROq01	CURITIBA/AC. JOÃO NEGRÃO PS IPE	77,78%

Tabela 7.7 Aspecto visual comparado com a taxa de reconhecidos corretos, onde da limiarização por Otsu para WuLu existe pouca diferença visual mas grande diferença na taxa de reconhecimento

A análise mais importante dos resultados quantitativos, tem por objetivo comprovar as 3 hipóteses formuladas anteriormente:

- **Primeira:** Aplicando-se técnicas adequadas de limiarização global em regiões de documentos complexos com características particulares (rotulação), obtêm-se incremento de qualidade dos conteúdos impressos em relação a abordagem tradicional, evitando-se o uso de limiarização local adaptativa que por ser muito lenta não é apropriada para grandes volumes.
- **Segunda:** Encontram-se métodos de limiarização global para melhorar conteúdos impressos de documentos complexos que são melhores que Otsu, tanto em termos de resultados como em termos de tempo de processamento o que pode ser comprovado pela limiarização múltipla.
- **Terceira:** Aplicando-se parâmetros adequados, é possível através da morfologia matemática binária aprimorar a qualidade de conteúdos impressos de documentos complexos limiarizados.

A primeira e segunda hipóteses podem ser comprovadas pelos dados da tabela 7.8 que para facilitar a análise tem os seus dados projetados no gráfico conforme mostra a Figura 7.11.

Para comprovar a primeira hipótese, verifica-se que em nenhum rótulo, a abordagem tradicional obteve resultados superiores aos resultados obtidos com limiarização global aplicada por região (rótulos). Este resultado confirma-se nos 4 métodos de limiarização utilizados onde a maior diferença foi para o método WuLu no rótulo NFE com 67,86 pontos percentuais.

Para comprovar a segunda hipótese, basta verificar (Tab. 7.8, Fig. 7.11) que para todos os rótulos, a limiarização pelo método de Otsu apresentou um resultado menor que os outros métodos aqui destacados. Sendo a maior diferença para o método WuLu no rótulo NFC com 50,64 pontos percentuais.

Rótulo	Tradicional	Otsu	Li-Lee	Lloyd	WuLu
AFE	01,03%	08,08%	12,84%	20,25%	33,72%
AFC	11,73%	11,73%	19,27%	26,67%	16,37%
NFE	03,13%	22,04%	33,72%	47,58%	70,99%
NFC	13,39%	22,91%	34,41%	42,38%	73,55%
PFC	30,99%	78,13%	79,99%	82,00%	83,87%

Tabela 7.8 Análise comparativa por rótulo da abordagem tradicional versus método de Otsu versus demais métodos de limiarização utilizados

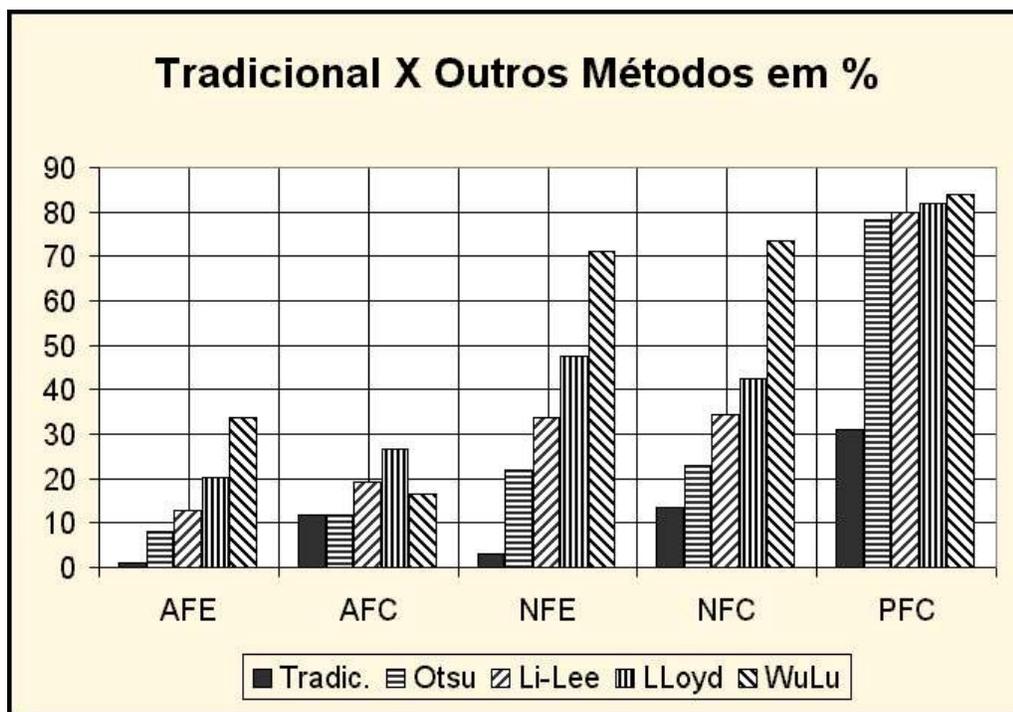


Figura 7.11 Análise comparativa por rótulo da abordagem tradicional versus método de Otsu versus demais métodos de limiarização utilizados

A terceira hipótese pode ser comprovada observando-se os resultados por rótulo da tabela 7.9, que para facilitar a análise, estão representados para cada método de limiarização de forma gráfica, conforme figuras 7.12 a 7.15. O caracter + grafado após cada método no cabeçalho da tabela 7.9, significa que as taxas de reconhecimento foram obtidas com os respectivos métodos combinados com parâmetros da morfologia matemática binária que apresentaram os melhores resultados. Pode-se observar que a morfologia binária traz um significativo incremento nas taxas de reconhecimento, na maioria dos casos, exceto para o método de WuLu nos rótulos AFC e NFC.

Rótulo	Otsu	Otsu+	Li-Lee	Li-Lee+	LLOYD	LLOYD+	WuLu	WuLu+
AFE	08,08%	37,87%	12,84%	42,47%	20,25%	43,80%	33,72%	35,23%
AFC	11,73%	41,94%	19,27%	48,81%	26,67%	50,66%	16,37%	16,37%
NFE	22,04%	65,81%	33,22%	75,80%	47,58%	77,29%	70,99%	72,52%
NFC	22,91%	64,54%	34,41%	76,25%	42,38%	77,90%	73,55%	73,55%
PFC	78,13%	89,69%	79,99%	90,33%	82,00%	91,07%	83,87%	91,84%

Tabela 7.9 Incrementos nas taxas de reconhecidos corretos com a utilização da morfologia matemática binária (melhores resultados) sobre os resultados dos respectivos métodos de limiarização

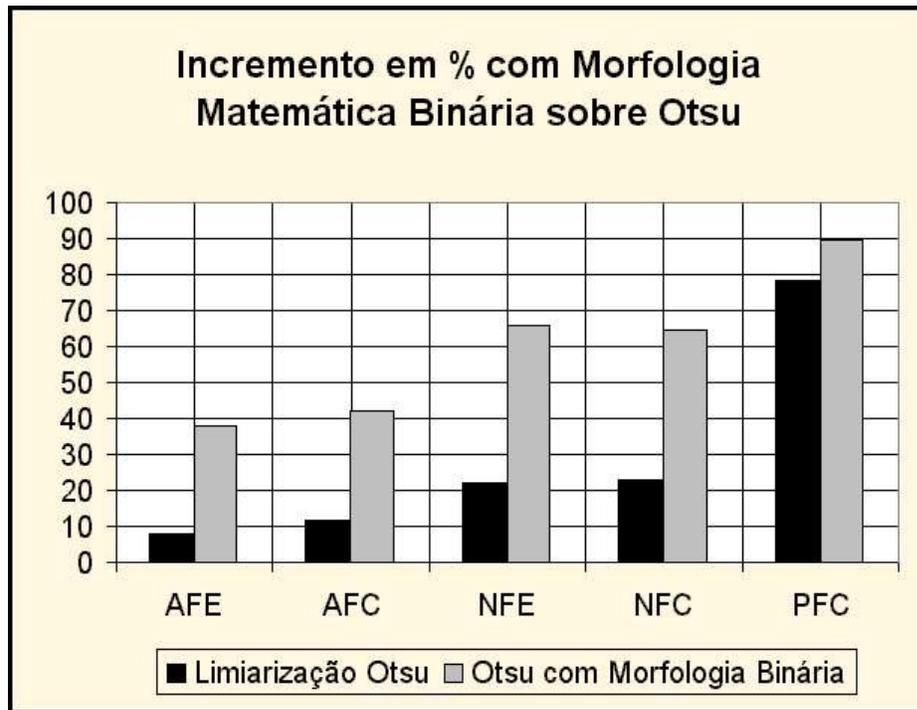


Figura 7.12 Incremento do melhor resultado da morfologia binária sobre a limiarização de Otsu

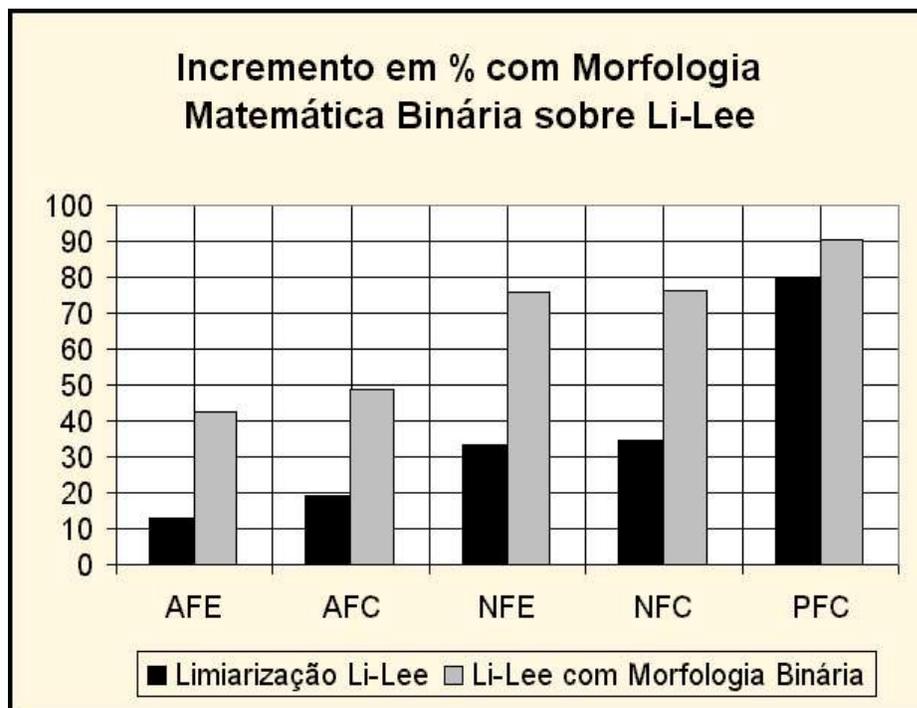


Figura 7.13 Incremento do melhor resultado da morfologia binária sobre a limiarização de Li-Lee

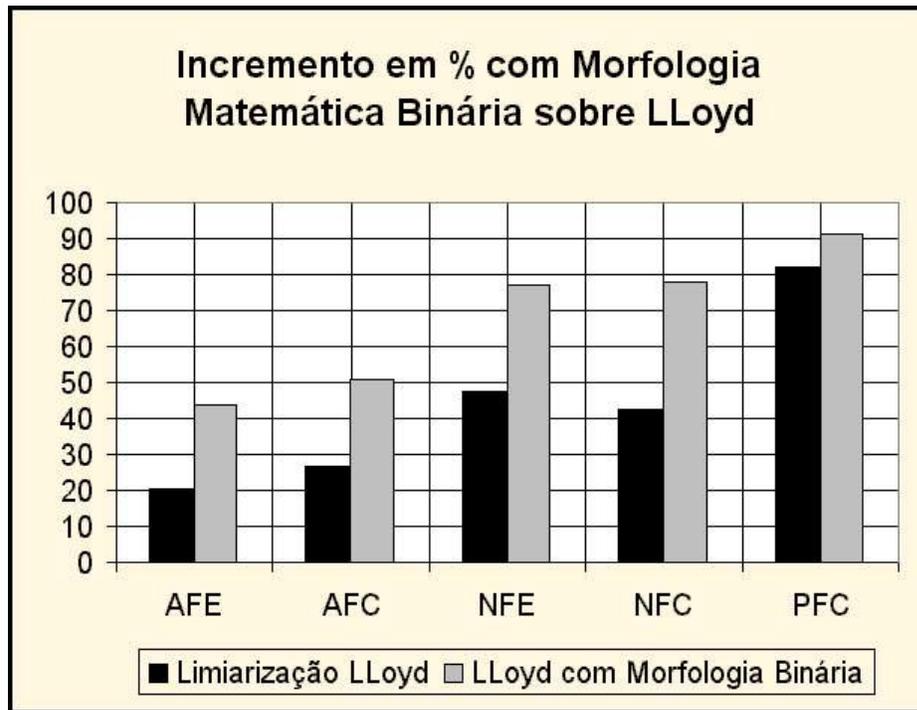


Figura 7.14 Incremento do melhor resultado da morfologia binária sobre a limiarização de LLoyd

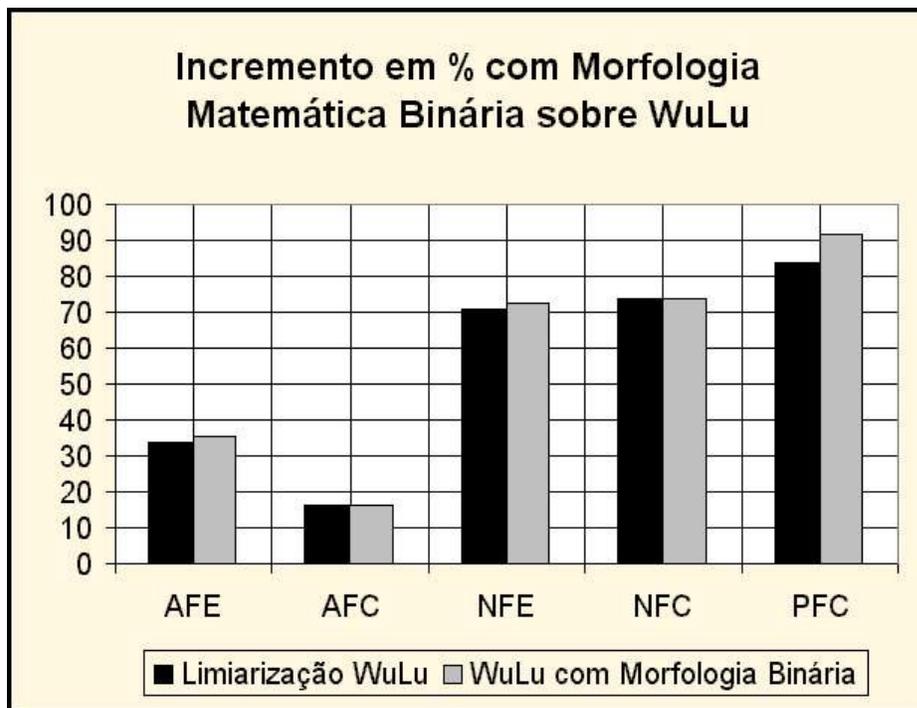


Figura 7.15 Incremento do melhor resultado da morfologia binária sobre a limiarização de WuLu

O baixo incremento da morfologia matemática binária sobre os resultados de WuLu requer uma análise complementar. O método de limiarização por entropia de WuLu [WU, 1998], foi formulado para aplicações com imagens médicas com baixo contraste como as de aparelhos de ultra-som. Por essa razão, com imagens de maior contraste o método de WuLu tende a ter efeitos mais erosivos, efeitos estes que nesta metodologia são realizados pela morfologia matemática binária. Por esse motivo, o método de WuLu, apresentou os melhores resultados em termos de limiarização, porém os menores incrementos quando combinado com a morfologia matemática binária, especialmente em regiões de alto contraste como AFC e NFC.

As duas linhas de resultado a seguir, extraídas da tabela 6.5, mostram uma particularidade que merece análise. As taxas de reconhecimento são as melhores da base de testes para o rótulo PFC e apresentam os mesmos resultados. Especialmente em situações como esta, pode haver dúvida em qual operador morfológico considerar em uma aplicação prática: abertura ou erosão.

Base	Rótulo	Métodos	Total-Car	%Rec-Corr	%Rec-Err	%Não-Rec
t	PFC	ENWU+ABEq01	3350	92,63	7,37	0
t	PFC	ENWU+EROq01	3350	92,63	7,37	0

Quando a aplicação requer somente a conversão do conteúdo da imagem em texto, descartando-se as imagens, a erosão deve ser escolhida por ser mais rápida. Quando a aplicação requer além da conversão, a preservação das imagens para futuras visualizações, a abertura pode ser necessária, porque apesar de ser mais demorada, preserva a espessura original dos caracteres.

Uma outra análise que pode ser feita sobre os resultados obtidos é uma comparação entre: Abordagem tradicional, Limiarização múltipla e Morfologia binária.

Na tabela 7.10 representada graficamente pela figura 7.16, os resultados da abordagem tradicional, são comparados com os melhores resultados obtidos com a metodologia nas fases da limiarização múltipla e da morfologia matemática binária sobre a limiarização múltipla. Dessa forma, os resultados colocam ainda mais em evidência as diferenças entre a abordagem tradicional e a metodologia proposta.

Rótulo	Abordagem Tradicional	Limiarização Múltipla	Morfologia Binária
AFE	01,03%	33,72%	43,80%
AFC	11,73%	26,67%	50,66%
NFE	03,13%	70,99%	77,29%
NFC	13,39%	73,55%	77,90%
PFC	30,99%	83,87%	91,84%

Tabela 7.10 Análise comparativa dos resultados obtidos entre a abordagem tradicional versus limiarização múltipla versus morfologia matemática binária

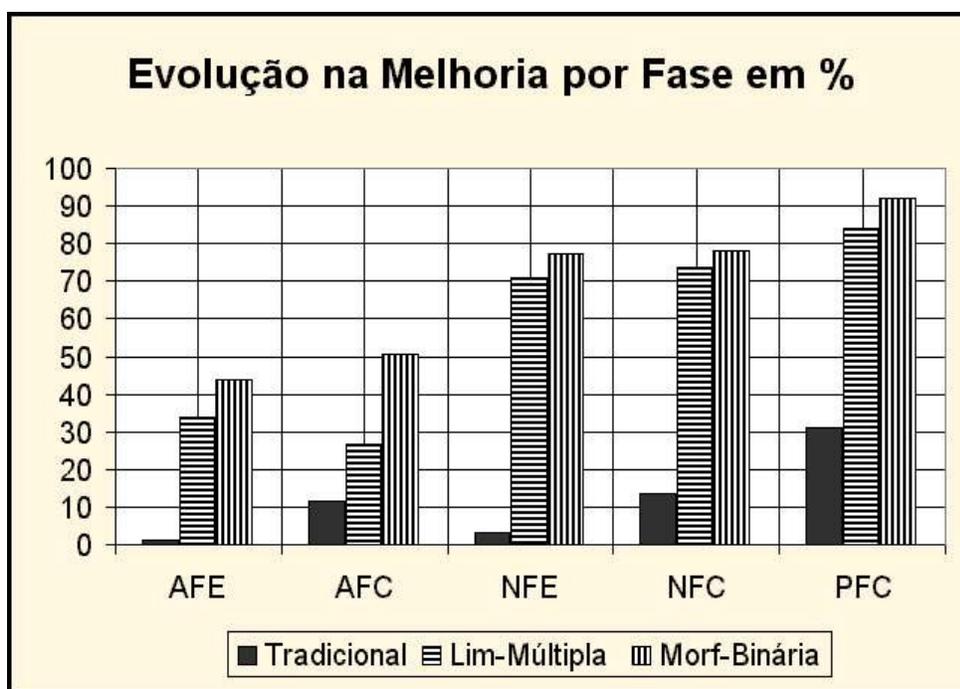


Figura 7.16 Análise comparativa dos resultados entre a abordagem tradicional, limiarização múltipla e morfologia matemática binária

A tabela 7.11 mostra as combinações de técnicas de processamento que obtiveram os melhores resultados em cada rótulo. Pode-se observar que apesar da limiarização LLoyd com erosão quadrado uma iteração prevalecer, para um mesmo documento podem ser necessárias técnicas diferentes para regiões diferentes. Esta constatação só foi possível porque a metodologia, apesar do esforço computacional, propôs testar e validar todas as combinações entre os 4 métodos de limiarização escolhidos e os parâmetros da morfologia matemática binária mais adequados.

Rótulo	Processamentos que obtiveram os melhores resultados	Rec-Corr	Tempo
AFE	Limiarização LLoyd mais erosão quadrado uma iteração	43,80%	1,05
AFC	Limiarização LLoyd mais erosão quadrado uma iteração	50,66%	1,05
NFE	Limiarização LLoyd mais erosão quadrado uma iteração	77,29%	1,05
NFC	Limiarização LLoyd mais erosão cruz uma iteração	77,90%	1,02
PFC	Limiarização WuLu mais abertura quadrado uma iteração	91,84%	1,25

Tabela 7.11 Técnicas de processamento que obtiveram os melhores resultados para cada rótulo, com taxas de reconhecidos corretos e tempos de processamento acumulados

Na tabela 7.11, a coluna tempo, representa a medida de tempo normalizada em função do tempo de Otsu (utilizado como referência neste trabalho). Para cada rótulo apresenta-se o tempo necessário para processar a limiarização combinada com os parâmetros da morfologia binária. Pode-se observar que para todos os rótulos os tempos de processamento são muito baixos. Por exemplo, para o rótulo NFC, o tempo necessário para se obter o melhor resultado, utilizando a limiarização iterativa de LLoyd com uma iteração da erosão cruz é praticamente igual ao tempo de Otsu.

Cabe ressaltar que por exemplo, para o rótulo PFC, o segundo melhor resultado é:

Rótulo	Processamentos que obtiveram os melhores resultados	Rec-Corr	Tempo
PFC	Limiarização WuLu mais erosão quadrado uma iteração	91,39%	1,07

Ou seja, um resultado com uma diferença de apenas 0,45% em relação ao melhor resultado mas com tempo de processamento significativamente menor.

Em relação ao reconhecimento, observa-se pelos resultados obtidos que, em geral, os caracteres borrados estão classificados como reconhecidos errados, o que pode ser observado no resultado “tradicional” da tabela 7.6. Os caracteres deteriorados estão, na maioria dos casos, classificados como não reconhecidos, como se pode verificar no resultado da tabela 7.12, cujos efeitos estão representados no exemplo da tabela 7.13. Aqui a deterioração ocorreu somente durante o processamento da imagem, mas em muitos casos ela também pode estar presente na imagem original.

Base	Rótulo	Métodos	Total-Car	%Rec-Corr	%Rec-Err	%Não-Rec
v	AFC	ENWU+EROc02	2823	0,96	13,89	85,15

Tabela 7.12 Resultado com baixa taxa de reconhecimento e alta taxa de caracteres não reconhecidos

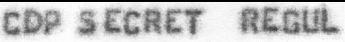
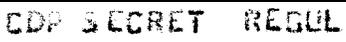
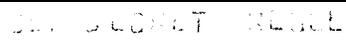
Imagem Original	Limiarização ENWU	ENWU+EROc02
		

Tabela 7.13 Efeitos da limiarização WuLu com duas iterações da erosão binária no rótulo AFC

Diante do que foi observado sobre o reconhecimento, e considerando-se que a situação ideal é ter 100% de reconhecidos corretos e 0% de não reconhecidos, os melhores resultados são aqueles que tendem para estes valores. Assim se duas abordagens para o mesmo rótulo, apresentam taxas similares para os reconhecidos corretos (tab. 7.14), a melhor abordagem será aquela com a menor taxa de não reconhecidos. Neste caso deve-se observar que se a deterioração existente na imagem original não se agravar pelas técnicas de processamento, a taxa de não reconhecidos deverá ser a mesma para as duas abordagens. Se isso ocorrer, os critérios para escolha da melhor abordagem podem ser por exemplo o tempo de processamento e/ou avaliação visual.

Base	Rótulo	Métodos	Total-Car	%Rec-Corr	%Rec-Err	%Não-Rec
v	AFC	TRADICIONAL	2823	11,73	48,53	39,74
v	AFC	OTSU	2823	11,73	54,20	34,07

Tabela 7.14 Abordagens apresentando mesma taxa de reconhecidos corretos e taxas diferentes para não reconhecidos

Para finalizar esta análise, apresenta-se uma relação de custo/benefício em termos de tempo de processamento e acréscimo na taxa de reconhecimento entre a abordagem tradicional e a metodologia proposta. Os resultados estão apresentados na tabela 7.15 onde os tempos para a abordagem tradicional foram considerados os de Otsu, uma vez que este método é muito popular nesta abordagem.

Rótulo (região)	Abordagem Tradicional	Metodologia Proposta	Acréscimo no Reconhecimento	Tempo Tradic.	Tempo Metod.	Aumento de tempo
AFE	01,03%	43,80%	42,52 vezes	1,00	1,05	0,05 vezes
AFC	11,73%	50,66%	04,31 vezes	1,00	1,05	0,05 vezes
NFE	03,13%	77,29%	24,69 vezes	1,00	1,05	0,05 vezes
NFC	13,39%	77,90%	05,81 vezes	1,00	1,02	0,02 vezes
PFC	30,99%	91,84%	02,96 vezes	1,00	1,25	0,25 vezes

Tabela 7.15 Acréscimos nas taxas de reconhecimento versus aumento de tempo de processamento, comparando a metodologia proposta com a abordagem tradicional

Pode-se observar que, no pior caso (rótulo PFC), com um acréscimo de tempo de 25% em relação ao tempo da abordagem tradicional, obteve-se quase o triplo de taxa de reconhecimento.

No melhor caso (rótulo AFE), com um acréscimo de apenas 5% no tempo de processamento em relação a abordagem tradicional, obteve-se um acréscimo na taxa de reconhecimento de 42,52 vezes.

No caso do rótulo NFC, com praticamente o mesmo tempo da abordagem tradicional, obteve-se um acréscimo na taxa de reconhecimento de quase 6 vezes.

Assim, considerando-se os resultados obtidos e as análises aqui apresentadas, pode-se afirmar que o importante incremento em termos de qualidade no reconhecimento por OCR e o reduzido tempo de processamento, definitivamente comprova que a metodologia proposta mostra-se pertinente e muito eficiente para a melhoria de conteúdos impressos de imagens de documentos complexos.

7.2.3 Análise dos Problemas Encontrados

O primeiro problema ocorreu quando se decidiu utilizar o OCR para avaliação da qualidade. Testes preliminares com documentos complexos indicaram viabilidade de uso do OCR para a avaliação da qualidade, porém alguns documentos tinham linhas com os mesmos caracteres ocupando regiões de fundo claro e escuro (fig. 7.17 A). Pelo critério de segmentação, estes caracteres seriam horizontalmente separados (fig. 7.17 B). Quanto as técnicas de processamento, não haveria nenhum problema (fig. 7.17 C), mas

para reconhecer via OCR seria inviável. Como neste trabalho, o objetivo principal foi avaliar técnicas de processamento de imagens onde o reconhecimento por OCR foi apenas um meio, optou-se por não incluir na base documentos com este problema.

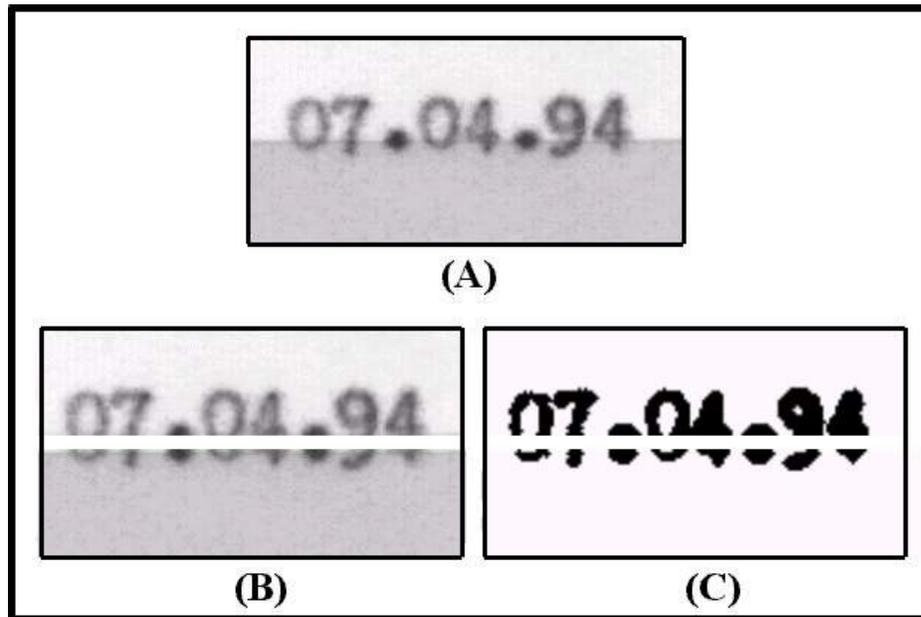


Figura 7.17 Problema para o reconhecimento via OCR: (A) imagem original, (B) imagem segmentada, (C) segmentos onde não é possível obter reconhecimento via OCR

O segundo problema ocorreu com a segmentação de alguns documentos. Como já foi mencionado, a segmentação foi efetuada de forma automática a partir de pontos de início e fim de cada região encontrados na imagem do documento. Para coincidir os pontos em todos os documentos optou-se por eliminar as margens, esquerda e superior, a partir de dados conhecidos a priori. Para eliminar a margem esquerda identificou-se a primeira borda da esquerda e para eliminar a margem superior identificou-se a primeira concentração de pixels que identifica o logotipo do banco. Para a margem esquerda não houve problema, porém alguns documentos das bases têm a margem superior muito borrada (fig. 7.18). Isto dificultou a identificação do ponto de corte. Como os casos foram poucos, a operação para eliminar as margens destes documentos foi realizada manualmente. A solução para esse tipo de problema, pode ser através de uma segmentação baseada na localização do logotipo. Para localizar o logotipo, Jesus [JESUS, 1999] propõe a utilização de uma abordagem morfológica para segmentação de logotipos em cheques bancários brasileiros sem conhecimento a priori.

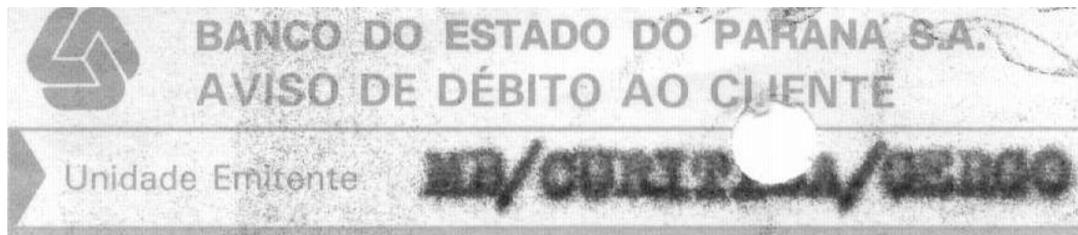


Figura 7.18 Parte superior esquerda de uma imagem da base, recortada manualmente em função da dificuldade de encontrar o ponto de corte na margem superior por excesso de ruído

O terceiro problema foi encontrado, analisando-se visualmente os resultados das imagens processadas. Em alguns casos, apesar do critério de segmentação utilizado, encontram-se na região segmentada contrastes entre conteúdos impressos (fig. 7.19 A). Neste caso o conjunto de caracteres “050-7” é mais “fraco” que o conjunto “138-4” e assim técnicas de processamento diferentes podem ser mais indicadas para cada conjunto. Conforme mostra a figura 7.19(B), para o conjunto de caracteres “138-4” um melhor resultado pode ser obtido com limiarização por entropia de WuLu com erosão enquanto que no conjunto “050-7” (fig. 7.19 C) o melhor resultado pode ser através da limiarização pela entropia de Li-Lee com abertura binária.

Em uma aplicação onde depois da conversão em texto as imagens são descartadas, isto não apresenta problema, uma vez que, neste caso, o interesse é tão somente na melhor taxa de reconhecimento. Mas quando a imagem deve ser preservada, como no caso do FTR, cuidados adicionais devem ser tomados porque com resultados como por exemplo o da figura 7.19(B) onde a erosão foi utilizada, existe deterioração do conteúdo da imagem, embora em geral isto se reflita na taxa de reconhecidos corretos versus taxa de não reconhecidos conforme pode ser confirmado na tabela 7.12.

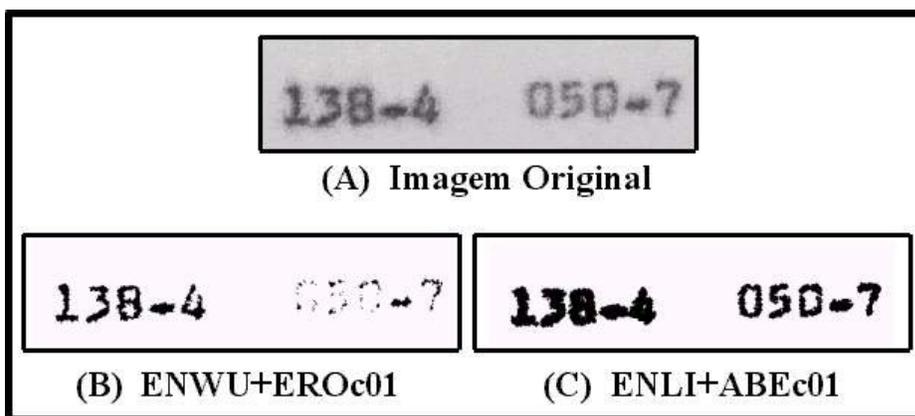


Figura 7.19 Problema de contraste no conteúdo impresso: (A) original, (B) com ENWU+EROC01 deteriora a parte “fraca”, (C) com ENLI+ABEC01 deixa borrada a parte “forte”

7.2.4 Aplicação da Metodologia em Outros Tipos de Documento

Pelos importantes resultados obtidos com a metodologia e sua facilidade de aplicação, acredita-se que o desenvolvimento de uma nova abordagem pode ser útil para diversos tipos de documento. Como os documentos das bases utilizadas neste trabalho são típicos para FTR em soluções de GED, testes foram realizados com 2 outros tipos de documentos. O primeiro é um documento simples sobre legislação que depois de ter seu conteúdo convertido para texto a imagem e descartada. O segundo é uma região de documento complexo com um valor a partir do qual deve ser gerado um conjunto de caracteres como texto para indexação. Os dois tipos diferentes de documentos são apresentados na tabela 7.16, na coluna “imagem original”.

Imagem Original	Abordagem Tradicional	Metodologia Proposta	Increment.
Art. 10. Os bens imóveis salvo, e mediante lei, se	Art. 10. Os bens imóveis salvo, e mediante lei, se 99,5% Reconhecimento	Art. 10. Os bens imóveis salvo, e mediante lei, se 99,8% Reconhecimento	0,3%
8.236,25	8.236,25 0,0% Reconhecimento	8.236,25 95% Reconhecimento	95%

Tabela 7.16 Aplicação da metodologia proposta em outros tipos de documento

O primeiro tipo de documento, testado com 100 páginas, obteve pela abordagem tradicional uma taxa de reconhecimento de 99,5% (tab. 7.16). Com a aplicação da metodologia proposta essa taxa subiu para 99,8%, tendo-se um incremento de 0,3%. Em princípio essa melhoria parece insignificante, mas se for considerado que, neste caso, isto representa aproximadamente 10 caracteres reconhecidos a mais por página, então o incremento passa a ter algum interesse.

O segundo tipo de documento, testado com 150 Notas Fiscais de um mesmo emitente, pela abordagem tradicional não foi possível obter absolutamente nenhum reconhecimento, isto pelo fato de o conteúdo estar impresso em um fundo contra fraude que não permite reconhecimento via OCR e que a abordagem tradicional não conseguiu eliminar (tab. 7.16). Pela metodologia proposta, o fundo foi eliminado, obtendo-se assim uma taxa média de reconhecimento de 95%, o que viabiliza uma indexação automática com pouca intervenção manual via digitação.

Assim apesar da necessidade de testes com rigor científico para diversos outros tipos de documento, por apresentar resultados satisfatórios para vários tipos de documento, a metodologia demonstra ser muito promissora.

Capítulo 8

Conclusões, Contribuições e Perspectivas

8.1 Conclusões Sobre o Trabalho

8.1.1 Considerações Gerais

Documentos são tão importantes para a evolução da humanidade que grande parte do conhecimento gerado, manipulado e repassado de geração para geração está registrado na forma de documentos. Desde a invenção da imprensa por Gutenberg no século 15, o registro de informações impressas em papel na forma de documentos tem crescido de forma geométrica. Com a chegada dos computadores no século 20, o papel começa dar lugar aos bits, onde muitos dos bilhões de documentos gerados em papel nestes últimos séculos estão sendo convertidos para o meio eletrônico. A principal característica de um documento é possuir um conteúdo que possa ser lido, seja pela visão humana quando se trata de documentos em papel, seja pela visão computacional quando se trata de documento eletrônico gerado a partir do papel. Assim a qualidade do conteúdo de um documento está diretamente relacionada a possibilidade de leitura desse conteúdo.

Muitas pesquisas com êxito têm sido realizadas na área de reconhecimento de padrões, o que tem permitido a “leitura computacional” com altas taxas de acerto em imagens eletrônicas de documentos simples. Porém para muitos documentos complexos, os resultados deixam a desejar, e na maioria dos casos desse tipo, a dificuldade se dá pelas características e qualidade desse tipo de documento. Para resolver estes problemas, as pesquisas têm se voltado para a área de processamento de imagens, onde muitas técnicas já foram formuladas, mas poucos testes com rigor

científico têm sido realizados com o objetivo de encontrar os melhores métodos para aprimorar a qualidade dos conteúdos impressos de documentos complexos com grandes volumes.

Este trabalho, com o específico objetivo de apresentar uma metodologia para melhorar a qualidade de conteúdos impressos em grandes volumes, mostrou que com a combinação de técnicas adequadas de processamento de imagens como a limiarização e a morfologia matemática binária, pode-se obter, com tempos viáveis, resultados muito satisfatórios em termos de qualidade, e por conseqüência um significativo aumento na taxa de reconhecimento de conteúdos impressos de documentos complexos em relação a principal abordagem até então utilizada.

Conclusões sobre os principais resultados obtidos com este trabalho, são a seguir apresentadas.

8.1.2 Conclusões Sobre os Resultados

Resultados obtidos com a aplicação da metodologia proposta, a partir de 402 documentos complexos, divididos em duas bases (teste e validação) e segmentados em 5 regiões com características particulares, mostraram um significativo incremento na qualidade em relação a abordagem tradicional, e com baixos tempos de processamento.

Combinando-se métodos de limiarização mais indicados com parâmetros da morfologia matemática binária mais adequados para os documentos da base, foram produzidos diversos resultados com incrementos na qualidade dos conteúdos impressos. Na avaliação da qualidade, realizada por reconhecimento óptico de caracteres, foram obtidos os seguintes melhores incrementos por região:

- Na região de baixo contraste, com caracteres alfabéticos impressos em fundo escuro, aplicando-se a limiarização por seleção iterativa de LLoyd combinada com uma iteração da erosão morfológica binária com elemento estruturante quadrado, o incremento foi de 42,77%, tendo-se um acréscimo de 42,52 vezes na taxa de reconhecimento a um custo de 5% de tempo adicional em relação a abordagem tradicional.

- Na região de alto contraste, com caracteres alfabéticos impressos em fundo claro, aplicando-se a limiarização por seleção iterativa de LLOYD combinada com uma iteração da erosão morfológica binária com elemento estruturante quadrado, o incremento foi de 39,29%, tendo-se um acréscimo de 4,31 vezes na taxa de reconhecimento a um custo de 5% de tempo adicional em relação a abordagem tradicional.
- Na região de baixo contraste, com caracteres numéricos impressos em fundo escuro, aplicando-se a limiarização por seleção iterativa de LLOYD combinada com uma iteração da erosão morfológica binária com elemento estruturante quadrado, o incremento foi de 74,16%, tendo-se um acréscimo de 24,69 vezes na taxa de reconhecimento a um custo de 5% de tempo adicional em relação a abordagem tradicional.
- Na região de alto contraste, com caracteres numéricos impressos em fundo claro, aplicando-se a limiarização por seleção iterativa de LLOYD combinada com uma iteração da erosão morfológica binária com elemento estruturante cruz, o incremento foi de 64,51%, tendo-se um acréscimo de 5,81 vezes na taxa de reconhecimento a um custo de apenas 2% de tempo adicional em relação a abordagem tradicional.
- Na região de baixo contraste, com caracteres pré-impressos em fundo claro, aplicando-se a limiarização por entropia de WuLu combinada com uma iteração da abertura morfológica binária com elemento estruturante quadrado, o incremento foi de 60,85%, tendo-se um acréscimo de 2,96 vezes na taxa de reconhecimento a um custo de 25% de tempo adicional em relação a abordagem tradicional. Como aqui, o conteúdo impresso não estava degradado pelo fato de estar pré-impresso, a melhoria se deu pelo baixo contraste, situação onde o método de WuLu é mais eficiente.
- Cabe ressaltar que o método de limiarização de Otsu continua sendo o mais indicado para documentos em geral. Porém este trabalho demonstrou que para documentos com características particulares, existem métodos melhores que o de Otsu.

8.2 Contribuições do Trabalho

8.2.1 Contribuições para a Pesquisa Científica

A principal contribuição para a pesquisa científica é a comprovação de 3 hipóteses formuladas no capítulo 6 e comprovadas no capítulo 7 que são:

- Em documentos complexos, aplicando-se técnicas de limiarização global em regiões com características particulares, pode-se obter um significativo incremento de qualidade dos conteúdos impressos em relação a abordagem tradicional. A abordagem tradicional com grandes volumes, limiariza os documentos sem considerar a regionalização.
- Encontram-se métodos de limiarização global para melhorar conteúdos impressos de documentos complexos, mais eficientes que o popular e rápido método de Otsu, tanto em termos de melhoria da qualidade como em termos de tempo de processamento.
- Aplicando-se parâmetros adequados da morfologia matemática binária sobre uma imagem em preto e branco pode-se, a baixo custo, obter um significativo incremento na qualidade dos conteúdos impressos de documentos complexos.

Outra contribuição é a proposta de automação da avaliação dos resultados com a inclusão da fase de digitação, que pode ser utilizada em futuras pesquisas com objetivos semelhantes.

8.2.2 Contribuições para Aplicação nas Empresas

Os resultados do trabalho têm importante contribuição para aplicações de gerenciamento eletrônico de documentos. Aumentar as taxas de reconhecimento de documentos eletrônicos gerados a partir do papel, significa reduzir custos e prazos o que muitas vezes pode significar a viabilização de alguns projetos. O aumento na taxa de reconhecimento na conversão de conteúdos impressos em imagens para texto, significa

reduzir a intervenção manual via digitação. Em soluções de Full Text Retrieval, qualquer incremento nas taxas de reconhecimento, significa mais caminhos de pesquisa para localização de imagens em bancos de dados. Para a indexação de imagens por conteúdos pré-determinados, a automação só é possível com altas taxas de reconhecimento para a geração dos índices.

Com base nos importantes resultados obtidos com este trabalho, foi desenvolvida uma abordagem para viabilizar documentos complexos para GED [KUBIÇA, 2001]. Esta abordagem foi apresentada no vigésimo quinto congresso internacional de gerenciamento eletrônico de documentos INFOIMAGEM2001, e já está sendo utilizada em projetos do Governo do Estado do Paraná.

8.3 Perspectivas para o Futuro

8.3.1 Perspectivas para a Pesquisa

Considerando-se os importantes resultados obtidos e o grande interesse na melhoria de conteúdos impressos, propõe-se, a partir deste trabalho, o desenvolvimento de novos estudos, como por exemplo:

- Testar a metodologia proposta com diversos outros tipos de documentos complexos com o objetivo de comprovar a sua abrangência.
- Ampliar e testar a metodologia proposta com mais um nível de combinação de técnicas incluindo o processamento em níveis de cinza como filtros e morfologia matemática em níveis de cinza.
- Rotulação automática com sistema de indexação e recuperação dos rótulos, evitando-se a nomeação que é dependente de Sistema Operacional.
- Implementar e testar outros métodos de limiarização e/ou outros recursos da morfologia matemática binária.
- Aprofundar a análise comparativa de forma a comprovar a relação de qualidade existente entre a avaliação visual e a avaliação por OCR.

- Integrar a ferramenta desenvolvida com o software de OCR com o objetivo de, através de um processamento baseado na computação evolutiva, automatizar a seleção dos melhores métodos e parâmetros de acordo com a evolução dos resultados da imagem que esta sendo processada.
- Incluir a variável tempo de forma a possibilitar que os melhores métodos sejam selecionados em função do volume de documentos e do tempo disponível para o processamento.

8.3.2 Perspectivas para as Empresas

Em função das grandes vantagens que podem ser obtidas pelo incremento nas taxas de reconhecimento e pelos promissores resultados obtidos com este trabalho, novas abordagens podem ser desenvolvidas, como por exemplo recursos de melhoria da qualidade incorporados em aplicações de GED. Estes novos recursos podem inclusive no futuro estar disponíveis nos próprios equipamentos de captura, utilizando computação evolutiva para selecionar de forma automática as melhores técnicas de processamento de imagens, ou até mesmo fazendo uso de redes neurais com o objetivo de compatibilizar métodos com características de documentos.

Hoje existe um grande interesse nas organizações pela gestão do conhecimento. Muito do conhecimento organizacional que deve ser compartilhado e preservado, está registrado na forma de documentos impressos. Converter estes importantes documentos para o meio eletrônico com qualidade satisfatória, será um grande desafio.

No futuro próximo, com a segurança que a assinatura digital pode proporcionar [KUBIÇA, 2001-2] e com a eminente legalidade dos documentos eletrônicos no Brasil, haverá grande interesse na conversão dos atuais documentos que hoje estão em meio analógico para o meio digital. A lei modelo das Nações Unidas [UNCITRAL, 2001] estabelece que uma imagem gerada a partir de um documento em papel, é na verdade uma cópia, o que hoje ainda inibe a eliminação efetiva do papel. Porém com a assinatura digital e a legalidade dos documentos eletrônicos, essas barreiras deverão desaparecer, e então haverá necessidade de geração de grandes volumes de imagens de documentos com qualidade satisfatória e boas taxas de reconhecimento. Para isso este trabalho oferece uma importante contribuição.

Capítulo 9

Referências Bibliográficas

- [AMATO, 2000] D'Amato D., **Imaging Systems: the Range of Factors Affecting Image Quality**, Guides to Quality in Visual Resource Imaging, July 2000.
<http://www.rlg.org/visguides/visguide3.html>
- [AURÉLIO, 1991] Aurélio B. L., **Novíssimo Dicionário da Língua Portuguesa**, Editora Nova Fronteira, 1991.
- [BAIRD, 1999] Baird H. S., **Document Image Quality: Making Fine Discriminations**, Xerox Palo Alto Research Center, Palo Alto, USA, 1999.
- [BAIRD, 2000] Baird H. S., **The State of the Art of Document Image Degradation Modelind**, Xerox Palo Alto Research Center, Palo Alto, USA, 2000.
- [BANON, 1994] Banon Gerald J., Barrera J., **Bases da Morfologia Matemática para a Análise de Imagens Binárias**, 9th Escola de Computação, 230p, Recife, 1994.
- [BAYER, 1997] Bayer T., Bohnacker U., Benz I., **Information Extraction from Paper Documents**, Daimler-Benz AG, Research ans Technology, Germany, 1997.
- [BERSEN, 1986] Bersen J., **Dynamic Thresholding of Gray-Level Images**, Proc. 8th Int'l Conf. on Pattern Recognition, pp. 1251-1255, Paris, France, 1986.
- [CAERE, 2001] ScanSoft, Inc., (USA) 9 Centennial Drive Peabody, MA 01960.
www.caere.com/products/ocr.asp

- [CANNON, 1999] Cannon M., Hochberg J., Kelly P., **Quality Assessment and Restoration of Typewritten Document Images**, Los Alamos National Laboratory, 1999.
- [CDIA, 2001] Certified Document Imaging Architect, The Rheinner Group Education Program, Edition 2.0, 2001.
www.rheinner.com
- [CHANG, 2000] Chang F., **Retrieving Information from Document Images: Problems and Solutions**, Institute of Information Science 20, Taiwan, 2000.
- [DEBORA, 2000] **Digital Access to Books of the Renaissance**, European project n.LB 5608A, September 2000.
- [DENGEL, 1990] Dengel A., **ANASTASIL: A System Low-Level and High-Level Geometric Analysis of Printed Documents**, German Research Center for Artificial Intelligence (DFKI), Germany, 1990.
- [EIKVIL, 1991] Eikvil L., Taxt T., Moen K., **A fast Adaptive Method for Binarization of Document Images**, Proceedings of the First International Conference on Document Analysis and Recognition, pp.435-443, Saint-Malo, France, 1991.
- [FACON, 1996] Facon J., **Mathematical Morphology: Theory and Examples**, in Portuguese, Curitiba, Brazil, October, 1996.
- [GORMAN, 1998] O’Gorman L., Kasturi R., **Document Image Analysis**, IEEE Computer Society Press, Los Alamitos, California, 1998.
- [HOBBY 1996] Hobby J. D., Baird H. S., **Degraded Character Image Restoration**, 1996
- [HOUAISS, 2001] Houaiss A., **Dicionário Houaiss da Língua Portuguesa**, Instituto Antônio Huaiss de Lexicografia, Objetiva, Rio de Janeiro, 2001.
- [HUANG, 1995] Huang L. K., Wang M. J. J., **Image Thresholding by Minimizing the Measures of Fuzziness**, Pattern Recognition, v.28, n1, pp. 41-51 1995.

- [JAWAHAR, 1997] Jawahar C. V., Bismas P. K., Ray K., **Investigations on Fuzzy Thresholding Based on Fuzzy Clustering**, Pattern Recognition, vol. 30, no. 10, pp.1605-1613, 1997.
- [JESUS, 1999] Jesus A., **Uma Abordagem Morfológica Para a Segmentação de Logotipos em Cheques Bancários Brasileiros**, Dissertação de Mestrado, Pontifícia Universidade Católica do Paraná, Curitiba, 1999.
- [JOHANNSEN, 1982] Johannsen G. Bille J., **A Threshold Selection Method using Information Measures**, Proceedings 6th International Conference. Pattern Recognition, Munich, Germany. pp. 140-143, 1982.
- [KANUNGO, 1996] Kanungo T., **Document Degradation Models and a Methodology for Degradation Model Validation**, Dissertation Doctor of Philosophy, University of Washington, USA, 1996.
- [KANUNGO, 1999] Kanungo T., Resnik P., **The Bible, Truth, and Multilingual OCR Evaluation**, Center for Automation Research, Institute for Advanced Computer Studies, University of Maryland, 1999.
- [KAPUR, 1985] Kapur N. J., Sahoo P. K., Wong A. C. K., **A New Method for Gray-level Picture Thresholding**, Computer Vision Graphics and Image Processing, v.29 pp. 273-285, 1985.
- [KATZ, 2000] Katz R. J., **Image Analysis and Supervised Learning in the Automated Differentiation of White Blood Cells from Microscopic Images**, Thesis of Master of Applied Science in Information Technology, Department of Computer Science, RMIT, 2000.
- [KENNEY, 1999] Kenney A. R., Sharpe L. H., **Illustrated Book Study: Digital Conversion Requirements of Printed Illustrations** Cornell University Library, July 1999. <http://www.library.cornell.edu>
- [KUBIÇA, 2001] Kubiça S., **Viabilizando Documentos Complexos Para GED**, Anais do vigésimo quinto congresso internacional de gerenciamento eletrônico de documentos – INFOIMAGEM2001, São Paulo, Brasil, 2001.

- [KUBIÇA, 2001-2] Kubiça S., **Uma Proposta para Utilização da Assinatura Digital Como Componente de Segurança no Tráfego e Armazenamento de Documentos Eletrônicos**, Anais do quarto simpósio internacional de gestão do conhecimento e gestão documentos, ISKM/DM2001, Curitiba, Brasil, 2001.
- [LE, 1995] Le D. X., Thoma G. R., Wechler H., **Classification of Binary Images into Textual or Nontextual Data Blocks Using Neural Network Models**, Machine Vision and Applications, Department of Computer Science, George Mason University, USA, 1995.
- [LI, 1993] Li C.H., Lee C.K., **Minimum Cross Entropy Thresholding**, Pattern Recognition, v.26, no 4, pp 616-626, 1993.
- [LLOYD, 1988] Mardia, K. V., Hainsworth T. J, **Spatial Thresholding Method for Image Segmentation**, IEEE, N° 06, VOL .10, 919-927, November 1988.
- [MAO, 2000] Mao S., Kanungo T., **PSET: A Page Segmentation Evaluation Toolkit**, Laboratory for Language and Media Processing, University of Maryland, USA, 2000.
- [MATRAKAS, 2000] Matrakas M. D., **Extração da Estrutura Física e Lógica de Documentos Comerciais**, Dissertação de Mestrado, Pontifícia Universidade Católica do Paraná, Curitiba, 2000.
- [MATTANA, 1999] Mattana M. F., **Avaliação por Reconhecimento da Qualidade da Segmentação por Binarização de Cheques Bancários**, Dissertação de Mestrado, Pontifícia Universidade Católica do Paraná, Curitiba, 1999.
- [MELLO, 1999] Mello C. A. B., Lins R. D., **A Comparative Study on OCR Tools**, Departamento de Informática, Universidade Federal de Pernambuco, Recife PE, Brazil, 1999.
- [MORI, 1992] Mori S., Suen C. Y., Yamamoto K., **Historical Review of OCR Research and Development**, Proceedings of the IEEE Vol.80, N.7, pp. 1029-1058, 1992.
- [NIBLACK, 1986] Niblack W., **An Introduction to Digital Image Processing**, Prentice Hall, Englewood Cliffs, N.J., 1986.

- [OTSU, 1979] Otsu N., **A Threshold Selection Method from Gray-Level Histograms**, IEEE Trans. Sys. Man and Cybernetics, v. SMC9, n°1, pp. 62-66, 1979.
- [PEARLESTINE, 2000] Pearlestine L. G., **Discrimination of an Invasive Plant from Aerial Digital Imagery**, Dissertation for the Degree of Doctor of Philosophy, University of Florida, 2000.
- [PUN, 1981] Pun T., **Entropic Thresholding – The New Approach**, Computer Graphics and Image Processing, vol.16, pp.210-239, 1981.
- [RANDRIAMASY, 1994] Randriamasy S., Vicent L., Wittner B., **An Automatic Benchmarking Scheme for Page Segmentation**, Harvard Robotics Lab, Cambridge MA, 1994.
- [RIDLER, 1978] Ridler T. W., Calvard S., **Picture Tresholding Using an Interactive Selection Method**, IEEE Trans on Systems Man and Cybernetics, Vol. SMC-8 pp. 630-632, 1978.
- [SAHOO, 1988] Sahoo P. K., Soltani S., Wong A. K. C., **A Survey of Thresholding Techniques**, Computer Vision, Graphics and Image Processing, vol.41, pp 233-260, 1988.
- [SERRA, 1982] Serra J., **Image Analysis and Mathematical Morphology**. Academic Press, London, 1982.
- [SRIHARI, 1996] Srihari S. N., Shin Y., Ramanaprasad V., Lee D., **A System to Read Names and Addresses on Tax Forms** , Proceedings of the IEEE Vol 84 n. 7 , July 1996.
- [SWAIN, 1997] Swain J. D., **Enhancing and Reconstructing Digitized Handwritihg**, Thesis of Master of Science in Computer Science, Virginia Polytechnic Institute and State University, 1997.
- [THRUSSEL, 1979] Thrussel H. J., **Picture Threshold using na Iterative Selection Method**, IEEE Transaction on Systems Man and Cybernetics, v.SMC 9 no.5 pp. 311, 1979.
- [TSUJIMOTO, 1992] Tsujimoto S., Asada H., **Major Components of a Complete Text Reading System**, IEEE Vol. 80 n. 7, July 1992.

- [UNCITRAL, 2001] United Nations Commission on International Trade Law, 2001.
www.uncitral.org
- [WEBSTER, 1998] **Webster's New World Dictionary**, Macmillan General Reference, Prentice Hall, 1998.
- [WU, 1998] Lu W., Sondge M., Hanging L., **An Effective Entropic Thresholding for Ultrasonic Image**, IEEE, pp. 1552-1554, 1998.
- [XUE, 1999] Xue J., Ding X., Liu C., Pan S., Kong H., **Destination Address Block Location on Handwritten Chinese Envelope**, Departament of Eletronic Engineering, Tsinghua University, Beijing, China, 1999.
- [YAGER, 1979] Yager R. R., **On the Measures of Fuzziness and Negation**, Part. 1: Membership in the Unit Interval Int Journal of Gen. Sya v.5 pp. 221-229, 1979.
- [YANIKOGLU, 1988] Yanikoglu B. A., Vicent L., **Pink Panther: A Complete Environment for Ground-Truthing and Benchmarking Document Page Segmentation**, IBM Almadem Research, San Jose, USA, 1998.