



UNIVERSIDADE FEDERAL DE PERNAMBUCO
CENTRO DE INFORMÁTICA
CURSO DE BACHARELADO EM ENGENHARIA DA COMPUTAÇÃO

LEANDRO HENRIQUE ESPÍNDOLA VIANA DE ALMEIDA

**MELHORIA DA QUALIDADE DE IMAGENS DE DOCUMENTOS COM
BACKGROUND COMPLEXO**

RECIFE

2012

UNIVERSIDADE FEDERAL DE PERNAMBUCO
CENTRO DE INFORMÁTICA
CURSO DE BACHARELADO EM ENGENHARIA DA COMPUTAÇÃO

LEANDRO HENRIQUE ESPÍNDOLA VIANA DE ALMEIDA

**MELHORIA DA QUALIDADE DE IMAGENS DE DOCUMENTOS COM
BACKGROUND COMPLEXO**

Monografia apresentada ao Centro de Informática (CIN) da Universidade Federal de Pernambuco (UFPE), como requisito parcial para conclusão do Curso de Engenharia da Computação, orientada pelo professor Carlos Alexandre Barros de Mello.

RECIFE

2012

UNIVERSIDADE FEDERAL DE PERNAMBUCO
CENTRO DE INFORMÁTICA
CURSO DE BACHARELADO EM ENGENHARIA DA COMPUTAÇÃO

LEANDRO HENRIQUE ESPÍNDOLA VIANA DE ALMEIDA

**MELHORIA DA QUALIDADE DE IMAGENS DE DOCUMENTOS COM
BACKGROUND COMPLEXO**

Monografia submetida ao corpo docente da Universidade Federal de Pernambuco, defendida e aprovada em 29 de Junho de 2012.

Banca Examinadora:

Carlos Alexandre Barros de Mello

Doutor

Orientador

George Darmiton da Cunha Cavalcanti

Doutor

Examinador 1

À minha bisavó Maria Sebastiana e ao meu
tio Oliveira Viana (*in memoriam*).

AGRADECIMENTOS

À Deus por ter abençoado minha vida com uma família maravilhosa onde encontrei o apoio necessário para enfrentar e superar os obstáculos que a vida apresentou até então.

Aos meus pais pelo exemplo de vida e força. Minha mãe Lúcia de Fátima que acreditou em meu potencial, mesmo quando nem mesmo eu acreditava, pelo amor incondicional dedicado ao longo de toda minha existência, pelos diários exemplos de força, determinação e por ser um referencial de inteligência, organização e superação em minha vida. Meu pai Roldão Viana por conhecer meu verdadeiro potencial, impor desafios que ajudaram a evoluir como ser humano e por estar sempre de coração aberto para ajudar.

Minha querida irmã Jéssica Espíndola pelo apoio nas horas difíceis, pela tolerância habitual que sempre teve com relação a minha personalidade, estando sempre pronta para ajudar e compartilhar momentos em família inesquecíveis.

Meus avós maternos Maria de Fátima e José Alexandre por terem concedido cuidados quando ainda bebê. Minha avó exercendo o papel de uma segunda mãe durante toda infância e adolescência e meu avô pela presença inestimável em meus primeiros anos do maternal.

Avós paternos Luiza Viana e Geraldo Viana (*in memoriam*) por sempre exibirem o mesmo sentimento de alegria sincera desde criança até hoje.

Aos padrinhos e tios Ruy Fernando e Maria Auxiliadora por ensinarem o verdadeiro significado do que é determinação, servindo assim de inspiração para não fraquejar perante as tribulações.

À tio Rinaldo Fabrício por ter ensinado que a busca pelo que é necessário deve sempre preceder a satisfações pessoais.

Tios Robério Alexandre e Maria Aparecida pelos sábios conselhos oferecidos, por servirem de farol, não apenas para mim, como também para toda a família e por mostrarem que a verdadeira Fé reside no coração e não apenas nas palavras.

Querida e amada tia Leyla Fabrícia pelas incontáveis noites de sono que perdeu quando me colocava para dormir quando bebê e ao seu ilustre esposo Francisco Lima (Chico) por sempre mostrar a mesma face alegre e sorridente, não se deixando abalar pelas diversas dificuldades impostas pela vida.

Aos estimados primos e primas: Elayne, Matheus, Filipe e Fernanda pelos momentos de diversão e alegria que proporcionaram e que este trabalho sirva de inspiração e motivação para a vida acadêmica de todos.

A Vanessa Alves por mostrar o que é amor verdadeiro, ajudando a superar meus defeitos, dividindo sua vida comigo e por ajudar na realização deste trabalho, sem você ele não teria sido feito.

Meus amigos Rodrigo Porto, Fagner Nascimento, Sérgio Sette, José Benito, João Guilherme, José Augusto, Pablo Alessandro, Victor Costa, Cássio Melo por todas as experiências boas e ruins ao longo do curso e por terem dado força para a conclusão do curso e o desenvolvimento deste trabalho. Em especial Hallan Cosmo, por me manter constantemente em busca da superação de meus limites tanto na vida acadêmica quanto na profissional.

Minhas amigas Aida Fernandes, Angélica Porto, Carla Fontes, Eline Bianchi e Fabiana Silva, exemplos na vida acadêmica, tendo comprometimento aliado à eficiência e por estarem sempre dispostas a enfrentar os problemas diários com a mesma simpatia e simplicidade.

À professora Maria da Trindade Sobral por ser até os dias de hoje um ícone em minha formação estudantil, sendo merecidamente a personificação de garra, determinação e amor pela educação.

Ao meu orientador Carlos Alexandre Barros de Mello pela excelente orientação ao longo deste trabalho e por ter dado uma visão e motivação inovadora à minha vida acadêmica, que continuará a guiar ao longo dos meus projetos, o meu muitíssimo Obrigado.

Professor Hélio Magalhães por ser um profissional exemplar na área acadêmica, transmitindo seus conhecimentos de forma resoluta e honesta, fazendo uso do chavão “*Read The Masters*”

para estimular seus alunos a buscarem o aprendizado diretamente com quem possui o conhecimento.

À minha atual gerente Adriana Góis e a Orivaldo Santana pela credibilidade depositada em meu potencial ainda como estagiário e por ter apresentado o contexto profissional da área de Processamento Digital de Imagens.

Enfim, a todos que contribuíram de forma direta e indireta ao longo da graduação, a todas as pessoas que não mencionei ao longo deste trabalho mais que me ajudaram a crescer como profissional e a todos que contribuíram para a formação do meu caráter, meus sinceros Agradecimentos.

“A visão é o mais avançado dos nossos sentidos,
de forma que não é de surpreender que as
imagens exerçam o papel mais importante na
percepção humana.”
Rafael C. Gonzalez

RESUMO

Este trabalho trata-se de uma investigação qualitativa e quantitativa sobre a eficiência de dois métodos de limiarização recentemente desenvolvidos. O intuito deste trabalho é iniciar um estudo sobre a eficiência de algumas técnicas de binarização recentemente propostas, que poderá ser utilizado como ponto de partida para uma investigação mais detalhada, a fim de se produzirem resultados mais interessantes. Os algoritmos investigados foram implementados e tiveram seus resultados confrontados com outros algoritmos frequentemente relacionados na literatura, utilizando-se as métricas desenvolvidas em competições de binarização. Foram utilizadas, como exemplares de teste, imagens retiradas destas competições que apresentam nível de ruído classificado entre médio e avançado. Todas as imagens possuem uma imagem-verdade associada, que serve como parâmetro de comparação para a eficiência do algoritmo. Os resultados são obtidos através da análise dos índices de performance obtidos a partir de tais comparações. A análise geral dos resultados sugere uma melhor eficiência, por parte das técnicas implementadas neste trabalho em relação às demais, para binarização de imagens de documentos. As duas técnicas estudadas foram mais eficientes do que pelo menos metade dos algoritmos apresentados em pelo menos um quesito de comparação. Em ambas as técnicas, sugestões de melhorias foram feitas para trabalhos posteriores.

Palavras-chave: Limiarização, Binarização, Algoritmos, Imagens, Processamento Digital,

ABSTRACT

This work is a qualitative and quantitative research on the effectiveness of two newly developed thresholding methods. The purpose of this paper is to initiate a study on the efficiency of some recently proposed binarization techniques that can be used as a starting point for further investigations in order to produce more interesting results. The algorithms were implemented and their results compared with other state-o-the-art algorithms, by means of metrics developed in binarization contests. As test specimens, some images were picked from these competitions that have noise level rated from medium to advanced. All images are associated with a ground truth, which serves as a benchmark for the efficiency of the algorithm. Results are obtained by analyzing rates of performance from such comparisons. The overall results suggest better efficiency by the techniques implemented in this work, for binarization of document images. Both techniques are more effective than at least half of the algorithms presented in at least one item for comparison. In both techniques, suggestions for improvements were made for future work.

Keywords: Thresholding, Algorithms, Images, Digital Processing

SUMÁRIO

Tabela de Siglas.....	13
Lista de Ilustrações.....	14
Lista de Tabelas.....	16
Tabela de Termos.....	17
1. INTRODUÇÃO.....	18
1.1 Motivação.....	18
1.2 Objetivos.....	18
1.3 Estrutura do Trabalho.....	19
2. APRESENTAÇÃO.....	20
2.1 Representação de Imagens Digitais.....	20
2.2 Imagens Monocromáticas e Multibandas.....	23
2.3 Limiarização e Binarização.....	24
2.4 Janelamento.....	25
2.5 <i>Background</i> Complexo.....	26
2.6 Histogramas.....	28
3. BINARIZAÇÃO DE IMAGENS DE DOCUMENTOS.....	29
3.1 Limiarização a Partir da Maximização de Brink.....	29
3.2 Limiarização Fuzzy C Means.....	30
3.3 Limiarização por Entropia de Johanssen.....	31
3.4 Limiarização por Entropia de Wulu.....	31
3.5 Limiarização Através da Projeção do Discriminante Linear de Fisher Sobre o Histograma Bidimensional.....	32
3.6 Limiarização Iterativa de Ye-Danielsson.....	33
3.7 Limiarização <i>Fuzzy</i> de Yager.....	34
3.8 Limiarização Iterativa de Rosenfeld por Relaxação.....	34
3.9 Limiarização Bimodal de Otsu.....	35
3.10 Limiarização de Niblack.....	36
3.11 Limiarização Adaptativa Híbrida de Sauvola.....	36

3.12	Limiarização por <i>Convex Hull</i>	37
3.13	Limiarização por Entropia de Pun.....	37
3.14	Outras Técnicas.....	38
3.14.1	Nível de Cinza Médio.....	38
3.14.2	Percentual de Preto.....	38
4.	ALGORITMOS DE BINARIZAÇÃO PARA IMAGENS DE DOCUMENTOS COM BACKGROUND COMPLEXO.....	39
4.1	Binarização de imagens de documentos utilizando máximos e mínimos locais.....	39
4.1.1	Análise do <i>Background</i> da Imagem.....	39
4.1.2	Seleção dos Pixels de Alto Contraste.....	41
4.1.3	Definição dos Parâmetros para Limiarização Local.....	42
4.1.4	Limiarização de Imagens de Documentos.....	44
4.1.5	Impacto da Largura do Texto na Limiarização.....	45
4.1.6	Análise de Pixels de Alto Contraste Fora das Bordas do Texto.....	49
4.2	Limiarização por níveis lógicos adaptativos.....	51
4.2.1	Definição dos Parâmetros para Limiarização.....	51
4.2.2	Análise de Níveis Lógicos.....	51
4.2.3	Análise de Níveis Lógicos Adaptativa.....	53
5.	EXPERIMENTOS E RESULTADOS.....	55
5.1	Teste I.....	58
5.1.1	Resultados do Teste I.....	63
5.2	Teste II.....	64
5.2.1	Resultados do Teste II.....	67
5.3	Teste III.....	68
5.3.1	Resultados do Teste III.....	75
5.4	Teste IV.....	76
5.4.1	Resultados do Teste IV.....	86

6. CONCLUSÃO.....	87
6.1 Contribuições.....	87
6.2 Trabalhos futuros.....	88
6.1.1 Sugestões sobre a técnica de limiarização por máximos e mínimos locais.....	88
6.1.2 Sugestões sobre a técnica de limiarização por análise de níveis lógicos.....	88
 7. REFERÊNCIAS.....	 90

TABELA DE SIGLAS

Sigla	Significado
DIBCO	Document Image Binarization Contest
RGB	R (Red – Vermelho), G (Green – Verde) e B (Blue – Azul).
SDM	Soft Decision Method
TBM	Text Binarization Method

LISTA DE ILUSTRAÇÕES

FIGURAS

Figura 2.1	Efeitos da redução da resolução sobre uma imagem.....	21
Figura 2.2	Diferentes quantizações para a mesma imagem.....	22
Figura 2.3	Imagem em preto e branco e em tons de cinza.....	23
Figura 2.4	Imagem multibanda e seus canais componentes.....	24
Figura 2.5	Janelas unidimensional e bidimensional.....	25
Figura 2.6	Exemplos de imagens de documentos com <i>background</i> complexo.....	27
Figura 2.7	Histogramas gerados a partir de imagens diferentes.....	28
Figura 3.1	Limiares de Brink sobre um histograma de uma imagem.....	30
Figura 4.1	Imagem de contraste gerada a partir de uma imagem de um documento.....	41
Figura 4.2	Limiarização de Otsu a partir da imagem original e imagem de contraste.....	42
Figura 4.3	Histograma de distâncias entre pixels de alto contraste.....	43
Figura 4.4	Resultado da limiarização utilizando máximos e mínimos locais.....	45
Figura 4.5	Imagem de contraste sobre uma imagem com diferentes tamanhos de texto.....	46
Figura 4.6	Resultado do algoritmo e <i>ground truth</i> correspondente.....	46
Figura 4.7	Estimativa da largura do texto.....	48
Figura 4.8	Resultado do algoritmo com tamanhos de janelas diferentes.....	48
Figura 4.9	Resultado da binarização utilizando limiar de intensidade de pico em 20%.....	50
Figura 4.10	Resultado da binarização utilizando limiar de intensidade de pico em 60%.....	50
Figura 4.11	<i>Ground truth</i> correspondente.....	50
Figura 4.12	Distribuição espacial das vizinhanças escolhidas para comparação, em relação ao pixel processado.....	52
Figura 5.1	Teste I.....	62
Figura 5.2	Teste II.....	66

Figura 5.3	Teste III.....	74
Figura 5.4	Teste IV.....	85

EQUAÇÕES

Equação 2.1	Limiarização e Binarização.....	25
Equação 3.1 e 3.2	Cálculo das Probabilidades Iniciais C Means.....	30
Equação 3.3	Limiarização por Entropia de Johanssen.....	31
Equação 3.4 e 3.5	Função de Limiarização Média Local de Fisher.....	32
Equação 3.6 e 3.7	Função de Limiarização Iterativa de Ye-Danielsson.....	33
Equação 3.8	Limiarização Fuzzy Yager.....	34
Equação 3.9	Probabilidade para Pixels Pretos na Limiarização Iterativa de Rosenfeld.....	35
Equação 3.10	Limiar de Niblack Controlado pelo <i>Bias</i>	36
Equação 3.11, 3.12 e 3.13	Cálculo da Entropia de Pun.....	38
Equação 4.1	Cálculo do Valor de um Pixel da Imagem de Contraste.....	40
Equação 4.2 e 4.3	Média e Desvio Padrão da Intensidade dos Pixels de Alto Contraste.....	44
Equação 4.4	Relação de Classificação Final dos Pixels.....	44
Equação 4.5	Intensidade Média Calculada Dentro de uma Janela.....	52
Equação 4.6	Comparação entre a Média da Janela e o Valor do Pixel.....	53
Equação 4.7	Cálculo do Valor do Pixel na Imagem Resultante.....	53
Equação 4.8	Parâmetro T Mais Próximo do Valor de Intensidade Mínima.....	54
Equação 4.9	Parâmetro T Mais Próximo do Valor da Intensidade Média.....	54
Equação 4.10	Parâmetro T Proporcional à Intensidade Média da Janela.....	54
Equação 5.1	Cálculo do <i>F-measure</i>	55
Equação 5.2	Cálculo do <i>Peak Signal-to-Noise Ratio</i>	56
Equação 5.3	Cálculo do <i>Negative Radio Metric</i>	56
Equação 5.3	Cálculo do <i>Misclassification Penalty Metric</i>	57

LISTA DE TABELAS

Tabela I	Resultados do Teste I.....	64
Tabela II	Resultados do Teste II.....	68
Tabela III	Resultados do Teste III.....	76
Tabela IV	Resultados do Teste IV.....	87

TABELA DE TERMOS

Termo	Significado
Background	Fundo da imagem
Foreground	Objeto relevante da imagem
Fuzzyness	Termo utilizado pelo autor
Ground Truth	Imagem verdade
Stroke Width	Largura do texto

1. INTRODUÇÃO

Este capítulo apresenta o leitor ao tema, trazendo as motivações e objetivos para este trabalho. A estrutura textual deste documento também é apresentada.

1.1 MOTIVAÇÃO

O Processamento Digital de Imagens concentra seus esforços em duas áreas de atuação importantes: melhoria da disposição dos dados da imagem para transmissão, armazenamento e representação automática por um computador, e melhoria da aparência visual da imagem, em caso de percepção humana. Em ambos os casos, é necessário classificar as informações contidas na imagem de maneira correta e eficiente, buscando-se sempre destacar a informação presente na imagem.

Para tal, um processo fundamental nessa análise é a operação de Limiarização, que consiste, em poucas palavras, em um método de classificação das informações contidas na imagem, utilizando-se limiares de classificação, que separam os pontos da imagem em grupos, que mais tarde podem ser interpretados adequadamente. A forma de Limiarização mais comum é a Binarização, que separa os objetos presentes na imagem, dos elementos de fundo, dividindo, assim, os pontos da imagem em dois conjuntos disjuntos.

Limiarizar uma imagem de maneira eficiente, porém, não é uma atividade trivial. A dificuldade maior reside em se encontrar o limiar ideal para classificar todos os elementos presentes na imagem de maneira correta, mesmo que estes objetos estejam sujeitos a ruídos diversos. A busca por este limiar perfeito acaba gerando várias técnicas de limiarização diferentes, que funcionam dentro de seus propósitos, mas ainda não constituem uma técnica de limiarização definitiva.

1.2 OBJETIVOS

Este trabalho tem por objetivo iniciar um estudo sobre técnicas de Limiarização de imagens de documentos, fazendo um levantamento inicial sobre as capacidades, vantagens e desvantagens das técnicas de limiarização presentes na literatura, de forma a difundir e

estimular ainda mais a busca pela técnica definitiva de limiarização de imagens digitais de documentos, comparando quantitativamente alguns procedimentos existentes e abrindo espaço para uma análise crítica que poderá servir como impulso para o surgimento de melhorias ou técnicas novas que apresentem um resultado mais interessante.

Para tanto, apresentamos ao leitor algumas técnicas de limiarização bastante utilizadas, familiarizando-o com alguns dos termos frequentemente utilizados na área de Processamento Digital de Imagens, além de investigar o comportamento de dois algoritmos de binarização, comparando-os com outras técnicas já estabelecidas na literatura, e apresentando os resultados utilizando-se da métrica para comparação de imagens desenvolvida e utilizada pela DIBCO [25].

Nesse sentido, neste trabalho consideramos como melhoria da qualidade da imagem a completa remoção de seu *background* (os tons relacionados ao papel no caso de imagens de documentos), resultando em uma imagem onde apenas os tons relacionados ao *foreground* (no caso, a tinta) permanecem, sendo convertidas para preto. Isso resulta em uma imagem de melhor qualidade em diversos sentidos: (i) em termos de leitura, a remoção dos tons do papel não possibilitam a confusão desses com os tons da tinta; (ii) em geral, ferramentas de reconhecimento óptico de caractere trabalham com imagens em preto-e-branco (onde preto é a tinta e branco o papel), assim, melhorias no processo de conversão para esse padrão de cores pode gerar melhores respostas de tais ferramentas.

1.3 ESTRUTURA DO TRABALHO

Capítulo 2: Apresentação de termos inerentes à área de Processamento Digital de Imagens que serão frequentemente utilizados neste trabalho.

Capítulo 3: Apresentação de algumas técnicas consagradas na literatura para binarização de imagens em geral.

Capítulo 4: Estudo inicial sobre duas técnicas recentes de binarização propostas especificamente para a binarização de imagens de documentos.

Capítulo 5: Comparação dos resultados obtidos pelas técnicas estudadas em relação às outras técnicas apresentadas neste trabalho, seguindo as métricas utilizadas para tal.

Capítulo 6: Conclusões e sugestões para melhorias sobre as técnicas estudadas.

2. APRESENTAÇÃO

Neste capítulo, são introduzidos alguns termos e conceitos utilizados ao longo deste trabalho. O Capítulo três apresenta, de forma resumida, algumas técnicas de binarização disponíveis na literatura. No Capítulo quatro, duas técnicas de binarização são investigadas em detalhes, e seus resultados são apresentados e comparados com as outras técnicas no Capítulo cinco, de forma visual e técnica. O Capítulo seis conclui o trabalho e apresenta sugestões para melhorias de performance em ambas as técnicas estudadas.

2.1 REPRESENTAÇÃO DE IMAGENS DIGITAIS

Quando observamos uma fotografia, ou uma cena do mundo real, recebemos de cada ponto do espaço um impulso luminoso que associa uma informação de cor a esse ponto. Dessa forma, um modelo matemático natural para descrever uma imagem é o de uma função definida em uma superfície bidimensional, que toma valores de um espaço de cor previamente definido [3]. Sendo assim, toda imagem pode ser definida como uma função de intensidade luminosa cujo valor nas coordenadas espaciais (x , y) fornece a intensidade da imagem naquele ponto [1].

Uma imagem digital corresponde a uma forma discreta de representação de uma imagem. A imagem digital pode ser definida como o resultado de uma operação de digitalização efetuado sob o domínio espacial, como definido em [1]. A operação de digitalização envolve os processos de amostragem e quantização.

Amostragem é a discretização do domínio de definição da imagem, em duas dimensões, criando uma matriz de amostras de tamanho $M \times N$. O número de linhas M da matriz de amostras é chamado de resolução vertical da imagem, enquanto que o número de colunas N da matriz é a resolução horizontal, como definido em [3]. O produto $M.N$ é definido como a resolução espacial, ou simplesmente a resolução da imagem. Percebe-se, portanto, que quanto maior a resolução da foto, maior a quantidade de amostras coletadas e mais informação estará contida na imagem digital, ao custo de produzir arquivos digitais de tamanho maior (Figura 2.1).

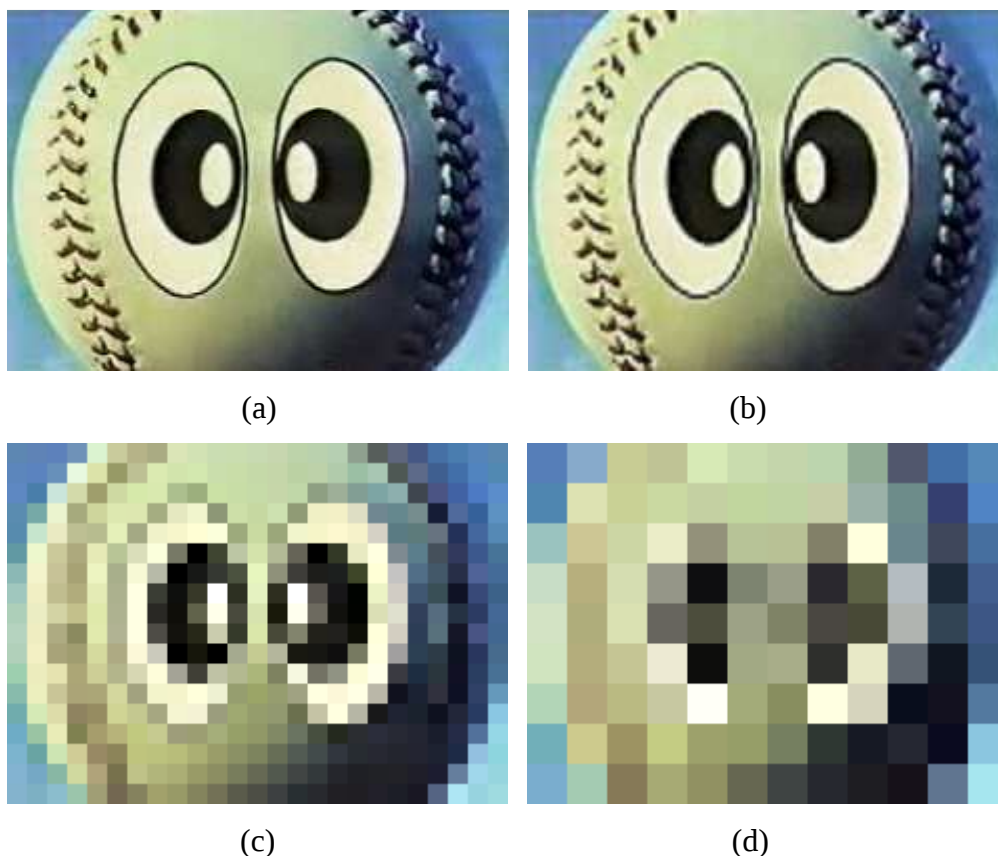


Figura 2.1: Efeitos da redução da resolução sobre uma imagem (em pixels). (a) 240 x 180, (b) 120 x 90, (c) 24 x 18, (d) 12 x 9. Imagens exibidas na mesma dimensão para ressaltar a perda de informação.

Quantização é uma operação que define, para cada ponto da matriz de amostragem, o valor máximo de intensidade, ou amplitude, permitido para o ponto em questão. É através da quantização que ocorre o processo de conversão de uma imagem com um conjunto contínuo de cores, em uma imagem com um conjunto discreto. Dessa forma, a quantização determina a resolução de cor da imagem digital. [3].

Geralmente, cada amostra da matriz possui um valor de intensidade diferente, sempre variando entre zero e $L-1$, onde L é o número máximo de níveis de intensidade definido pela quantização da imagem. Em se tratando de Processamento Digital de Imagens, a quantização geralmente é definida em função do número de bits utilizados para representar a intensidade da amostra, que sempre será menor do que a quantidade de cores presentes no ambiente real. (Figura 2.2).

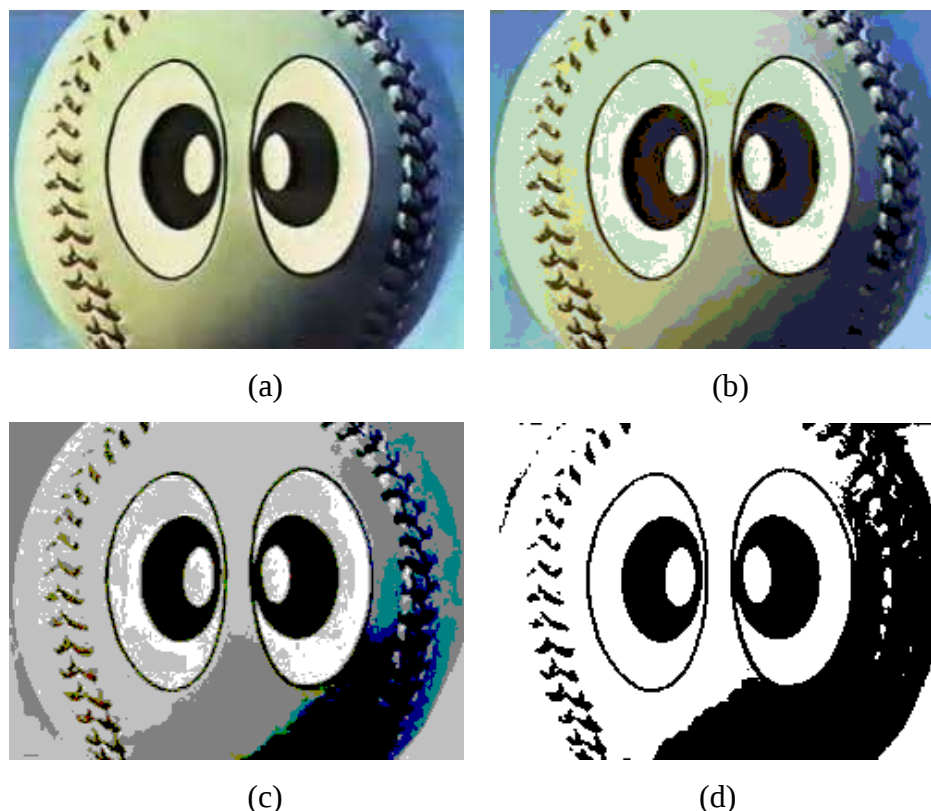


Figura 2.2: Diferentes quantizações para a mesma imagem. (a) Imagem com vinte e quatro bits de cor, (b) oito bits, (c) quatro bits e (d) um bit.

Sendo assim, é possível perceber que cada elemento $A_{i,j}$ da matriz de amostras representa um par de coordenadas (x, y) e possui pelo menos um valor de intensidade $f(x, y)$ associado. Esses elementos são chamados de *elementos pictóricos*, ou simplesmente *pixels* [2]. Um *pixel* constitui a unidade atômica de informação de uma imagem, e sobre eles, são efetuadas diversas operações, principalmente sobre o valor de sua intensidade. De maneira que uma imagem, cuja matriz de amostras possui dimensões $M \times N$, possui $M.N$ *pixels*. Ao longo deste trabalho, os termos *imagem* e *pixel* serão utilizados para representar uma imagem digital e seus elementos básicos.

Em uma imagem, cada pixel possui um ou mais valores de intensidade associados. Tais valores estão compreendidos em um intervalo, geralmente entre zero e um valor máximo definido, que comumente assume o valor de uma potência de dois. Ou seja, $L = 2^b$, onde L é o número de níveis de intensidade permitidos para o pixel da imagem. Neste caso, o número inteiro b , que representa a quantidade de bits necessária para armazenar as informações de cor da imagem, é conhecido como Profundidade da imagem.

2.2 IMAGENS MONOCROMÁTICAS E MULTIBANDAS

Uma imagem monocromática é uma imagem que apresenta apenas um canal para determinação da intensidade do pixel. Neste caso, a informação de intensidade do pixel informa o nível de presença ou ausência de luz no pixel correspondente. Uma imagem monocromática pode ser classificada como uma imagem *em tons de cinza*, já que a informação sobre a intensidade não é precisa o suficiente para fornecer informação de cor. Costuma-se fazer confusão entre imagem *em tons de cinza* e imagem *em preto e branco*. Para este trabalho, porém, esses conceitos são totalmente distintos. Imagem *em tons de cinza* é uma imagem ordinária, monocromática, geralmente com oito bits de profundidade e que pode apresentar mais de duas cores (geralmente, preto, branco e tons de cinza, como o próprio nome sugere). Imagem *em preto e branco* é uma imagem que sempre apresenta apenas duas cores (Figura 2.3).

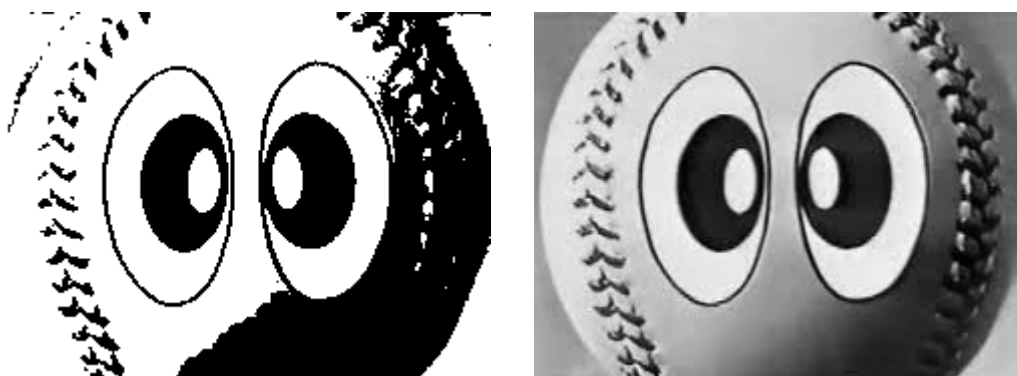


Figura 2.3: (esquerda) Imagem em preto e branco e (direita) imagem em tons de cinza.

Em uma imagem digital multibanda, existe mais de um canal de intensidade para o mesmo pixel, o que lhe garante mais precisão no armazenamento da informação e permite a composição de algum sistema de cor. Sendo assim, uma imagem colorida é uma imagem multibanda em que a cor de cada pixel é determinada pela composição de mais de um canal de intensidade [1]. Um modelo de representação de cores bastante simples e muito funcional é o modelo RGB, que apresenta, para cada pixel, três canais de informação sobre a intensidade, de maneira que a cor do pixel é representada pela composição da informação contida nos três canais, R (*Red* – Vermelho), G (*Green* – Verde) e B (*Blue* – Azul) (Figura 2.4).

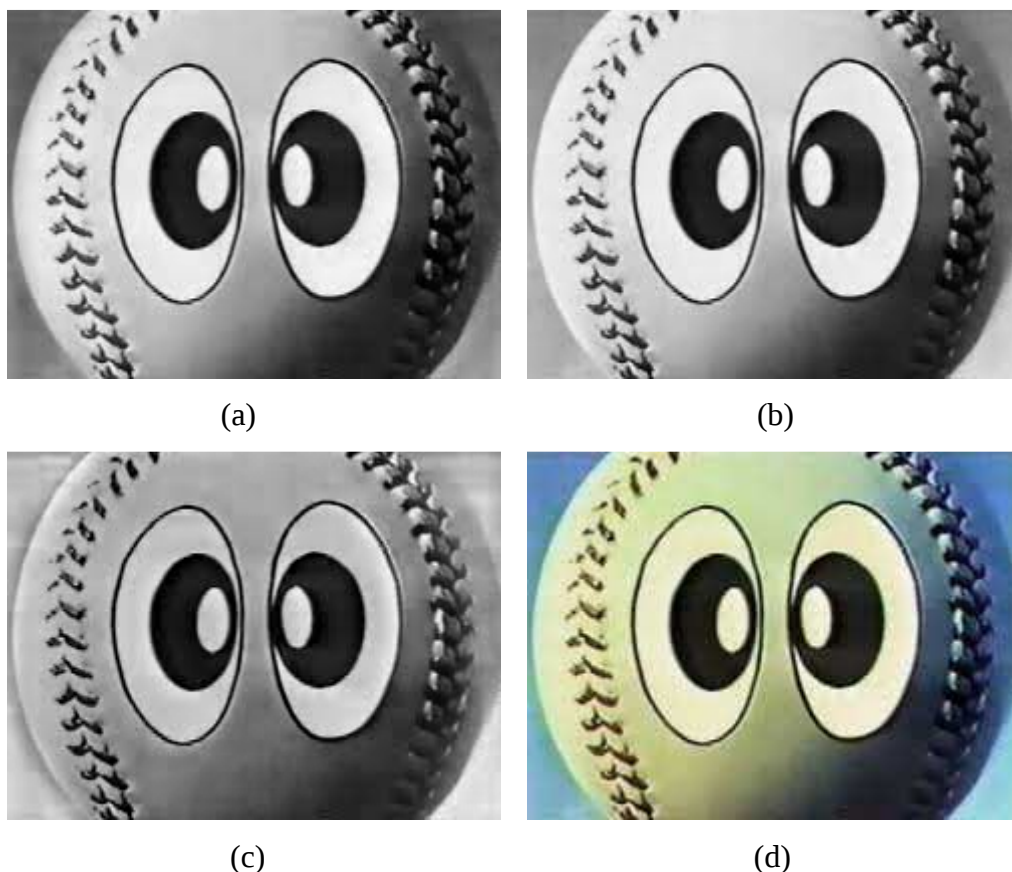


Figura 2.4: Intensidade dos canais (a) vermelho, (b) verde, (c) azul e (d) imagem formada a partir da composição dos três canais.

2.3 LIMIAÇÃO E BINARIZAÇÃO

Uma operação de extrema importância para o Processamento Digital de Imagens e ainda foco principal deste trabalho é a operação de Limiarização de uma imagem. Trata-se, pois, de uma forma de representar as informações contidas na imagem de forma mais compacta e intuitiva, sem causar prejuízo no que diz respeito à informação contida na imagem original. A limiarização classifica os pixels da imagem em conjuntos, que podem ser interpretados de maneira adequada, a depender do objetivo pelo qual se propõe a limiarização.

Um intuito de uma limiarização, por exemplo, pode ser separar o que é fundo do que é útil, definindo-se um, ou mais de um valor de intensidade limiar que servirá de parâmetro de comparação com os pixels da imagem. Esta classificação, porém, é uma das possíveis interpretações que uma limiarização pode ter. A Equação 2.1 representa uma forma de limiarização.

$$g(x, y) = \begin{cases} 1, & \text{se } f(x, y) > T \\ 0, & \text{se } f(x, y) \leq T \end{cases} \quad (2.1)$$

onde $g(x, y)$ é o valor de intensidade do pixel na imagem binarizada, $f(x, y)$ é o valor de intensidade do pixel na imagem original e T é o valor do limiar. Como a função $g(x, y)$ só possui dois valores possíveis para a intensidade do pixel, a limiarização é denominada Binarização e a imagem resultante é, de fato, *em preto e branco*.

Quando o valor do limiar T é uma constante aplicada por toda a imagem, o processo é conhecido como *limiarização global*. Em geral, a limiarização global não é adequada, pois as imagens costumam apresentar variações nos níveis de cinza dos objetos e do fundo, além de estarem sujeitas à iluminação não uniforme, ruídos, entre outros fatores [1]. Nesses casos, talvez seja mais interessante mudar o valor do limiar de acordo com suas características locais. Nesse caso, o processo é chamado de *limiarização local* ou *limiarização adaptativa*. O uso de um método de limiarização local geralmente envolve operações de janelamento.

2.4 JANELAMENTO

Aplicar um janelamento sobre uma imagem significa aplicar uma técnica ou avaliar certas características da imagem apenas sobre um conjunto específico de pixels, repetindo o processo várias vezes até que todos os pixels da imagem tenham sido analisados. Por percorrer a imagem várias vezes, é comum na literatura esta operação ser denominada de “Janelamento deslizante” ou “Janela deslizante”. A operação de janelamento é bastante utilizada em limiarizações locais, que geralmente utilizam características de vizinhança para determinação do limiar local. A principal vantagem do janelamento é obter informações a respeito da imagem de forma local, diminuindo a influência de ruídos presentes na imagem durante a coleta de informações dos pixels.

As janelas utilizadas para percorrer a imagem podem ser *unidimensionais* (Figura 2.5 - esquerda) ou *bidimensionais* (Figura 2.5 - direita), e o número de pixels dentro da janela é indicado através da dimensão da janela. As janelas unidimensionais possuem apenas uma dimensão ($1 \times N$), que indica a quantidade de pixels que estão dentro da janela em uma mesma linha. Já as janelas bidimensionais possuem duas dimensões ($M \times N$), que indicam a

quantidade de linhas e colunas que compõem a janela. Neste caso, se o número de linhas for igual ao número de colunas ($M = N = k$), a janela é dita como janela *quadrada*, de dimensão k .

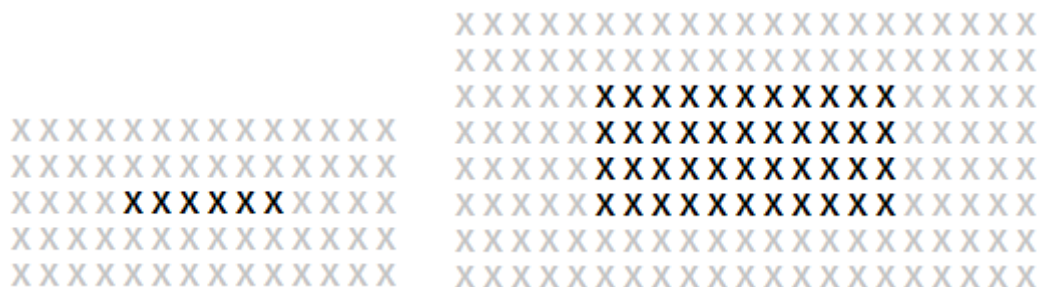


Figura 2.5: (esquerda) Janela unidimensional de dimensão cinco e (direita) janela bidimensional de dimensão quatro por onze.

2.5 BACKGROUND COMPLEXO

Neste trabalho, consideramos imagens de documentos com *background* complexo quando o papel do documento (considerado aqui como o *background*) possui ruídos que o tornam não uniforme. Esses ruídos podem vir de manchas (como provocadas por marcas de fita adesiva, manchas de gordura dos dedos, etc.), presença de diferentes cores no próprio papel, problemas de iluminação na digitalização, etc. Todos esses fatores tornam mais complexa a tarefa de separação entre a tinta e o papel.

Idealmente, as imagens de documentos devem apresentar *background* uniforme. Isso, porém, não é regra, uma vez que documentos antigos estão sujeitos ao desgaste causado pelo tempo e pelas circunstâncias de uso. Quando se digitaliza um documento antigo, ele naturalmente traz consigo um nível de ruído de fundo considerável, que assume forma e disposição aleatórias, sendo assim, uma imagem com *background* complexo.

A Figura 2.6 apresenta várias imagens com diferentes níveis de ruído de fundo, devido a circunstâncias de uso do documento original, com o aparecimento de manchas, ou devido à presença de elementos de texto de outras páginas, agregados com o tempo. Todas elas, no contexto deste trabalho, são consideradas imagens com *background* complexo.

turn me of
 h Icar saw
 I buy much
 y for the same
 tas appear
 had to command
 the same 5000

Gintaro 12h
 11 Jan
 C. W. Short M.D.
 Lexington

Modello 15.

Indicazioni di urgenza

Ufficio Telegrafico di ROMA

TELEGRAMMA

Il telegramma può essere scritto in qualsiasi lingua, ma la lingua deve essere indicata nel campo "LINGUA".

Il telegramma deve essere scritto in modo che la lettura ne sia facile.

DESTINATARIO (1) Arthur Assis Caldas Domicio Jacaragay

DESTINATARIO (2) Dandara Lima Rodrigo Benissimo

TITO Liviana Canier

DATA DELLA PERMANENZA 1900

INDICAZIONE SPECIALE

ROMA, 1900 - Tip. F. Gaudenzi - 18/24

59,7
 enquanto fazer (comessa rosaria) o
 cas que quixer d'ella entre os nossos
 amigos
 J.M. 3
 Joaquin Voluer
 London 31 Maio 1883
 Rev. 3. Junho
 Rev. 30 Julho

TO
CHARLES II.
 KING OF GREAT BRITAIN,
 AND
 THE DOMINIONS THEREUNTO BELONGING:
ROBERT BARCLAY,
 A servant of Jesus Christ, called of God to the Dis-
 pensation of the Gospel now again revealed, and,
 after a long and dark night of Apostacy, commanded
 to be preached to all nations, wisheth health and sal-
 vation.

Eu tive imensa votacão em Bernam-
 buco, em diversas secções da Capital
 fui o mais votado, roubaram-me
 porém grande votacão em todos os
 districtos. Foi uma verdadeira res-
 comotagem - uma vana do Maskelyne
 & Cook de fundos falsos e ven-
 tillogos
 O Jornal foi comprado pelo Ro-

Figura 2.6: Exemplos de imagens de documentos com background complexo.

2.6 HISTOGRAMAS

O *histograma* de uma imagem digital com níveis de intensidade variando dentro de um intervalo definido entre zero e um valor máximo ($L-1$) é uma função de contagem $h(r_k)$, onde r_k é o k -ésimo valor de intensidade e $h(r_k)$ é o número de pixels da imagem com intensidade r_k [2]. Em outras palavras, o histograma de uma imagem conta o número de ocorrências de todos os valores de intensidade presentes na imagem, desprendendo-se das coordenadas espaciais (Figura 2.7). A manipulação de histogramas, além de fornecer informações estatísticas úteis sobre a imagem, pode produzir efeitos de realce, através de operações de equalização, além de ser útil também em operações de compressão e segmentação.

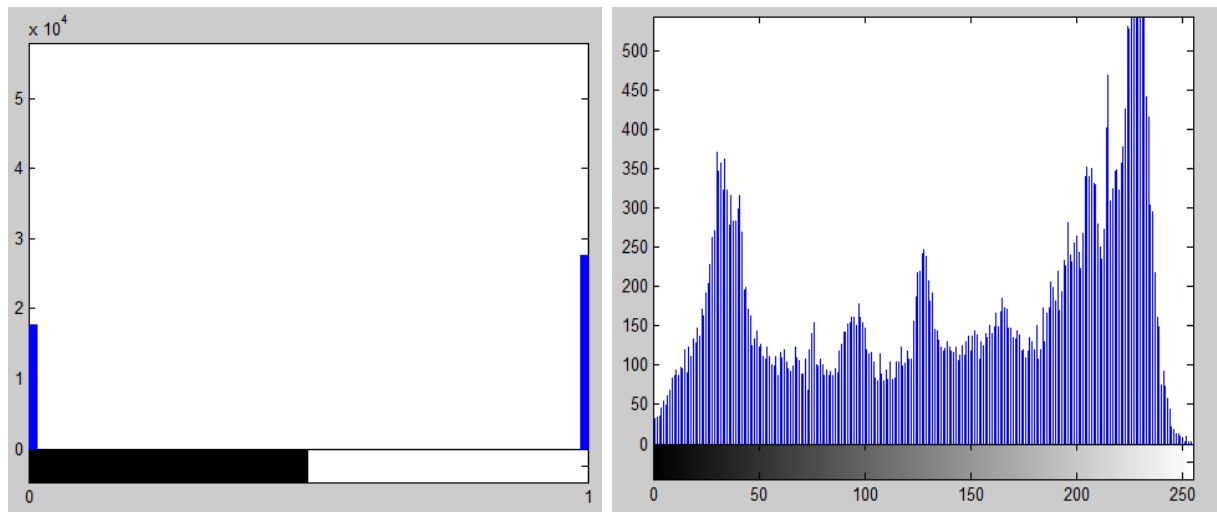


Figura 2.7: Histogramas típicos gerados a partir de imagens diferentes: (esquerda) gerado a partir da Figura 2.3 - esquerda e (direita) gerado a partir da Figura 2.3 - direita. O eixo horizontal apresenta os valores de intensidade presentes, enquanto que o eixo vertical indica o número de ocorrências de cada valor de intensidade da imagem.

3. BINARIZAÇÃO DE IMAGENS DE DOCUMENTOS

Existem muitas formas diferentes de se produzir uma imagem binarizada. Apesar de todas as estratégias chegarem a um resultado semelhante, são várias as possibilidades para a definição do critério de limiar a ser usado na binarização. Algumas técnicas, como veremos a seguir, se utilizam de informações obtidas a partir de histogramas, outras usam informações locais para mudar dinamicamente o valor do limiar, só para exemplificar. Cada algoritmo possui peculiaridades que o tornam mais eficiente em alguns casos, porém mais custoso em outros. Este capítulo apresenta algumas técnicas utilizadas para binarização de imagens de documentos, que servirão de base de comparação para verificar a qualidade das imagens produzidas pelos algoritmos implementados neste trabalho.

3.1 LIMIARIZAÇÃO A PARTIR DA MAXIMIZAÇÃO DE BRINK

A limiarização através do uso do algoritmo de Brink, descrito em [4], utiliza um conjunto de equações, denominadas de critério de correlação de limiar de Brink para estabelecer dois limiares para a imagem, T_1 e T_2 , sendo $T_1 < T_2$, que são utilizados para se efetuar uma limiarização global na imagem. Durante a varredura, todos os pixels da imagem são classificados, de acordo com o valor da sua intensidade I , em comparação com os limiares encontrados, criando-se assim, três conjuntos distintos:

Na Figura 3.1, é possível perceber que os valores calculados pelo critério de Brink dividem o histograma da imagem em três setores. Os pixels com intensidade abaixo do valor de T_1 (representado pela linha vermelha no histograma) são classificados como integrantes do conjunto um. As intensidades intermediárias entre T_1 e T_2 compõem o grupo dois. E os pixels com intensidade acima de T_2 (linha verde) estão no grupo três. Os pixels dos conjuntos um e três pode ter suas intensidades diretamente alteradas para os valores máximo (branco) ou mínimo (preto), enquanto os pixels do conjunto dois não estão claramente definidos, se fazendo necessário uma análise local. Neste caso, sobre cada pixel desse conjunto, é vistoriada uma janela de vizinhança bidimensional quadrada, de tamanho fixo vinte e cinco, à procura de algum pixel que esteja dentro da janela e que faça parte do conjunto três. Se a janela contém um pixel do conjunto três, então a intensidade do pixel vistoriado (que pertence

originalmente ao grupo dois e é o pixel central da janela) será máxima (branco), caso contrário, será mínima (preto).

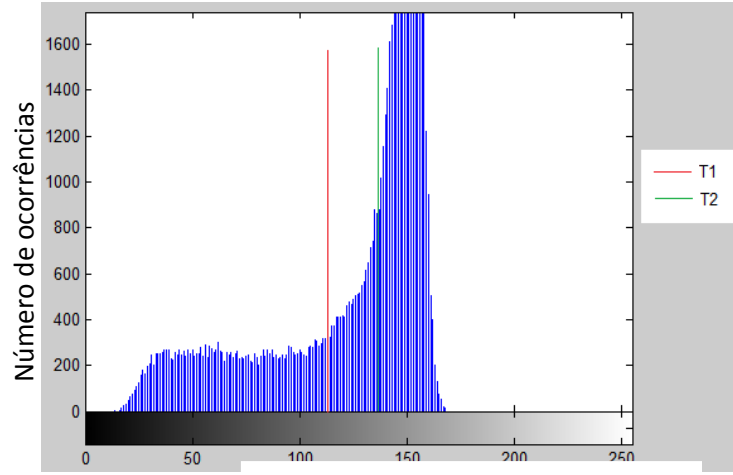


Figura 3.1: Limiares de Brink (T1 e T2) sobre um histograma de uma imagem. Conjunto um: contém os pixels que possuem valor de intensidade menor que T1. Conjunto dois: contém os pixels da imagem com valor de intensidade intermediário, ou seja, $T1 < I < T2$. Conjunto três: contém os pixels que possuem valor de intensidade maior que T2.

3.2 LIMIARIZAÇÃO FUZZY C MEANS

De acordo com [5], a limiarização Fuzzy C Means é uma técnica que combina análise de histogramas com probabilidades para se estabelecer o valor do limiar que será utilizado na binarização. Para cada nível de cinza no histograma, são calculados os valores intermediários μ_o e μ_b , (Equações 3.1) que servem de base para o cálculo de duas probabilidades iniciais, de acordo com as Equações 3.2: a probabilidade ρ_o de se encontrar o nível de cinza em questão em um elemento de fundo, e a probabilidade ρ_b de se encontrar o nível de cinza em um elemento de texto.

$$\mu_o = \frac{\sum_0^t j \cdot h(j)}{\sum_0^t h(j)} \quad \mu_b = \frac{\sum_{t+1}^{255} j \cdot h(j)}{\sum_{t+1}^{255} h(j)} \quad (3.1)$$

$$\rho_o = \frac{\mu_o}{\mu_o + \mu_b} \quad \rho_b = \frac{\mu_b}{\mu_o + \mu_b} \quad (3.2)$$

Com as probabilidades iniciais definidas, a cada iteração, deve-se calcular o nível médio de tom de cinza para o texto e para o *background*, a partir do histograma, bem como atualizar o valor das probabilidades, até que a diferença entre elas seja considerada insignificante, ou seja, abaixo de um valor pré-estabelecido. O valor do limiar que será utilizado para binarizar o documento é calculado a partir da média aritmética dos valores de nível médio dos tons de cinza encontrados para o texto e para o *background*, ao final das iterações.

3.3 LIMIAÇÃO POR ENTROPIA DE JOHANNSEN

De acordo com o método de Johannsen [6], para cada valor de tom de cinza T presente na imagem, são calculados dois valores de entropia, de acordo com as Equações 3.3, que representam a entropia dos pixels de intensidade máxima (S_w) e dos pixels de intensidade mínima (S_b).

$$S_b(t) = \log\left(\sum_{i=0}^t p_i\right) + \frac{1}{\sum_{i=0}^t p_i} \cdot \left[E(p_t) + E\left(\sum_{i=0}^{t-1} p_i\right) \right] \quad (3.3)$$

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

O limiar será escolhido como sendo o valor T para o qual a entropia total da imagem, entendida como a soma das duas entropias, S_w e S_b , calculadas anteriormente, tenha valor mínimo.

3.4 LIMIAÇÃO POR ENTROPIA DE WULU

O método de limiarização por entropia de Wulu, como descrito em [11] e [15] tem seu estudo baseado em figuras geradas a partir de aparelhos de ultrassom, que apresentam imagens com pouco contraste onde, normalmente, a informação encontra-se misturada ao fundo e a ruídos inerentes à imagem. Assim, como a maioria das técnicas de entropia, a

técnica de WuLu tem por objetivo minimizar a diferença entre a entropia do objeto e a entropia do fundo.

3.5 LIMIAÇÃO ATRAVÉS DA PROJEÇÃO DO DISCRIMINANTE LINEAR DE FISHER SOBRE O HISTOGRAMA BIDIMENSIONAL

Normalmente, os algoritmos que utilizam informações de histograma para extrair informações de limiarização se baseiam apenas na projeção unidimensional dos níveis cinza da imagem. Entretanto, em [7], os autores utilizam o método de projeção do discriminante linear de Fisher, que faz uso de um histograma bidimensional para obter mais informações a respeito da imagem.

Para a construção do histograma bidimensional, para cada pixel da imagem é calculado o valor médio do nível de cinza da vizinhança do pixel em questão. Isso cria um mapeamento entre o nível de cinza do pixel e o nível médio de sua vizinhança. O conjunto de todos os pixels mapeados, definido pelos autores como “palavra”, cria o histograma utilizado para se extrair a informação de limiarização.

Assumindo que, na Equação 3.4, a função $g(x, y)$ representa o nível médio local sobre uma janela quadrada de dimensão n , podemos obter o histograma médio local, que vai registrar a ocorrência local média de um nível de cinza, a partir da projeção de ocorrência local média do histograma bidimensional. A partir do histograma local, determina-se um parâmetro de limiarização T , aplicando-se alguma técnica de limiarização comum (como o método de Otsu [16], por exemplo) sobre o histograma. O parâmetro T é o limiar da função de limiarização média local, de acordo com a equação 3.5.

$$g(x, y) = \frac{1}{n^2} \sum_{s=-n/2}^{n/2} \sum_{t=-n/2}^{n/2} f(x + s, y + t) \quad (3.4)$$

$$f_T(x, y) = \begin{cases} 0, & \text{se } g(x, y) < T \\ 1, & \text{se } g(x, y) \geq T \end{cases} \quad (3.5)$$

3.6 LIMIAÇÃO ITERATIVA DE YE-DANIELSSON

A técnica de limiarização iterativa de Ye-Danielsson [8] calcula um valor de limiar diferente a cada pixel processado. Para entendermos a técnica, devemos apresentar o algoritmo de Ridler-Calvard-Thrussel.

Este algoritmo faz uso de um histograma unidimensional dos níveis de cinza da imagem. A partir do histograma, uma sequência de valores é calculada a fim de se obter o limiar de Ridler-Calvard-Thrussel, tendo início a partir do valor do pixel de interesse t_0 , definido em função dos valores presentes no histograma, de acordo com a Equação 3.6.

Note que o limite superior do somatório na Equação 3.6 para o cálculo de t é o valor máximo para a intensidade de um pixel em uma imagem de tons de cinza de oito bits. O valor de t é utilizado como limite superior do somatório utilizado para os cálculos de outras duas variáveis, μ_t e v_t definidas de forma semelhante à variável t . O valor do limiar de Ridler-Calvard-Thrussel é, então, calculado como definido na Equação 3.7.

A cada novo pixel processado, os valores de t , μ_t e v_t são recalculados até que o valor do limiar se repita entre duas ou mais iterações consecutivas (convergência de valores).

A técnica de Ye-Danielsson utiliza o limiar final encontrado pelo algoritmo de Ridler-Calvard-Thrussel como valor inicial. A partir dele, calcula-se o limiar a partir de uma equação do segundo grau, que depende das variáveis t , μ_t e v_t , e tem raízes conhecidas. O limiar final da técnica de Ye-Danielsson é a raiz positiva da equação. No caso de Ye-Danielsson, o limiar é recalculado até que ocorra convergência de valores, ou ocorra uma raiz imaginária. Neste último caso, o limiar utilizado será a raiz real positiva calculada na iteração anterior.

$$t_0 = \frac{\sum_{i=0}^{255} i \cdot h(i)}{\sum_{i=0}^{255} h(i)} \quad (3.6)$$

$$t = \frac{\mu_t + v_t}{2} \quad (3.7)$$

3.7 LIMIAZIZAÇÃO FUZZY DE YAGER

A técnica de limiarização de Yager [9] utiliza lógica *Fuzzy* para discriminar a probabilidade de um pixel selecionado pertencer ao objeto ou ao *background*. Para tanto é necessário definir os valores de nível de cinza dos objetos (μ_o) e do fundo (μ_b), e classificar todos os pixels da imagem, dependendo do valor de nível de cinza do pixel selecionado. Se um pixel estiver na classe de pixels de *foreground*, a diferença entre o seu nível de intensidade e o nível de intensidade de referência da classe será menor do que a diferença do seu nível para o nível de referência da classe *background*, fato que se reflete no cálculo da probabilidade μ_x , definida pela Equação 3.8, onde t é o valor de limiar, e C é a diferença entre o máximo e o mínimo nível de cinza. Para um pixel pertencer à classe dos objetos ($X > t$), a sua probabilidade $\mu_x(X)$ deve ser entre $\frac{1}{2}$ e 1.

Calculada a probabilidade da função, faz-se necessário calcular o *fuzzyness*, com valor máximo limitado à unidade. Uma forma de medição é utilizar a função de Shannon para cálculo da entropia. O limiar utilizado para a limiarização é o valor de t que proporcione o menor valor de *fuzzyness*.

$$\mu_x(g) = \begin{cases} \frac{1}{1 + |g - \mu_b|/C}, & \text{se } g \leq t \\ \frac{1}{1 + |g - \mu_o|/C}, & \text{se } g > t \end{cases} \quad (3.8)$$

3.8 LIMIAZIZAÇÃO ITERATIVA DE ROSENFELD POR RELAXAÇÃO

O método de relaxamento de Rosenfeld, apresentado em [10], é um método iterativo em que o limiar é calculado como uma função de acordo com a vizinhança da iteração anterior, a partir de uma configuração inicial, chamada de configuração de confiança.

Uma possibilidade de se estabelecer uma configuração de confiança é estabelecer o valor médio do nível de cinza da imagem como fator de comparação. Logo, são definidas as

probabilidades iniciais de configuração para cada pixel da imagem, classificando-os entre “possivelmente brancos” ou “possivelmente pretos”, a depender do valor da sua intensidade em relação ao valor médio obtido.

Após a classificação inicial, é necessário verificar a compatibilidade desta classificação com a vizinhança local do pixel em questão. Por exemplo: se um pixel é preto e toda a sua vizinhança é branca, essa configuração sugere que o pixel preto deveria ser branco, já que a compatibilidade deste pixel com sua vizinhança é muito baixa.

A compatibilidade de uma região ao redor de um pixel i pode ser definida como a compatibilidade C média de seus oito vizinhos, onde C é uma função de compatibilidade qualquer, como, por exemplo, a média dos pixels de uma classe. A classificação dos pixels através de sua compatibilidade com a vizinhança frequentemente está associada a um par de probabilidades p e q , que são modificados a cada iteração, e correspondem à probabilidade do pixel ser classificado como *foreground* ou *background*. A Equação 3.8 ilustra o cálculo da probabilidade para pixels do *foreground*. Uma equação semelhante é utilizada para o *background*. Quando não ocorrerem alterações adicionais em passos subsequentes, ou quando as probabilidades atingirem um valor estabelecido como limite, o pixel é, então, classificado.

$$p_i^{k+1} = \frac{p_i^k (1 + Q_i^k)}{p_i^k (1 + Q_i^k(\text{branco})) + q_i^k (1 + Q_i^k(\text{preto}))} \quad (3.9)$$

3.9 LIMIAZIZAÇÃO BIMODAL DE OTSU

O método de limiarização de Otsu [16] é uma técnica bastante utilizada na binarização de imagens, por apresentar, sob certas circunstâncias, um alto grau de eficiência e performance. A técnica utiliza métodos estatísticos e análise de discriminante para classificar os pixels da imagem de níveis de cinza em duas classes, C_0 e C_1 , que mais tarde serão convertidas em preto e branco. Durante o processo de cálculo do limiar de Otsu, porém, várias outras informações que podem ser relevantes são fornecidas, como as porções das áreas ocupadas pelas classes C_0 e C_1 , ou a estimativa dos níveis médios das classes na imagem original em níveis de cinza, por exemplo.

Além de prover informações adicionais ao limiar, o método de Otsu tem por principal característica o fato de ser um método simples e de uso genérico. A eficiência do método, contudo, está diretamente relacionada com o nível de uniformidade entre os objetos e o *background*, já que o limiar ótimo é escolhido de forma automática, utilizando-se as propriedades globais do histograma.

3.10 LIMIARIZAÇÃO DE NIBLACK

O algoritmo de Niblack [18] é abordado rapidamente em [19] e é utilizado como base para o desenvolvimento do algoritmo de Sauvola em [17]. Interessante também observar a comparação entre os algoritmos de Niblack e Sauvola, proposta por [20]. Em sua proposta original, o algoritmo de Niblack calcula o valor do limiar a partir de uma janela deslizante retangular, de tamanho $n \times n$ sobre a imagem em tons de cinza. O limiar calculado é uma função da média e do desvio padrão de todos os pixels dentro da janela, centrada no pixel para o qual se deseja calcular o limiar. O funcionamento adequado da binarização depende do tamanho da janela retangular e da influência do desvio padrão sobre o cálculo, controlada a partir do valor da constante k (chamada de *bias*), na Equação 3.10.

$$T = m + k.s \quad (3.10)$$

3.11 LIMIARIZAÇÃO ADAPTATIVA HÍBRIDA DE SAUVOLA

O método de limiarização desenvolvido por Sauvola e Pietaksinen, definido em [17] e comentado em [19] é uma abordagem híbrida para a limiarização de imagens em tons de cinza. Para cada pixel processado, duas abordagens diferentes são aplicadas, a fim de se calcular o valor do limiar utilizado. Um método que visa binarizar as figuras e *background*, chamado de *soft decision method* (SDM) e outro para os objetos de texto encontrados na imagem, chamado de *text binarization method* (TBM). O resultado combinado dos dois algoritmos produz o limiar utilizado para o pixel processado.

A abordagem inicial do algoritmo envolve a análise de uma superfície local da imagem, no intuito de se determinar qual algoritmo deve ser utilizado primeiro para o pixel

em questão. A partir dessa análise, a imagem é então dividida em setores, onde cada setor será binarizado com o algoritmo mais propício (SDM para figuras ou TBM para textos). No caso específico do TBM, o autor utiliza uma versão modificada do algoritmo de Niblack [18], que originalmente utiliza a média e o desvio padrão computada em uma pequena janela de vizinhança. Na modificação proposta, o desvio padrão tem influência maior no cálculo do limiar, que, segundo [17], melhora a atuação do algoritmo sobre *background* de imagens com nível de cinza constantemente maior do que o limiar estabelecido.

A interpolação dos resultados dos algoritmos para produzir a imagem limiarizada final pode ser feita de duas formas. Uma opção mais simples seria computar o limiar para um conjunto definido e igualmente espaçado de pixels e usar interpolação para os demais pixels que não tiveram seu limiar diretamente calculado. Assim, diminuindo o custo computacional, tem-se um resultado satisfatório. Outra possibilidade é se utilizar a média de n pixels candidatos (pixels que tiveram seus limiares calculados) e utilizar essa média diretamente como limiar nos pixels restantes.

3.12 LIMIARIZAÇÃO POR *CONVEX HULL*

O método de limiarização por *Convex Hull*, descrito brevemente em [19], traduzido como “casca convexa” é um método de limiarização a partir de comparação entre as concavidades presentes no histograma da imagem, e uma “cobertura” construída sobre o histograma, que serve como parâmetro de comparação. A casca cria, para cada ponto do histograma, uma informação adicional de distância entre as concavidades do histograma e a casca. O limiar escolhido será o ponto do histograma para o qual a distância do seu vale até a casca seja a maior possível.

3.13 LIMIARIZAÇÃO POR ENTROPIA DE PUN

O método de limiarização por entropia proposto por [13] é mais um método baseado no estudo da entropia da imagem para se obter um limiar adequado para efetuar a binarização de uma imagem com oito bits de profundidade. Para tal, o método faz uso da entropia dos

pixels pretos (Equação 3.12) e brancos (Equação 3.13) para encontrar um valor de limiar t que maximiza entropia total da imagem (Equação 3.11).

$$H = H_b + H_w \quad (3.11)$$

$$H_b = - \sum_{i=0}^t p_i \log p_i \quad (3.12)$$

$$H_w = - \sum_{i=t+1}^{255} p_i \log p_i \quad (3.13)$$

onde p_i é a probabilidade de ocorrência do i -ésimo nível de cinza.

3.14 OUTRAS TÉCNICAS

3.14.1 Nível de Cinza Médio

A técnica de limiarização através do nível de cinza médio (Mean Gray Level) é uma técnica bastante antiga e simples, mas com razoável eficácia. O limiar utilizado para binarizar a imagem é o nível de cinza obtido através da média da intensidade de todos os pixels presentes na imagem.

3.14.2 Porcentagem de Preto

A técnica de limiarização por percentual de preto utiliza um valor de entrada como base para calcular o valor do limiar. O valor de entrada é o percentual de preto que se espera obter na imagem binarizada ao final do processo, e o limiar utilizado é aquele capaz de produzir uma imagem que mais se aproxime do percentual fornecido.

4. ALGORITMOS DE BINARIZAÇÃO PARA IMAGENS DE DOCUMENTOS COM BACKGROUND COMPLEXO

Este capítulo está dividido em duas partes, sendo que cada uma delas está relacionada a um dos algoritmos de binarização de imagens de documentos que foram implementados durante este estudo. Cada parte é composta por subdivisões, destacando o passo-a-passo do desenvolvimento do algoritmo correspondente.

4.1 BINARIZAÇÃO DE IMAGENS DE DOCUMENTOS UTILIZANDO MÁXIMOS E MÍNIMOS LOCAIS

Esta técnica, descrita em [21] e [22], analisa o *background* da imagem, criando uma imagem de contraste a partir da imagem do documento original e utilizando essa nova imagem para detectar os pixels de alto contraste dentro de uma janela local, que ocorrem com maior frequência nas bordas do texto. Em seguida são aplicados limiares locais, estimados a partir dos pixels de alto contraste encontrados. As próximas seções dão mais detalhes sobre os procedimentos.

4.1.1 Análise do *Background* da Imagem

Em imagens de documentos, a análise de *background* consiste em destacar o conteúdo do texto em relação aos ruídos de fundo. A análise do *background* da imagem é um passo importante para a correta binarização da imagem. Separar o que é informação relevante e o que é ruído de *background* facilita o trabalho dos métodos de binarização. Por exemplo, um método bastante comum para análise do *background* envolve a extração de informação de bordas a partir de uma imagem gradiente criada a partir da convolução da imagem original com um filtro. Um dos filtros mais simples é o filtro de Sobel, utilizado para detecção de bordas, verticais e horizontais. Cada pixel na imagem gradiente representa a mudança de intensidade no mesmo ponto na imagem original, em relação ao pixel anterior em uma dada direção – horizontal ou vertical. Quanto mais alto o valor do pixel na imagem de contraste, maior é a probabilidade de ele ser considerado um pixel de borda, por tanto, um pixel com informação relevante. Apesar de robusto, a detecção de bordas pela imagem gradiente é

bastante sensível à variação de brilho e contraste da imagem, já que, para a construção da imagem gradiente, leva em consideração somente a diferença absoluta entre pixels dentro de uma janela de vizinhança [2].

O filtro de Sobel, como já citado, é apenas uma dentre várias possíveis estratégias utilizadas para analisar o *background* de imagens. A técnica descrita em [21] e [22], por exemplo, não faz uso da imagem de gradiente. Para a análise de *background*, é criada uma imagem de contraste calculada a partir dos pixels com valor máximos e mínimos dentro de uma janela local na imagem fonte. A janela sugerida para a obtenção destes valores é uma janela quadrada, de tamanho três, centrada no pixel sob avaliação. Após a análise da janela, é possível computar o valor de intensidade do pixel correspondente na imagem de contraste calculando o valor da Equação 4.1.

$$D(x, y) = \frac{f_{\max}(x, y) - f_{\min}(x, y)}{f_{\max}(x, y) + f_{\min}(x, y) + \epsilon} \quad (4.1)$$

A Equação 4.1 define o cálculo do valor do pixel de contraste usando máximo e mínimo local. (x, y) são as coordenadas de um pixel $P(i, j)$ em processamento. $D(i, j)$ é o valor do pixel $P(i, j)$ na imagem de contraste, $f_{\max}(x, y)$ e $f_{\min}(x, y)$ são os valores de intensidade máxima e mínima dentro da janela de vizinhança de $P(i, j)$. ϵ é um valor infinitesimal, apenas para impedir que ocorra divisão por zero.

Observa-se que o numerador da Equação 4.1 é similar ao gradiente da imagem, como definido no capítulo introdutório deste trabalho. Porém, o denominador age como um fator de normalização, para amenizar as interferências do brilho e contraste que afetam a imagem gradiente. À medida que o gradiente aumenta, ou seja, ocorreu uma mudança de intensidade brusca, o denominador também aumenta, fazendo com que o valor final da equação esteja normalizado em relação ao valor do gradiente.

A Figura 4.1 - esquerda representa uma imagem de teste sob forte influência de ruído. Já a Figura 4.1 - direita é uma imagem de contraste, criada a partir da imagem de teste para ilustrar o efeito da imagem de contraste para detecção das bordas do texto.

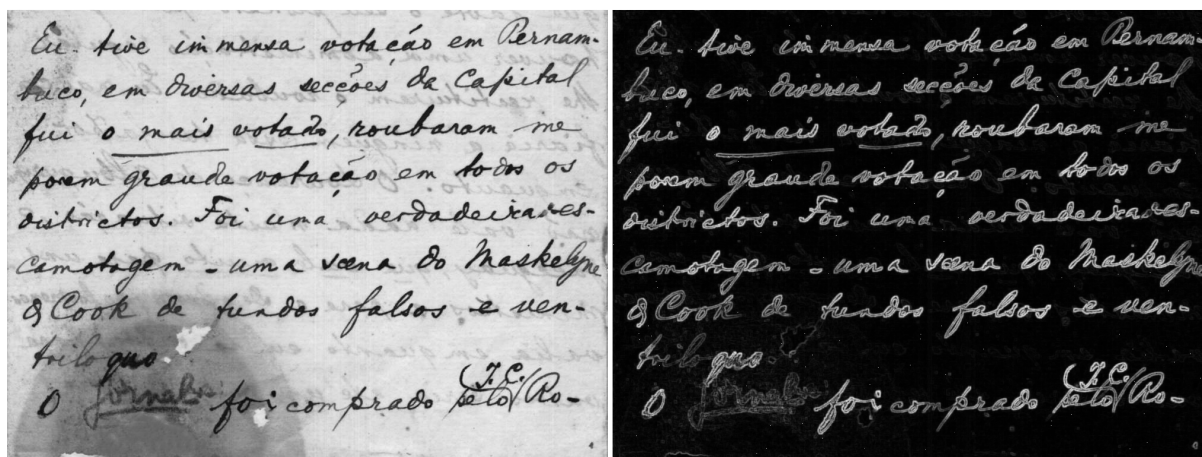


Figura 4.1: (esquerda) Imagem de um documento manuscrito sob forte influência de ruído e (direita) a correspondente imagem de contraste gerada pelo algoritmo.

4.1.2 Seleção dos Pixels de Alto Contraste

Com a imagem de contraste devidamente construída, a detecção dos pixels de alto contraste pode ser efetuada de maneira adequada. Visualmente, estes pixels se encontram, preferencialmente, nas bordas do texto, não representando, por enquanto, a imagem devidamente binarizada, apesar de servirem como base de cálculo para geração da imagem final. O processo de detecção de pixels de alto contraste envolve, como já citado, o cálculo do valor numérico da Equação 4.1 para todos os pixels da imagem a ser binarizada e, como naturalmente deveria ser, não é um processo imune a erros. De fato, a ocorrência de pixels de alto contraste indesejáveis aumenta de acordo com o tamanho da fonte utilizada no texto presente na imagem. Como demonstrado na Seção 4.1.5 deste trabalho, pixels com alto valor de contraste fora da borda do texto apresentam uma dificuldade a mais para a correta limiarização da imagem.

Os pixels de alto contraste são gerados aplicando-se um método de limiarização global sobre a imagem de contraste. Neste trabalho, para tal tarefa, foi utilizado o método de limiarização de Otsu [16]. A imagem binarizada deverá conter todos os pixels de borda de texto presentes na imagem do documento original. (Figura 4.2 - direita)

É interessante observar que, apesar de já se apresentar como uma imagem binarizada, a Figura 4.2 - direita não é a imagem limiarizada resultante do algoritmo. Ela representa os

pixels de alto contraste presentes na imagem, destacados na imagem de contraste e que serão utilizados para o cálculo dos pixels de texto na imagem final. Apesar de também utilizar o método de limiarização global de Otsu, a imagem binária gerada a partir da imagem de contraste apresenta significativas diferenças em relação a uma imagem de Otsu gerada a partir da imagem original, por oferecer mais importância aos pixels de borda. A Figura 4.2 ilustra tais diferenças, visualmente.

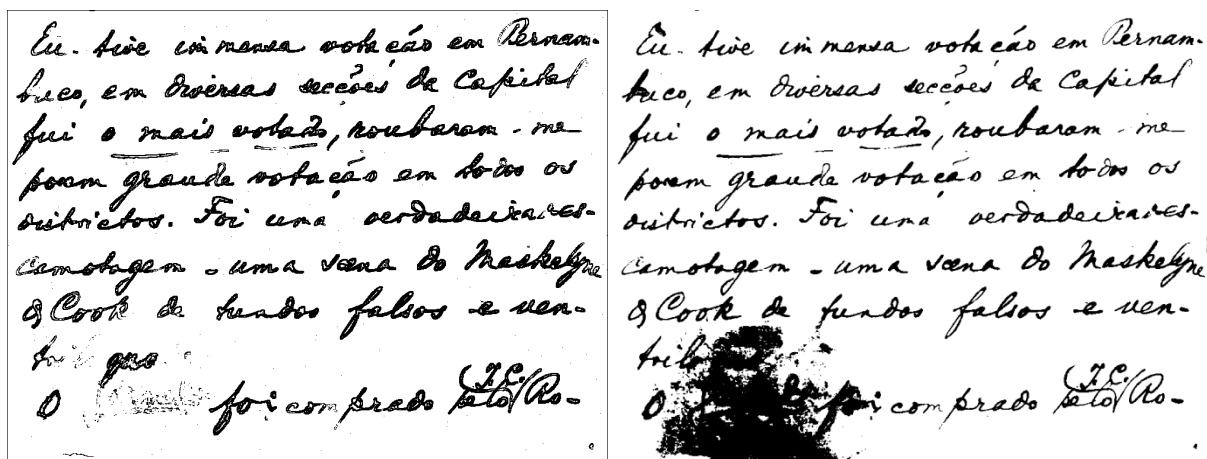


Figura 4.2: (esquerda) Resultado da limiarização de Otsu, a partir da imagem de contraste da Figura 4.1 e (direita) resultado da limiarização de Otsu obtida a partir da imagem original.

4.1.3 Definição dos Parâmetros para Limiarização Local

A estimativa da largura do texto é um passo importante para a binarização correta da imagem, pois ela determina o tamanho da janela que será utilizada nos passos seguintes do algoritmo. Um erro aqui pode comprometer o resultado da binarização como um todo, já que uma janela de tamanho inadequado pode comprometer a limiarização local, que pode classificar erroneamente os pixels de texto como pixels de *background*.

A estimativa é feita em cima da imagem de contraste, utilizando-se uma janela unidimensional de tamanho mínimo três. A partir do janelamento, os pixels de maior contraste são localizados linha por linha, permitindo gerar um histograma (Figura 4.3) com registro das distâncias entre as posições dos pixels selecionados. A largura do texto é estimada a partir da análise deste histograma, já que o valor de distância com maior ocorrência deverá

corresponder, idealmente, à distância entre as duas bordas do texto, que corresponde à largura estimada.

Por ser um processo global de análise, o processo de estimativa da largura do texto, como demonstrado na Seção 4.1.5 deste capítulo, não oferece garantias de que a largura estimada de fato corresponde à largura real do texto presente na imagem.

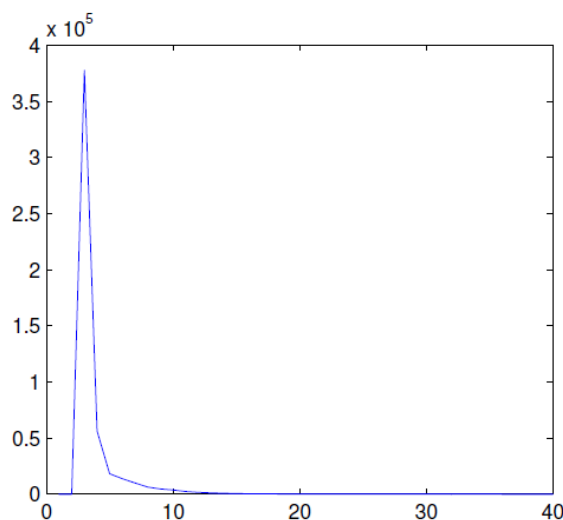


Figura 4.3: Exemplo de histograma que mapeia as ocorrências mais frequentes das distâncias entre pixels de alto contraste. (retirado de [21])

Com a largura do texto determinada, dois parâmetros importantes para a correta limiarização podem ser definidos: o tamanho da janela de limiarização e o limiar de números de borda. No caso do tamanho da janela, geralmente, a limiarização local do documento não sofre interferência se o tamanho da janela utilizada for maior que o real tamanho do texto. Porém, o mesmo não ocorre nos casos onde a janela tinha tamanho menor do que a largura real do texto. Como demonstrado na Seção 4.1.5, o resultado da binarização pode ser comprometido.

Nos testes realizados, o tamanho da janela de limiarização foi o dobro da largura de texto estimada. O limiar de números de borda que é utilizado para se obter a imagem final é definido como o número mínimo de pixels de alto contraste encontrados dentro de uma janela bidimensional. O tamanho da janela é o mesmo utilizado para a limiarização do documento. Como estamos interessados apenas no número de pixels de borda, esse número pode ser obtido a partir da imagem gerada pela binarização de Otsu, que, como já foi exposto, deverá conter apenas os pixels de borda de texto.

4.1.4 Limiarização de Imagens de Documentos

Uma vez que todos os parâmetros foram definidos – tamanho da janela de limiarização e o limiar de pixels de borda – os pixels da imagem do documento podem ser classificados em pixels de texto, ou pixels de *background*. Para tanto, é necessário estabelecer alguns valores estatísticos a respeito do pixel processado em relação à sua vizinhança na imagem original e na imagem de Otsu. As Equações 4.2 e 4.3 descrevem a média e o desvio padrão da intensidade dos pixels de alto contraste detectados a partir da imagem gerada pelo algoritmo de Otsu.

$$E_{médio} = \frac{\sum_{vizinhança} I(x, y) \times [1 - E(x, y)]}{N_e} \quad (4.2)$$

$$E_{dp} = \sqrt{\frac{\sum_{vizinhança} \{[I(x, y) - E_{médio}] \times [1 - E(x, y)]\}^2}{N}} \quad (4.3)$$

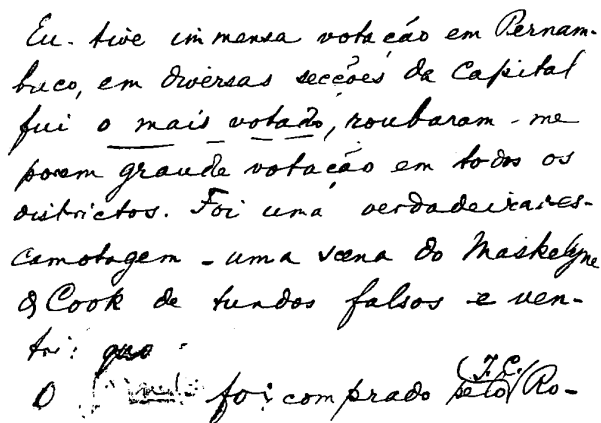
Nas Equações 4.2 e 4.3, E_{dp} representa o valor do desvio padrão, enquanto $E_{médio}$ representa o valor médio da intensidade dos pixels na janela de vizinhança ao redor do pixel $P(i, j)$. $I(i, j)$ corresponde ao valor da intensidade nesse mesmo pixel na imagem original, enquanto $E(i, j)$ corresponde ao mesmo valor na imagem gerada por Otsu. N e N_e representam o número de pixels, total e de alto contraste, localizados dentro da janela de vizinhança.

Com os valores da média e desvio padrão, podemos classificar os pixels da imagem original a partir da Equação 4.4:

$$R(x, y) = \begin{cases} 0, & N_e \geq N_{\min} \ \& \ I(x, y) \leq E_{médio} + E_{dp}/2 \\ 1, & \text{em caso contrário} \end{cases} \quad (4.4)$$

Em 4.4, $N_{\min}$ é o limiar de pixels de borda determinado anteriormente e $R(i, j)$ é o valor de $P(i, j)$ na imagem final. Note que a relação especifica o *background* como branco (intensidade nula) e o texto como preto (intensidade unitária). O resultado da aplicação deste

método, que corresponde à imagem final produzida pelo algoritmo, sobre a imagem da Figura 4.1 - esquerda pode ser visto na Figura 4.4.



Eu. tive imensa votação em Bernam-
buc, em diversas secções da Capital
fui o mais votado, roubaram-me
porém grande votação em todos os
districtos. Foi uma verdadeira res-
camotagem - uma scena do Mackelne
& Cook de fundos falsos e ven-
to: que
O ~~qual~~ foi comprado pelo Ro-

Figura 4.4: Resultado da limiarização utilizando máximos e mínimos locais

4.1.5 Impacto da Largura do Texto na Limiarização

A eficácia do método, como dito anteriormente, depende fundamentalmente da correta determinação da largura do texto presente na imagem. Esse efeito se torna crítico quando a imagem sobre a qual se deseja obter a binarização apresenta elementos de texto com tamanhos diferentes. A Figura 4.5 - esquerda foi utilizada para ilustrar o impacto da largura estimada do texto sobre a limiarização.

Observando os caracteres presentes, pode-se notar que existem vários caracteres de tamanhos diferentes na imagem. Como a análise do histograma da imagem de bordas faz uma análise global, a largura do texto estimada para a imagem inteira possui um valor que nem sempre corresponde ao valor real da largura da letra em alguns pontos do texto. De fato, para a imagem em questão, a largura escolhida teve um valor intermediário em relação às larguras presentes na imagem.



Figura 4.5: (esquerda) Aplicação do método em uma imagem com a presença de textos de tamanhos diferentes e (direita) imagem de contraste correspondente.

Para a Figura 4.5, a largura do texto estimada (*Stroke Width*), considerando todos os tipos de texto encontrados na imagem foi igual a seis. A imagem limiarizada resultante do algoritmo e o correspondente *Ground Truth* estão ilustrados na Figura 4.6.



Figura 4.6: (esquerda) Imagem limiarizada resultante da aplicação do algoritmo e (direita) *Ground Truth* correspondente.

Comparando as Figuras 4.6 (esquerda) e (direita) visualmente, pode-se perceber que os textos que apresentam tamanho de fonte menor, e por tanto, mais próximo do valor do *Stroke Width* estimado, foram mais facilmente identificados como texto do que as letras com fonte maior, fato determinante para o resultado final da limiarização.

Observando cuidadosamente a distribuição dos pixels de alto contraste na imagem criada a partir da Figura 4.6 - esquerda, é possível perceber, numericamente, a diferença entre os tamanhos utilizados. Podemos dividir a figura em diferentes setores e analisar o comportamento do algoritmo de estimativa da largura do texto em cada setor, e comparar a

sua estimativa com um valor de largura obtido manualmente através de qualquer editor de imagens. Trata-se, contudo, de uma comparação grosseira, uma vez que apenas alguns pontos da imagem são analisados, não caracterizando esses valores como tamanho real da fonte utilizada. O intuito da amostragem é de apenas servir como base para encontrar possibilidades para melhorias no algoritmo.

Observando a Figura 4.7 - esquerda, que é uma parte da imagem de alto contraste (Figura 4.5 - direita) é possível perceber que as bordas das letras foram, de fato, corretamente identificadas. Contudo, existem manchas (ruídos) na imagem, tanto na região entre as letras, como no interior das bordas de cada letra. De fato, a presença desses pontos de ruído tem influência direta na estimativa do *Stroke Width*, já que, como apresentado na Seção 4.1.3 deste capítulo, a estimativa da largura do texto é feita varrendo-se a imagem, linha por linha, mapeando as distâncias entre os pontos de alto contraste e identificando a distância mais frequente na imagem. Ruídos, como os presentes na Figura 4.5 - direita podem comprometer essa análise, se não forem ignorados, como destacado na Seção 4.1.6 deste trabalho.

Para a Figura 4.5 - esquerda, a largura do texto estimada pelo algoritmo esteve próxima da média da largura dos textos menores, que apresentam largura variando entre três e dez pixels na imagem original, porém muito distante da largura do texto presente na Figura 4.7, que varia entre vinte e seis e trinta e três pixels. Como destacado na Seção 4.1.3, a largura do texto é vital para a determinação do tamanho da janela a ser utilizada para enquadrar o texto. Não ocorrem problemas se o tamanho da janela for maior que a largura do texto real, por isso, os textos menores são enquadrados normalmente, mesmo possuindo largura menor do que a estimada. O problema do texto na Figura 4.7 - esquerda é que a sua largura é maior do que a largura estimada para o texto, e assim, também é maior do que a janela utilizada para limiarização da imagem, resultando em uma classificação inadequada dos pixels que estão na região interior de cada letra do texto presente na imagem, que são erroneamente classificados como *background*, como ilustrado na Figura 4.8 - esquerda.

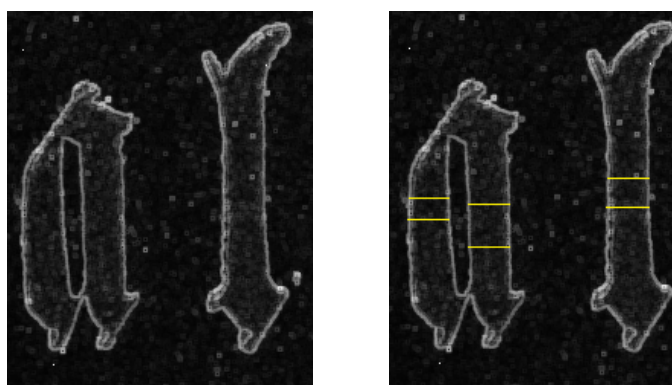


Figura 4.7: (esquerda) Imagem de pixels de alto contraste utilizada para estimar a largura do texto e (direita) largura do texto medida manualmente a partir de algumas amostras de largura.

Se, a partir de algumas amostras de largura, como na Figura 4.7 - direita, forçarmos o algoritmo a utilizar uma largura de texto mais adequada para a largura da letra utilizada no texto da Figura 4.5 - esquerda, ao invés de usar a largura estimada, o resultado da limiarização muda consideravelmente para o caso do texto com letras de tamanho maior do que a largura estimada. No exemplo ilustrado pela Figura 4.7 - direita, a largura estimada manualmente possuía valor entre vinte e oito e trinta e três, em média. O valor “forçado” utilizado como largura de texto para o algoritmo, em substituição ao estimado, foi trinta e dois. Com esse novo valor de largura, mantendo-se todos os demais parâmetros e métodos de cálculo, os pixels internos a cada letra, que foram classificados como *background* na limiarização inicial, agora passaram a ser considerados como pixels de texto, e o resultado da limiarização “forçada” é muito mais próximo do *ground truth* para a imagem, como ilustrado na Figura 4.8 - direita. Isso mostra que apenas mudando o tamanho da janela, é possível corrigir pequenas falhas no algoritmo, sem que isso implique em grandes mudanças.



Figura 4.8: (esquerda) Resultado da limiarização comprometido pela má classificação dos pixels devido ao uso de uma janela de tamanho inadequado e (direita) limiarização da mesma imagem com tamanho de janela calculado manualmente.

4.1.6 Análise de Pixels de Alto Contraste Fora das Bordas do Texto

A finalidade da imagem de contraste, como já foi descrito nas seções anteriores, é dar suporte para o método de limiarização global para detecção dos pixels presentes nas bordas do texto da imagem, que são utilizados para estimativa da largura do texto e definição do tamanho da janela de limiarização. Entretanto, como na Figura 4.5 - direita, é possível perceber que outros pixels na imagem, que não fazem parte da borda, também podem ser classificados como pixels de alto contraste. Isso acontece devido principalmente a não uniformidade do *background* da imagem, permitindo a ocorrência de variações de cor entre pixels vizinhos, que acaba por, durante a criação da imagem de contraste, se refletindo no valor do pixel de contraste.

O efeito da presença de pixels de contraste fora da borda do texto afeta diretamente a limiarização da imagem, já que ele tem influência direta no cálculo da largura estimada do texto, que vai servir de base para o cálculo da janela de limiarização, como descrito nas seções anteriores.

O cálculo dos picos de contraste leva em consideração a diferença entre o maior pixel de contraste presente na janela e o seu vizinho imediatamente anterior. O resultado dessa operação é a intensidade de um pixel de alto contraste se o valor resultante estiver acima de um determinado limiar, chamado de limiar de intensidade de pico, previamente determinado, e a posição do pixel é dada pela posição do maior pixel utilizado na operação. Com essa análise feita por toda a imagem, é possível construir um vetor de pixels de alto contraste, de onde será extraída a distância mais frequente entre dois picos que será utilizada como largura de texto estimada.

Sendo assim, foi possível perceber, experimentalmente, que a escolha adequada do limiar de intensidade de pico definido para esta operação pode também ter influência na limiarização final. As Figuras 4.9 e 4.10 apresentam imagens obtidas a partir da limiarização do algoritmo, fazendo uso de dois limiares de intensidade diferentes, durante a criação do vetor de picos de alto contraste. A razão para tal dependência reside no fato de que considerando um limiar mais baixo, mais pixels serão considerados como “alto contraste”, e

por tanto, mais valores são agregados ao vetor de picos. De maneira que, quando as distâncias entre picos forem tomadas, deve haver mais pixels próximos uns dos outros, favorecendo a ocorrência mais frequente de distâncias menores, que acaba por fazer com que a janela também tenha um tamanho reduzido. Porém, como demonstrado na Seção 4.5, o efeito de janelas menores sobre textos de largura maior é prejudicial à binarização, uma vez que não possibilita o correto enquadramento do texto durante o processo de classificação do pixel, comprometendo o resultado da limiarização.

O limiar de intensidade de pico, durante este trabalho, foi definido como um percentual sobre o valor de intensidade máxima (255 para uma imagem de oito bits, por exemplo). Para ilustrar o efeito desse percentual sobre a binarização, o limiar de intensidade foi fixado em 20% do valor máximo de intensidade, gerando uma janela de tamanho seis e produzindo a imagem final ilustrada pela Figura 4.9. Como o tamanho da janela não foi adequado, já que a largura do texto possui tamanho um pouco maior, é possível notar, na região interna de cada letra do texto na figura, pixels que foram erroneamente classificados como *background*. A mesma imagem, quando processada sobre um limiar de 60%, estimou uma largura de texto de tamanho doze, bem mais próximo do que o texto, com a vantagem de ter sido maior do que a largura real média, resultando na imagem da Figura 4.10. Nela, é possível notar que os erros de classificação dos pixels internos ao texto foram minimizados, e o resultado se aproximou bastante do *ground truth* (Figura 4.11).



Figura 4.9: Resultado da binarização, utilizando limiar de intensidade de pico em 20%.



Figura 4.10: Resultado da binarização, utilizando limiar de intensidade de pico em 60%.



Figura 4.11: *Ground truth* correspondente.

4.2 Limiarização Por Níveis Lógicos Adaptativos

A técnica de limiarização por níveis lógicos, proposta por [23], se propõe a analisar o nível de intensidade do pixel processado, caso a imagem seja livre de ruído, ou o seu valor suavizado, em caso de ruído incidente sobre a imagem, comparando-o com médias locais obtidas a partir janelas centradas em pixels vizinhos selecionados. Para cada pixel, mais de uma comparação é feita para cada pixel processado, e o seu resultado por ser considerado como uma “derivada” para o valor da intensidade. Para tanto, faz-se importante determinar o tamanho da janela de vizinhança bidimensional, bem como a localização dos pixels vizinhos escolhidos para análise. As seções seguintes descrevem estes procedimentos em detalhes.

4.2.1 Definição dos Parâmetros para Limiarização

Assim como no algoritmo de binarização através de máximos e mínimos, descrito na seção anterior, a técnica de níveis lógicos efetua uma análise local para definir o valor da intensidade do pixel na imagem final. Para tanto, a espessura da letra do texto avaliado (*stroke width*) novamente aparece como um parâmetro importante a ser determinado. A proposta original do autor, descrita em [23], faz uma análise do histograma de comprimento de carreira das distâncias entre pixels de maior valor dentro de uma janela unidimensional de para determinar a espessura da letra. A distância com maior frequência no histograma deverá corresponder à espessura da letra utilizada no texto e, por tanto, será o parâmetro utilizado pelo algoritmo. O valor da espessura da letra é importante para o desenvolvimento do algoritmo, já que pode comprometer o processo de análise, por não enquadrar pixels suficientes, ou por incluir informações além do necessário na hora de se efetuarem os cálculos para a definição do valor do pixel.

4.2.2 Análise de Níveis Lógicos

A análise de níveis lógicos original estabelece uma comparação lógica para definir o valor do pixel processado na imagem final. Isto é feito comparando-se o valor da intensidade do pixel processado com médias de vizinhanças locais obtidas previamente. Para tanto, a obtenção da posição central das janelas de vizinhança é um passo fundamental para a correta

implementação do algoritmo, e isso depende fundamentalmente da determinação da espessura do texto, definida no item anterior.

Para uma imagem digital de um documento, com espessura de letra conhecida SW , o primeiro passo para análise de nível lógico envolve selecionar corretamente a posição dos pixels que servirão de referencia central para o processo de janelamento. A disposição dos pixels centrais, em relação ao pixel processado está ilustrada na Figura 4.12.

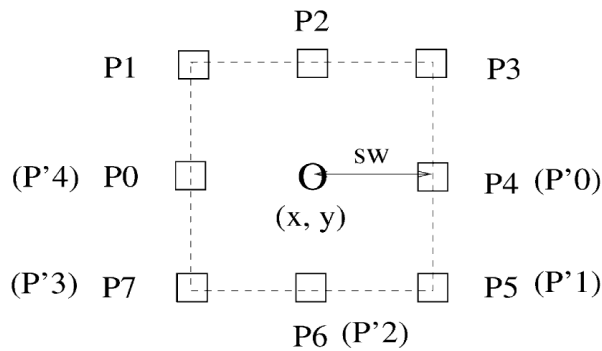


Figura 4.12: Distribuição espacial das vizinhanças escolhidas para comparação, em relação ao pixel processado (retirada de [21]).

O ponto central da figura representa o pixel processado e os oito pixels selecionados são pontos centrais P , distantes do ponto O de pelo menos SW pixels (na horizontal e vertical). Para cada ponto P , faz-se necessário calcular o valor médio dos pixels dentro de uma janela de tamanho $(2.SW + 1)$, centrado em P . Sendo assim, oito valores médios são obtidos na vizinhança do ponto processado. A Equação 4.5 ilustra esse cálculo.

$$ave(P) = \sum_{i=-SW}^{SW} \sum_{j=-SW}^{SW} \frac{f(P_x - i, P_y - j)}{(2 \times SW + 1)^2} \quad (4.5)$$

onde $ave(P)$ é a média calculada para uma janela de tamanho $(2.SW + 1)$ centrada no ponto P escolhido, (P_x, P_y) indicam as coordenadas do ponto P e f é o valor de intensidade do pixel na imagem original.

Para cada valor médio obtido, é efetuada uma comparação com o nível de intensidade do pixel processado, descrita pela Equação 4.6, para se determinar se a diferença é superior ou

não a um determinado parâmetro de binarização T . Esse parâmetro, na técnica original descrita em [24], seria um valor pré-determinado, independente do texto presente na imagem sob análise.

$$L(P) = ave(P) - g(x, y) > T \quad (4.6)$$

onde $g(x, y)$ corresponde ao nível de intensidade do pixel na imagem original, $L(P)$ é a condição lógica que confirma se a diferença entre a média de uma dada janela de vizinhança $ave(P)$ e o valor de g é maior que um parâmetro T , previamente determinado.

Ao final, os pontos de vizinhança são comparados 4 a 4, levando-se em conta sempre dois pontos vizinhos em sentido horário e dois diretamente opostos a estes, em relação ao ponto central processado. Cada combinação de quatro pontos pode ser chamada de grupo de comparação. Um exemplo de grupo de comparação seriam os pontos $\{P_0, P_1, P_4$ e $P_5\}$ da Figura 4.12. Os dois primeiros são vizinhos em sentido horário e os dois últimos estão diametralmente opostos aos primeiros.

Se, em pelo menos um grupo de comparação, a Equação 4.6 é verdade para todos os seus pontos, o pixel processado é marcado como objeto de interesse, ou *foreground*. Do contrário, é marcado como *background*, de acordo com a Equação 4.7.

$$b(x, y) = \begin{cases} 1, & \text{se } \bigvee_{i=0}^3 [L(P_i) \wedge L(P'_{i+1}) \wedge L(P_{i+1}) \wedge L(P'_{i+2})] \text{ é verdade} \\ 0, & \text{caso contrário} \end{cases} \quad (4.7)$$

onde $b(x, y)$ é o valor do pixel na imagem binarizada resultante.

4.2.3 Análise de Níveis Lógicos Adaptativa

Segundo [24], o parâmetro T , utilizado para definir a diferença entre a média local e o valor do pixel processado pode ser definido dinamicamente, a partir de cada uma das janelas processadas. A melhoria proposta calcula um parâmetro específico para cada valor de pixel processado, diminuindo assim a influência de ruídos localizados e mudanças bruscas de *background* sobre a avaliação do valor do pixel processado. O cálculo dinâmico de T leva em

conta o comportamento dos pixels dentro da janela processada para melhor se ajustar à janela atual.

Para tanto, em cada uma das janelas centradas em $P(i, j)$, é necessário determinar os valores dos pixels de intensidade máxima e mínima dentro da janela, bem como a diferença absoluta entre o valor deste pixel, seja ele máximo ou mínimo, e o valor da média da janela obtida. Se a diferença entre o pixel de valor máximo e a média for maior do que a diferença entre o pixel de valor mínimo e a média, significa que a média de pixels naquela janela tende para um valor mais baixo, que pode ser traduzido em uma janela com pixels de valores reduzidos de intensidade. Assim, o valor do pixel mínimo teria mais peso sobre o cálculo de P do que a média. Se, por outro lado, a diferença entre o pixel de valor mínimo e a média for mais significativa, a janela contém mais pixels de alta intensidade, e a média teria mais peso sobre o cálculo do parâmetro. Em um último caso, o parâmetro apenas levaria em conta apenas a média obtida pela janela. As Equações 4.2.3, 4.2.3.1 e 4.2.3.2 abaixo descrevem este comportamento matematicamente.

Caso I: a diferença entre o pixel de valor máximo e a média é maior do que a diferença entre o pixel de valor mínimo e a média: parâmetro T mais próximo do valor mínimo.

$$\begin{aligned} & \text{Se } |f_{SW-\max(x,y)} - \text{ave}(P)| > |f_{SW-\min(x,y)} - \text{ave}(P)|, \\ T &= \alpha \times \left[\frac{2}{3} \times f_{SW-\min(x,y)} + \frac{1}{3} \times \text{ave}(P) \right], \quad (0,3 < \alpha < 0,8) \end{aligned} \quad (4.8)$$

Caso II: a diferença entre o pixel de valor mínimo e a média é maior do que a diferença entre o pixel de valor máximo e a média: parâmetro T mais próximo da média.

$$\begin{aligned} & \text{Se } |f_{SW-\max(x,y)} - \text{ave}(P)| < |f_{SW-\min(x,y)} - \text{ave}(P)|, \\ T &= \alpha \times \left[\frac{1}{3} \times f_{SW-\min(x,y)} + \frac{2}{3} \times \text{ave}(P) \right], \quad (0,3 < \alpha < 0,8) \end{aligned} \quad (4.9)$$

Caso III: nenhuma das situações anteriores ocorre: parâmetro T proporcional à média da janela.

$$T = \alpha \times \text{ave}(P), \quad (0,3 < \alpha < 0,8) \quad (4.10)$$

5. EXPERIMENTOS E RESULTADOS

Em 2009, durante a Conferência Internacional sobre Análise e Reconhecimento de Documentos (ICDAR-09) foi organizado pela primeira vez um concurso sobre binarização de documentos (DIBCO-09) [25]. O objetivo geral desse concurso é identificar os avanços propostos na área de binarização de imagens de documentos, utilizando-se de métricas de avaliação de performance.

Este capítulo apresenta os resultados obtidos a partir do uso das duas técnicas estudadas no capítulo anterior, bem como os resultados obtidos pelas técnicas apresentadas no Capítulo dois, comparando seus resultados utilizando as medidas utilizadas no DIBCO [25]. Todas as figuras apresentam uma imagem original e seu *ground truth* fornecido pela competição, sendo esta a imagem ideal a ser alcançada.

As tabelas que apresentam os resultados obtidos pelos algoritmos apresentam diversas medidas utilizadas para a avaliação dos algoritmos. Neste trabalho, destacamos cinco delas para efetuar as comparações, desprezando as demais [26]:

- ✓ **F-Measure:** quantifica a relação entre os pixels pertencentes ao texto corretamente classificados, os pertencentes ao texto que foram classificados como *background* e os que pertenciam ao *background* e foram classificados como texto (Equação 5.1). Para um resultado ideal, o valor do F-Measure é 100.

$$F_{Measure} = \frac{2 \cdot Recall \cdot Precisão}{Recall + Precisão} \quad (5.1)$$

onde

$$Recall = \frac{TP}{TP + FN}$$

$$Precisão = \frac{TP}{TP + FP}$$

TP, FP, TN e FN e representam as ocorrências de Verdadeiro Positivo (pixel classificado como texto corretamente), Falso Positivo (pixel classificado como texto erroneamente), Verdadeiro Negativo (pixel classificado como fundo corretamente) e Falso Negativo (pixel classificado como fundo erroneamente).

- ✓ **PSNR (*Peak Signal-to-Noise Ratio*)**: indica a similaridade entre duas imagens. Quanto maior o seu valor, mais idênticas serão as imagens. É calculado sobre o valor do erro médio quadrático (MSE – *Mean Square Error*) (Equação 5.2).

$$PSNR = 10 \cdot \log \frac{C^2}{MSE} \quad (5.2)$$

onde

$$MSE = \frac{\sum_{y=1}^H \sum_{x=1}^W [g(x, y) - g'(x, y)]^2}{H \cdot W}$$

O valor C representa o maior valor do domínio de cor da imagem. H e W correspondem à altura e largura da imagem e g e g' são as imagens a serem comparadas.

- ✓ **NRM (*Negative Radio Metric*)**: avalia o índice de erro do algoritmo durante a classificação dos pixels, levando em conta as classificações erradas presentes na imagem. Diferentemente do F-Measure e do PSNR, uma boa binarização apresenta um NRM mais baixo, sendo zero seu valor ideal (Equação 5.3).

$$NRM = \frac{NR_{FN} + NR_{FP}}{2}$$

onde

$$NR_{FN} = \frac{FN}{FN+TP} \quad (5.3)$$

$$NR_{FP} = \frac{FP}{FP+TN}$$

- ✓ **MPM (Misclassification Penalty Metric):** medida de punição aplicada sobre o algoritmo incidente sobre os erros de classificação de pixels. A punição é aplicada levando-se em conta a distância do pixel mal classificado em relação à borda do objeto mais próximo presente no *ground truth*. Um valor MPM baixo denota que o algoritmo é eficiente em encontrar as bordas dos objetos presentes na imagem (Equação 5.4).

$$MPM = \frac{MP_{FN} + MP_{FP}}{2} \quad (5.4)$$

onde

$$MP_{FN} = \frac{\sum_{i=1}^{N_{FN}} d_{FN}^i}{D}$$

$$MP_{FP} = \frac{\sum_{j=1}^{N_{FP}} d_{FP}^j}{D}$$

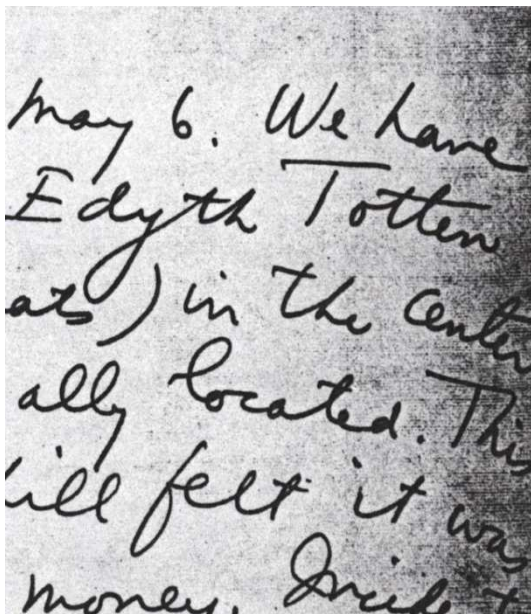
d_{FN}^i e d_{FP}^j representam as distâncias do i-ésimo falso negativo e j-ésimo falso positivo a partir do contorno de um segmento de texto retirado do *ground truth*. D é um fator de normalização. N_{FP} e N_{FN} representam o total de ocorrências de falso positivo e falso negativo.

- ✓ **Total:** esta não é uma medida oficial utilizada pela DIBCO, mas representa a soma de outras quatro medidas (*precision*, *recall*, *accuracy* e *specificity*) não apresentadas na tabela. Como cada uma das medidas tem valor máximo de 1, a soma das quatro tende a 4. Assim, quanto mais próximo de 4 for a soma, melhor o resultado da binarização.

As próximas seções apresentam alguns testes realizados durante a produção deste trabalho. As imagens aqui apresentadas foram escolhidas a partir do conjunto de imagens fornecido pela DIBCO para testes, entre os anos de 2009 e 2011. Para os Testes I, II e III, foram utilizadas imagens com características semelhantes, já que todas são de texto manuscrito, porém com disposição de *background* diferente. Já para o Teste IV, a imagem escolhida foi, dentre todas as imagens de documentos impressos, aquela que apresentou o resultado mais próximo do ideal.

5.1 TESTE I

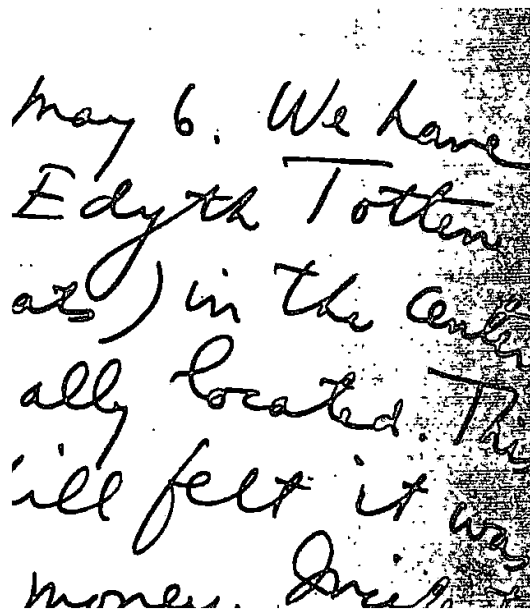
Nesta seção, aplicamos as técnicas estudadas sobre uma imagem de um documento manuscrito com *background* não uniforme, porém próximo de um comportamento degrade, retirada da base de dados utilizada pelo DIBCO. Todos os resultados obtidos são comparados visualmente (Figura 5.1) e quantitativamente (Tabela I), a partir do *ground truth* associado.



(a) Figura original.

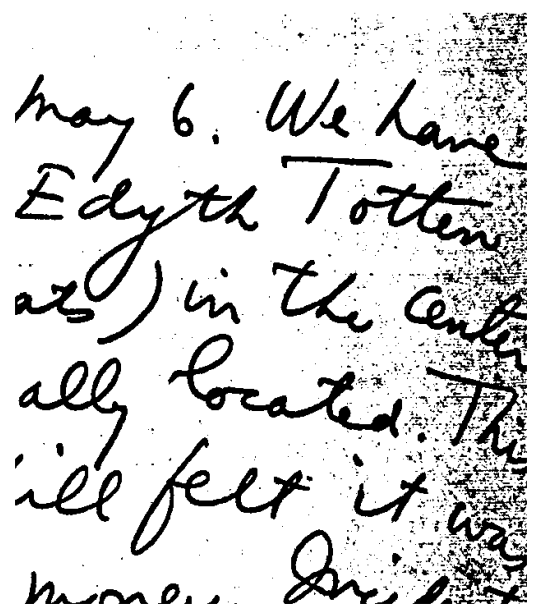
May 6. We have
Edyth Totten
as) in the center
ally located. This
ill felt it was
moner. Drick +

(b) Ground truth.



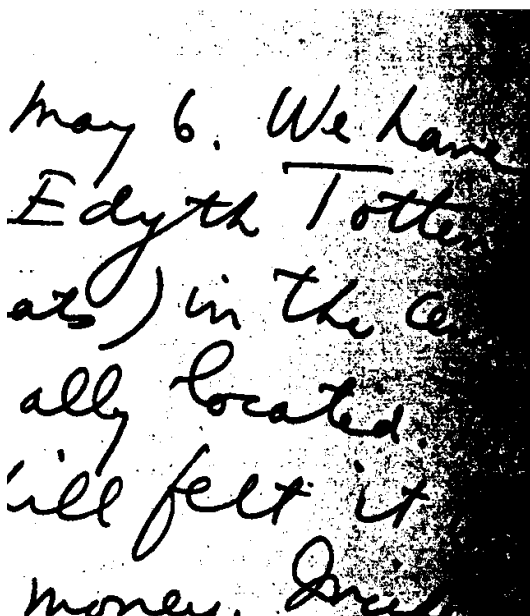
may 6. We have
Edyth Totten
as) in the center
ally located. This
ill felt it was
money. I will

(c) Máximos e mínimos locais.



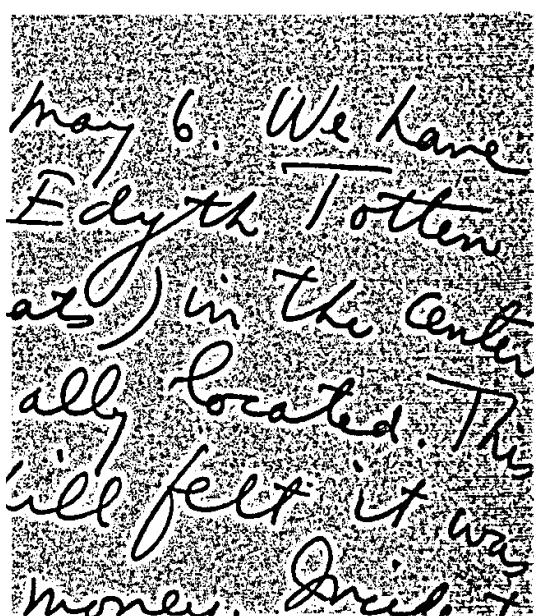
may 6. We have
Edyth Totten
as) in the center
ally located. This
ill felt it was
money. I will

(d) Níveis lógicos.



may 6. We have
Edyth Totten
as) in the center
ally located. This
ill felt it was
money. I will

(e) Complex hull.



may 6. We have
Edyth Totten
as) in the center
ally located. This
ill felt it was
money. I will

(f) Niblack.

May 6. We have
Edyth Totten
as) in the center
ally located. This
ill felt it was
money. Drills

(g) Sauvola.

May 6. We have
Edyth Totten
as) in the center
ally located. This
ill felt it was
money. Drills

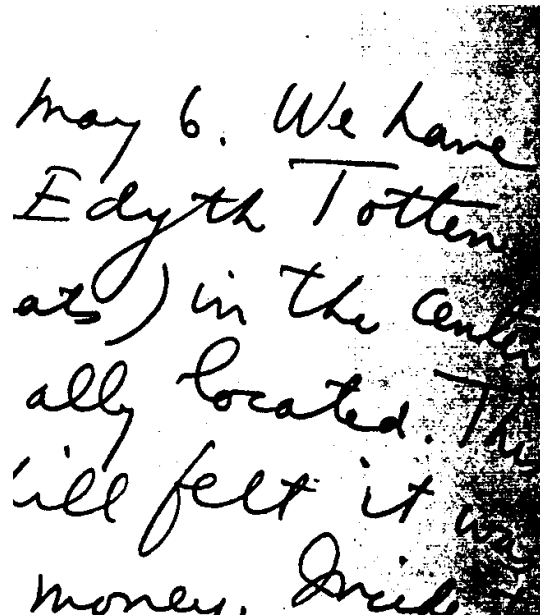
(h) Brink.

May 6. We have
Edyth Totten
as) in the center
ally located. This
ill felt it was
money. Drills

(i) C-means.

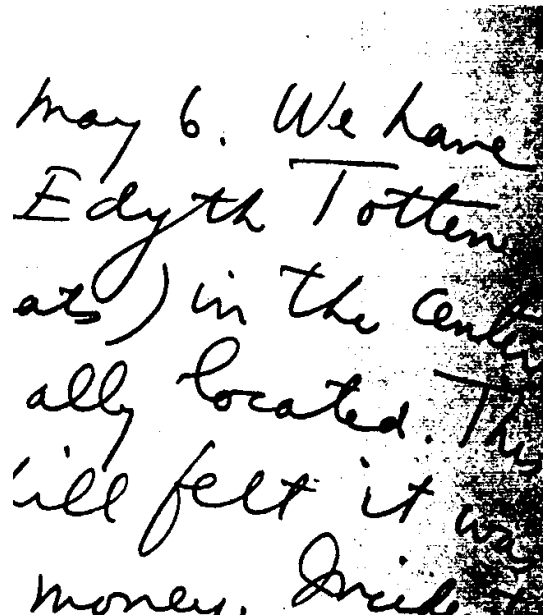
May 6. We have
Edyth Totten
as) in the center
ally located. This
ill felt it was
money. Drills

(j) Fisher.



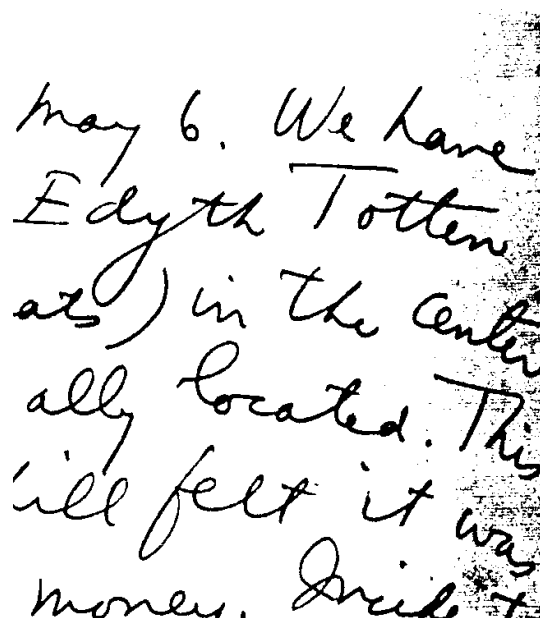
May 6. We have
Edyth Totten
as) in the center
ally located. This
ill felt it was
money. Irid

(l) Nível de cinza médio.



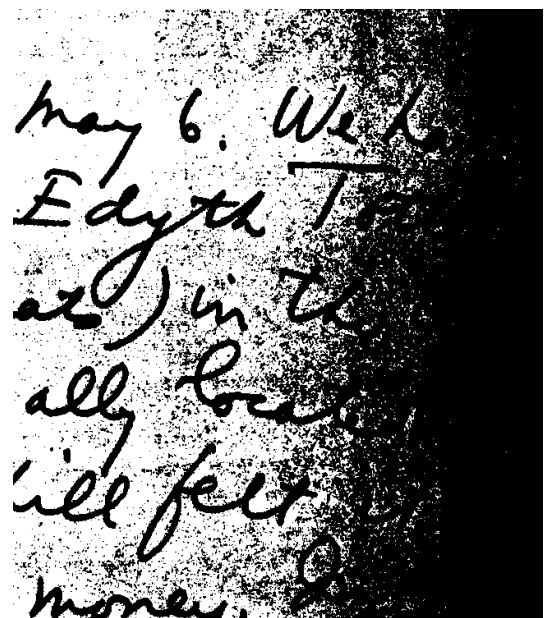
May 6. We have
Edyth Totten
as) in the center
ally located. This
ill felt it was
money. Irid

(m) Otsu.



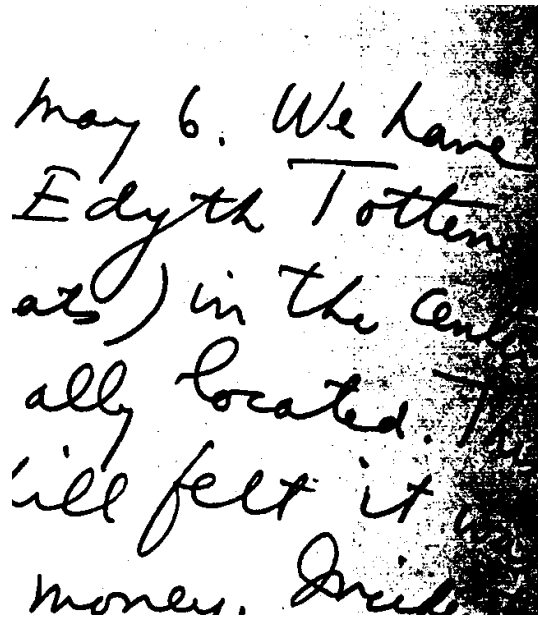
May 6. We have
Edyth Totten
as) in the center
ally located. This
ill felt it was
money. Irid

(n) Porcentagem de preto.



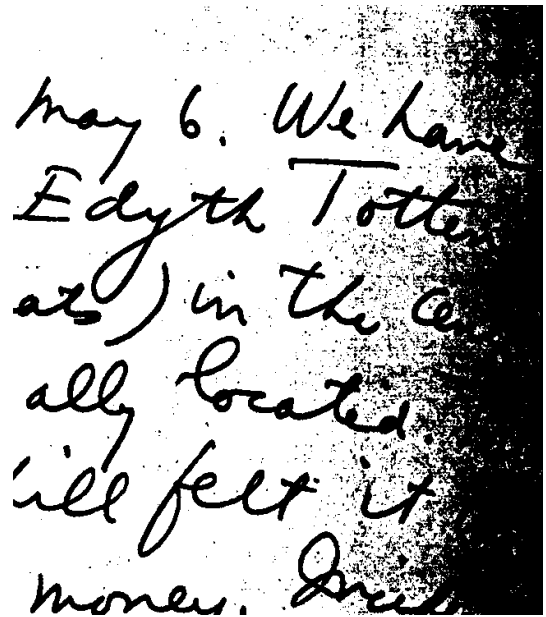
May 6. We have
Edyth Totten
as) in the center
ally located. This
ill felt it was
money. Irid

(o) Pun.



May 6. We have
Edyth Totten
at) in the centrally
located. The
ill felt it in
money. Drick

(p) Yager.



May 6. We have
Edyth Totten
at) in the centrally
located. The
ill felt it in
money. Drick

(q) Ye-Danielsson.

Figura 5.1: Resultados da aplicação dos algoritmos de binarização.

5.1.1 Resultados do Teste I

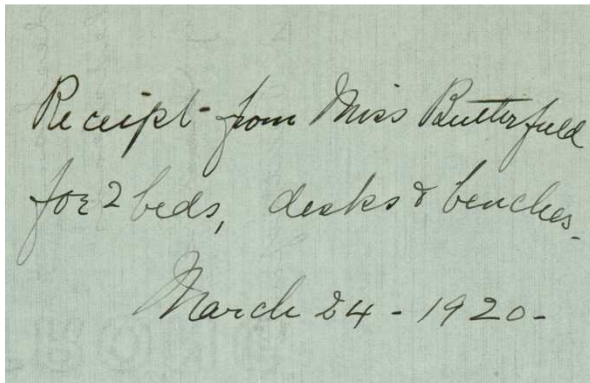
	Total	True-Pos.	False-Pos.	True-Neg.	False-Neg.	F-Measure	NRM	MPM	MSE	PSNR
Máximos e mínimos (Fig. 5.1c)	3,461	51.982	20.643	397.867	8.743	77,963	0,097	33,197	0,061	60,255
Níveis lógicos (Fig. 5.1d)	3,545	56.357	21.653	396.857	4.368	81,244	0,062	27,673	0,054	60,783
Complex hull (Fig. 5.1e)	3,029	60.583	85.523	332.987	142	58,582	0,103	115,717	0,179	55,608
Niblack (Fig. 5.1f)	2,691	53.709	112.195	306.315	7.016	47,398	0,192	162,436	0,249	54,173
Sauvola (Fig. 5.1g)	2,484	56.726	157.875	260.635	3.999	41,206	0,222	230,069	0,338	52,845
Brink (Fig. 5.1h)	3,240	59.090	55.130	363.380	1.635	67,553	0,079	81,187	0,118	57,396
C-means (Fig. 5.1i)	3,185	59.763	63.279	355.231	962	65,042	0,084	90,875	0,134	56,858
Fisher (Fig. 5.1j)	3,240	59.090	55.130	363.380	1.635	67,553	0,079	81,187	0,118	57,396
Johanssen (*)	*	*	*	*	*	*	*	*	*	*
Nível de cinza médio (Fig. 5.1l)	3,347	55.918	36.981	381.529	4.807	72,799	0,084	57,929	0,087	58,726
Otsu (Fig. 5.1m)	3,382	54.102	30.320	388.190	6.623	74,548	0,091	48,313	0,077	59,261
Porcentagem de preto (Fig. 5.1n)	3,512	42.519	5.995	412.515	18.206	77,846	0,157	105,505	0,050	61,098
Pun (Fig. 5.1o)	2,388	60.725	190.757	227.753	0	38,901	0,228	231,701	0,398	52,132
Wulu (*)	*	*	*	*	*	*	*	*	*	*
Yager (Fig. 5.1p)	3,290	58.048	47.298	371.212	2.677	69,907	0,079	71,705	0,104	57,949
Ye-Danielsson (Fig. 5.1q)	3,039	60.564	84.129	334.381	161	58,967	0.10184	114,241	0,176	55,679

Tabela I: Demonstração dos resultados obtidos a partir da comparação das imagens binarizadas produzidas pelos algoritmos com a imagem *ground truth* correspondente à original.

De acordo com a Tabela I, a técnica de níveis lógicos, estudada neste trabalho, apresenta resultados superiores em relação a todas as outras técnicas em quatro das cinco medidas selecionadas. Os algoritmos de Johanssen e Wulu não foram capazes de identificar a presença de texto na imagem utilizada, por tanto os valores dos seus resultados e suas imagens não foram considerados para efeito desta comparação.

5.2 TESTE II

Nesta seção, aplicamos todos os algoritmos à outra imagem de documento manuscrito, também retirada da base de imagens do DIBCO, com *background* mais uniforme e sem grande presença de ruídos no corpo do texto. Os resultados visuais estão indicados na Figura 5.2 e a análise quantitativa pode ser feita a partir dos dados da Tabela II.



(a) Imagem original.

Receipt - from Miss Butterfield
for 2 beds, desks & benches.
March 24 - 1920.

(b) Ground truth.

Receipt - from Miss Butterfield
for 2 beds, desks & benches.
March 24 - 1920.

(c) Máximos e mínimos locais.

Receipt - from Miss Butterfield
for 2 beds, desks & benches.
March 24 - 1920.

(d) Níveis lógicos.

Receipt from Miss Butterfield
for 2 beds, desks & benches.
March 24 - 1920.

(e) Complex hull.

Receipt from Miss Butterfield
for 2 beds, desks & benches.
March 24 - 1920.

(f) Niblack.

Receipt from Miss Butterfield
for 2 beds, desks & benches.
March 24 - 1920.

(g) Sauvola.

Receipt from Miss Butterfield
for 2 beds, desks & benches.
March 24 - 1920.

(h) Brink.

Receipt from Miss Butterfield
for 2 beds, desks & benches.
March 24 - 1920.

(i) C-means.

Receipt from Miss Butterfield
for 2 beds, desks & benches.
March 24 - 1920.

(j) Fisher.

Receipt from Miss Butterfield
for 2 beds, desks & benches.
March 24 - 1920.

(l) Nível de cinza médio.

Receipt from Miss Butterfield
for 2 beds, desks & benches.
March 24 - 1920.

(m) Otsu.

Receipt from Miss Butterfield
for 2 beds, desks & benches.
March 24 - 1920.

(n) Porcentagem de preto.

Receipt from Miss Butterfield
for 2 beds, desks & benches.
March 24 - 1920.

(o) Pun.

Receipt from Miss Butterfield
for 2 beds, desks & benches.
March 24 - 1920.

(p) Yager.

Receipt from Miss Butterfield
for 2 beds, desks & benches.
March 24 - 1920.

(q) Ye-Danielsson.

Figura 5.2: Resultados da aplicação dos algoritmos de binarização.

5.2.1 Resultados do Teste II

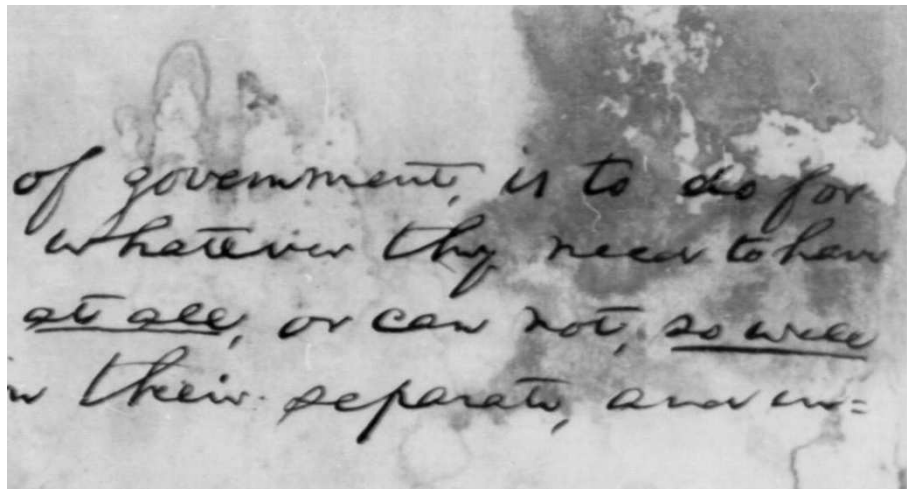
	Total	True-Pos.	False-Pos.	True-Neg.	False-Neg.	F-Measure	NRM	MPM	MSE	PSNR
Máximos e mínimos (Fig. 5.2c)	3,680	30.916	430	907.134	12.778	82,399	0,146	0,076	0,014	66,705
Níveis lógicos (Fig. 5.2d)	3,780	35.718	983	906.581	7.976	88,856	0,092	0,039	0,009	68,391
Complex hull (Fig. 5.2e)	3,700	42.877	14.251	893.313	817	85,055	0,017	4,975	0,016	66,133
Niblack (Fig. 5.2f)	2,457	40.196	273.042	634.522	3.498	22,523	0,190	164,315	0,291	53,496
Sauvola (Fig. 5.2g)	2,174	42.026	408.381	499.183	1.668	17,011	0,244	250,534	0,431	51,785
Brink (Fig. 5.2h)	3,786	35.487	592	906.972	8.207	88,970	0,094	0,034	0,009	68,470
C-means (Fig. 5.2i)	3,319	43.402	54.539	853.025	292	61,287	0,033	28,183	0,058	60,524
Fisher (Fig. 5.2j)	3,786	35.487	592	906.972	8.207	88,970	0,094	0,034	0,009	68,470
Johanssen (*)	*	*	*	*	*	*	*	*	*	*
Nível de cinza médio (Fig. 5.2l)	3,526	23.895	8	907.556	19.799	70,698	0,227	0,086	0,021	64,946
Otsu (Fig. 5.2m)	3,829	38.296	1.520	906.044	5.398	91,716	0,063	0,049	0,007	69,514
Porcentagem de preto (Fig. 5.2n)	3,319	43.402	54.539	853.025	292	61,287	0,033	28,183	0,058	60,524
Pun (Fig. 5.2o)	2,197	43.684	416.757	490.807	10	17,330	0,230	237,958	0,438	51,715
Wulu (*)	*	*	*	*	*	*	*	*	*	*
Yager (Fig. 5.2p)	3,726	32.509	184	907.380	11.185	85,117	0,128	0,039	0,012	67,357
Ye-Danielsson (Fig. 5.2q)	3,837	41.720	4.913	902.651	1.974	92,376	0,025	0,622	0,007	69,534

Tabela II: Demonstração dos resultados obtidos a partir da comparação das imagens binarizadas produzidas pelos algoritmos com a imagem *ground truth* correspondente à original.

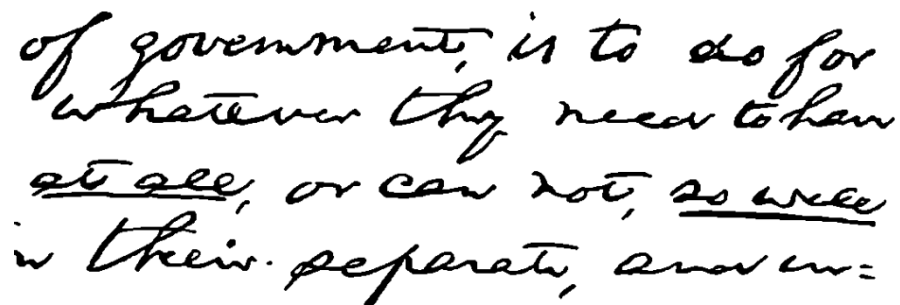
De acordo com a Tabela II, apenas o algoritmo de Ye-Danielsson obteve o melhor índice em mais de uma categoria. A técnica de níveis lógicos foi a segunda melhor no índice MPM. Novamente, assim como no Teste I, os algoritmos de Johanssen e Wulu não foram capazes de identificar a presença de texto nessa imagem e produziram imagens brancas. Seus resultados visuais e quantitativos foram novamente excluídos.

5.3 TESTE III

Para o Teste III, utilizamos uma imagem de um documento manuscrito com *background* sem um comportamento previsível, como na imagem do Teste I e sem a característica uniforme da imagem utilizada no Teste II. Os resultados visuais são apresentados na Figura 5.3 e comparados quantitativamente na Tabela III.



(a) Imagem original



(b) Ground truth

of government, is to do for
whatever they need to have
at all, or can not, as well
in their separate, answer

(c) Máximos e mínimos locais

of government, is to do for
whatever they need to have
at all, or can not, as well
in their separate, answer

(d) Níveis lógicos

of government,
whatever they
at all, or can
in their separate, answer

(e) Complex hull

of government, is to do for
whatever they need to have
at all, or can not, as well
in their separate, and in

(f) Niblack

of government, is to do for
whatever they need to have
at all, or can not, as well
in their separate, and in

(g) Sauvola

of government,
whatever they
at all, or can not, as well
in their separate, and in

(h) Brink

of government,
whatever they
at all, or can at all
in their separate, and in

(i) C-means

of government,
whatever they
at all, or can at all
in their separate, and in

(j) Fisher

of government,
whatever they
at all, or can at all
in their separate, and in

(l) Johanssen

of government, if it does
whatever they think to be
at all, or can not, so will
in their separate, and in:

(m) Nível de cinza médio

of government, if it does
whatever they think to be
at all, or can not, so will
in their separate, and in:

(n) Otsu

of government, if it does
whatever they think to be
at all, or can not, so will
in their separate, and in:

(o) Porcentagem de preto

of government, is to determine
whatever they need to have
at all, or can do, so well
in their separate, and in

(p) Pun

of government, is to determine
whatever they need to have
at all, or can do, so well
in their separate, and in

(q) Wulu

of government, is to determine
whatever they need to have
at all, or can do, so well
in their separate, and in

(r) Yager

of government, is to defer
whatever they need to have
at all, or can not, so well
in their separate, unco-ordinated

(s) Ye-Danielsson

Figura 5.3: Resultados da aplicação dos algoritmos de binarização.

5.3.1 Resultados do Teste III

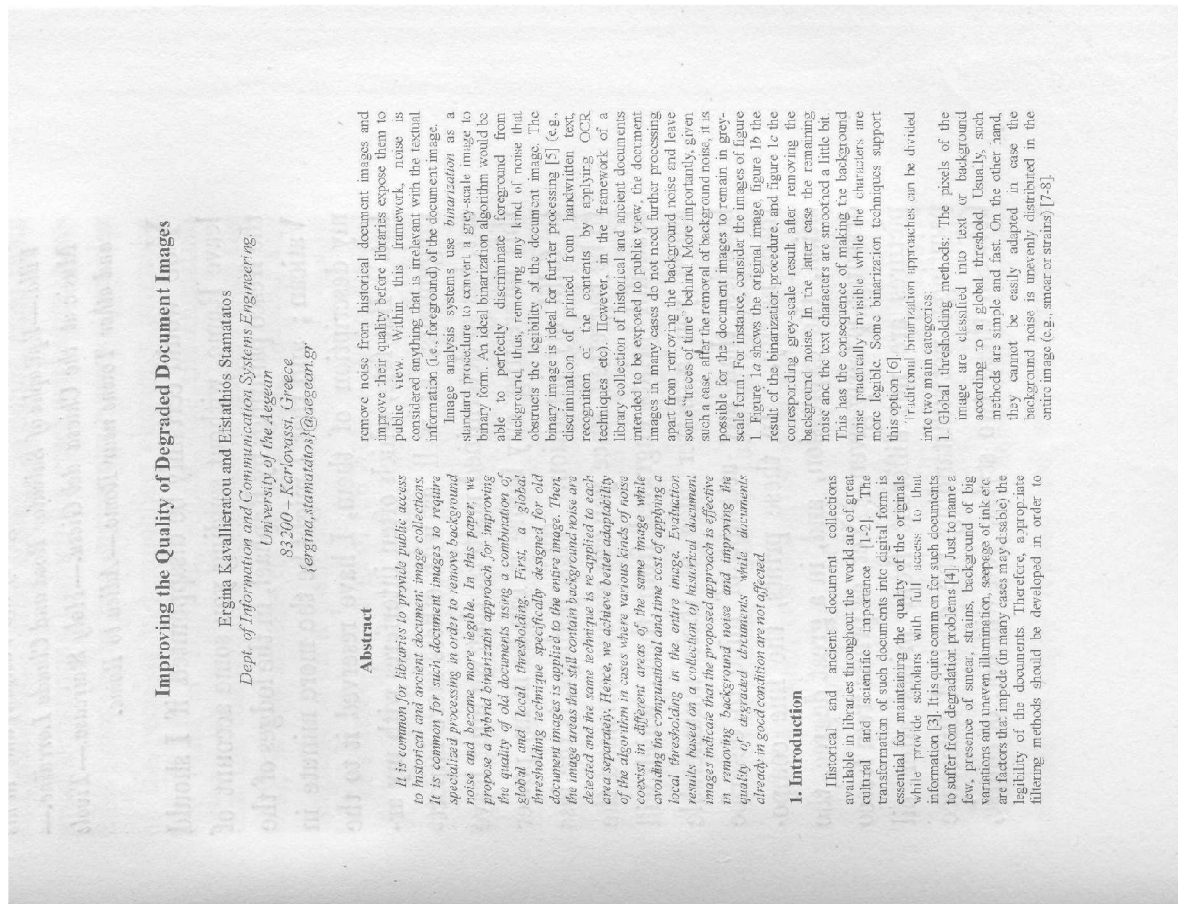
	Total	True-Pos.	False-Pos.	True-Neg.	False-Neg.	F-Measure	NRM	MPM	MSE	PSNR
Máximos e mínimos (Fig. 5.3c)	3,676	33.721	12.777	587.117	256	83,805	0,014	0,572	0,021	65,000
Níveis lógicos (Fig. 5.3d)	3,613	44.499	1.999	567.696	19.677	80,415	0,155	0,674	0,034	62,791
Complex hull (Fig. 5.3e)	2,880	46.443	55	391.150	196.223	32,122	0,404	152,521	0,310	53,222
Niblack (Fig. 5.3f)	2,881	42.253	4.245	430.422	147.951	35,701	0,394	42,226	0,240	54,327
Sauvola (Fig. 5.3g)	2,673	44.094	2.404	327.737	259.636	25,180	0,431	169,330	0,413	51,967
Brink (Fig. 5.3h)	3,021	45.956	542	450.319	137.054	40,047	0,375	61,837	0,217	54,765
C-means (Fig. 5.3i)	2,969	46.233	265	429.007	158.366	36,825	0,387	87,636	0,250	54,147
Fisher (Fig. 5.3j)	3,029	45.900	598	453.423	133.950	40,557	0,373	59,817	0,212	54,862
Johanssen (Fig. 5.3l)	3,136	43.965	2.533	500.608	86.765	49,614	0,334	18,797	0,141	56,642
Nível de cinza médio (Fig. 5.3m)	3,219	41.040	5.458	534.807	52.566	58,585	0,286	6,885	0,092	58,515
Otsu (Fig. 5.3n)	3,252	40.306	6.192	543.048	44.325	61,475	0,268	5,418	0,080	59,116
Porcentagem de preto (Fig. 5.3o)	3,342	38.088	8.410	561.163	26.210	68,753	0,211	2,385	0,055	60,758
Pun (Fig. 5.3p)	2,663	46.497	1	288.365	299.008	23,723	0,433	685,942	0,472	51,394
Wulu (Fig. 5.3q)	3,192	41.763	4.735	526.427	60.946	55,980	0,301	9,013	0,104	57,976
Yager (Fig. 5.3r)	3,390	36.773	9.725	568.521	18.852	72,017	0,178	1,787	0,045	61,591
Ye-Danielsson (Fig. 5.3s)	3,495	32.150	14.348	582.075	5.298	76,597	0,083	0,917	0,031	63,318

Tabela III: Demonstração dos resultados obtidos a partir da comparação das imagens binarizadas produzidas pelos algoritmos com a imagem *ground truth* correspondente à original.

A partir da Tabela III, é possível perceber que a técnica de Máximos e Mínimos Locais foi superior aos outros algoritmos em todas as métricas utilizadas. Já as técnicas de Sauvola e Pun apresentaram resultados muito ruins.

5.4 TESTE IV

Para o terceiro experimento, utilizamos uma imagem de um documento impresso, retirada da base do DIBCO com nível de ruído médio e fundo com presença de falso-texto. Todos os algoritmos foram capazes de identificar o texto presente na imagem, alguns até com significativa precisão, como indicado na Figura 5.4. Os resultados da comparação quantitativa dos resultados estão indicados na Tabela IV.



(a) Imagem original.

Improving the Quality of Degraded Document Images

Ergina Kavalieratou and Efsthios Stamatiou
Dept. of Information and Communication Systems Engineering,
University of the Aegean
83200 – Karlovassi, Greece
{ergina.stamatiou}@aegean.gr

Abstract

It is common for libraries to provide public access to historical and ancient document image collections specialized processing in order to remove background noise and become more legible. In this paper, we propose a hybrid binarization approach for improving the quality of old documents using a combination of global and local thresholding. First, a global thresholding technique specifically designed for old document images is applied to the entire image. Then the image areas that still contain background noise are detected and the same technique is re-applied to each area separately. Hence, we achieve better adaptability of the algorithm in cases where various kinds of noise coexist in different areas of the same image while avoiding the computational and time cost of applying a local thresholding in the entire image. Evaluation results based on a collection of historical document images indicate that the proposed approach is effective in removing background noise and improving the quality of degraded documents while documents already in good condition are not affected.

1. Introduction

Historical and ancient document collections available in libraries throughout the world are of great cultural and scientific importance [1-2]. The transformation of such documents into digital form is essential for maintaining the quality of the originals while provide scholars with full access to that information [3]. It is quite common for such documents to suffer from degradation problems [4]. Just to name a few, presence of smears, stains, background of big variations and uneven illumination, seepage of ink etc are factors that impede in many cases may disable the legibility of the documents. Therefore, appropriate filtering methods should be developed in order to

remove noise from historical document images and improve their quality before libraries expose them to public view. Within this framework, noise is considered anything that is irrelevant with the textual information (i.e. foreground) of the document image.

Image analysis systems use *binarization* as a standard procedure to convert a grey-scale image to binary form. An ideal binarization algorithm would be able to perfectly discriminate foreground from background, thus, removing any kind of noise that obstructs the legibility of the document image. The binary image is ideal for further processing [5] (e.g., discrimination of printed from handwritten text, recognition of the contents by applying OCR techniques etc). However, in the framework of a library collection of historical and ancient documents intended to be exposed to public view, the document images in many cases do not need further processing apart from removing the background noise and leave some "traces of time" behind. More importantly, given such a case, after the removal of background noise, it is possible for the document images to remain in grey-scale form. For instance, consider the images of figure 1. Figure 1a shows the original image, figure 1b the result of the binarization procedure, and figure 1c the corresponding grey-scale result after removing the background noise. In the latter case the remaining noise and the text characters are smoothed a little bit. This has the consequence of making the background noise practically invisible, while the characters are more legible. Some binarization techniques support this option [6].

Traditional binarization approaches can be divided into two main categories: 1. Global thresholding methods: The pixels of the image are classified into text or background according to a global threshold. Usually, such methods are simple and fast. On the other hand, they cannot be easily adapted in case the background noise is unevenly distributed in the entire image (e.g., smears or stains) [7-8].

(c) Máximos e mínimos locais

Improving the Quality of Degraded Document Images

Ergina Kavalieratou and Efsthios Stamatiou
Dept. of Information and Communication Systems Engineering,
University of the Aegean
83200 – Karlovassi, Greece
{ergina.stamatiou}@aegean.gr

Abstract

It is common for libraries to provide public access to historical and ancient document image collections specialized processing in order to remove background noise and become more legible. In this paper, we propose a hybrid binarization approach for improving the quality of old documents using a combination of global and local thresholding. First, a global thresholding technique specifically designed for old document images is applied to the entire image. Then the image areas that still contain background noise are detected and the same technique is re-applied to each area separately. Hence, we achieve better adaptability of the algorithm in cases where various kinds of noise coexist in different areas of the same image while avoiding the computational and time cost of applying a local thresholding in the entire image. Evaluation results based on a collection of historical document images indicate that the proposed approach is effective in removing background noise and improving the quality of degraded documents while documents already in good condition are not affected.

1. Introduction

Historical and ancient document collections available in libraries throughout the world are of great cultural and scientific importance [1-2]. The transformation of such documents into digital form is essential for maintaining the quality of the originals while provide scholars with full access to that information [3]. It is quite common for such documents to suffer from degradation problems [4]. Just to name a few, presence of smears, stains, background of big variations and uneven illumination, seepage of ink etc are factors that impede in many cases may disable the legibility of the documents. Therefore, appropriate filtering methods should be developed in order to

(b) Ground truth.

Improving the Quality of Degraded Document Images

Ergina Kavallieratou and Efsthios Stamatiadis

Dept. of Information and Communication Systems Engineering,

University of the Aegean

83200 - Karlovassi, Greece

{ergina.stamatiadis}@aegean.gr

Abstract

It is common for libraries to provide public access to historical and ancient document image collections. It is common for such document images to require specialized processing in order to remove background noise and become more legible. In this paper, we propose a hybrid binarization approach for improving the quality of old documents using a combination of global and local thresholding. First, a global thresholding technique specifically designed for old document images is applied to the entire image. Then, the image areas that still contain background noise are detected and the same technique is re-applied to each area separately. Hence, we achieve better adaptability of the algorithm in cases where various kinds of noise coexist in different areas of the same image while avoiding the computational and time cost of applying a local thresholding in the entire image. Evaluation results based on a collection of historical document images indicate that the proposed approach is effective in removing background noise and improving the quality of degraded documents while documents already in good condition are not affected.

1. Introduction

Historical and ancient document collections available in libraries throughout the world are of great cultural and scientific importance [1-2]. The transformation of such documents into digital form is essential for maintaining the quality of the originals while provide scholars with full access to that information [3]. It is quite common for such documents to suffer from degradation problems [4]. Just to name a few, presence of smears, stains, background of ink etc. are factors that impede in many cases may disable the legibility of the documents. Therefore, appropriate filtering methods should be developed in order to

remove noise from historical document images and improve their quality before libraries expose them to public view. Within this framework, noise is considered anything that is irrelevant with the textual information (i.e. foreground) of the document image.

Image analysis systems use binarization as a standard procedure to convert a grey-scale image to binary form. An ideal binarization algorithm would be able to perfectly discriminate foreground from background, thus, removing any kind of noise that obstructs the legibility of the document image. The binary image is ideal for further processing [5] (e.g., discrimination of printed from handwritten text, recognition of the contents by applying OCR techniques etc.). However, in the framework of a library collection of historical and ancient documents intended to be exposed to public view, the document images in many cases do not need further processing apart from removing the background noise and leave some "traces of time" behind. More importantly, given such a case, after the removal of background noise, it is possible for the document images to remain in grey-scale form. For instance, consider the images of figure 1. Figure 1a shows the original image, figure 1b the result of the binarization procedure, and figure 1c the corresponding grey-scale result after removing the background noise. In the latter case the remaining noise and the text characters are smoothed a little bit. This has the consequence of making the background noise practically invisible while the characters are more legible. Some binarization techniques support this option [6].

Traditional binarization approaches can be divided into two main categories:

1. Global thresholding methods: The pixels of the image are classified into text or background according to a global threshold. Usually, such methods are simple and fast. On the other hand, they cannot be easily adapted in case the background noise is unevenly distributed in the entire image (e.g., smears or stains) [7-8].

Abstract

It is common for libraries to provide public access to historical and ancient document image collections. It is common for such document images to require specialized processing in order to remove background noise and become more legible. In this paper, we propose a hybrid binarization approach for improving the quality of old documents using a combination of global and local thresholding. First, a global thresholding technique specifically designed for old document images is applied to the entire image. Then, the image areas that still contain background noise are detected and the same technique is re-applied to each area separately. Hence, we achieve better adaptability of the algorithm in cases where various kinds of noise coexist in different areas of the same image while avoiding the computational and time cost of applying a local thresholding in the entire image. Evaluation results based on a collection of historical document images indicate that the proposed approach is effective in removing background noise and improving the quality of degraded documents while documents already in good condition are not affected.

1. Introduction

Historical and ancient document collections available in libraries throughout the world are of great cultural and scientific importance [1-2]. The transformation of such documents into digital form is essential for maintaining the quality of the originals while provide scholars with full access to that information [3]. It is quite common for such documents to suffer from degradation problems [4]. Just to name a few, presence of smears, stains, background of ink etc. are factors that impede in many cases may disable the legibility of the documents. Therefore, appropriate filtering methods should be developed in order to

Improving the Quality of Degraded Document Images

Ergina Kavallieratou and Efsthios Stamatiadis

Dept. of Information and Communication Systems Engineering,

University of the Aegean

83200 - Karlovassi, Greece

{ergina.stamatiadis}@aegean.gr

remove noise from historical document images and improve their quality before libraries expose them to public view. Within this framework, noise is considered anything that is irrelevant with the textual information (i.e. foreground) of the document image.

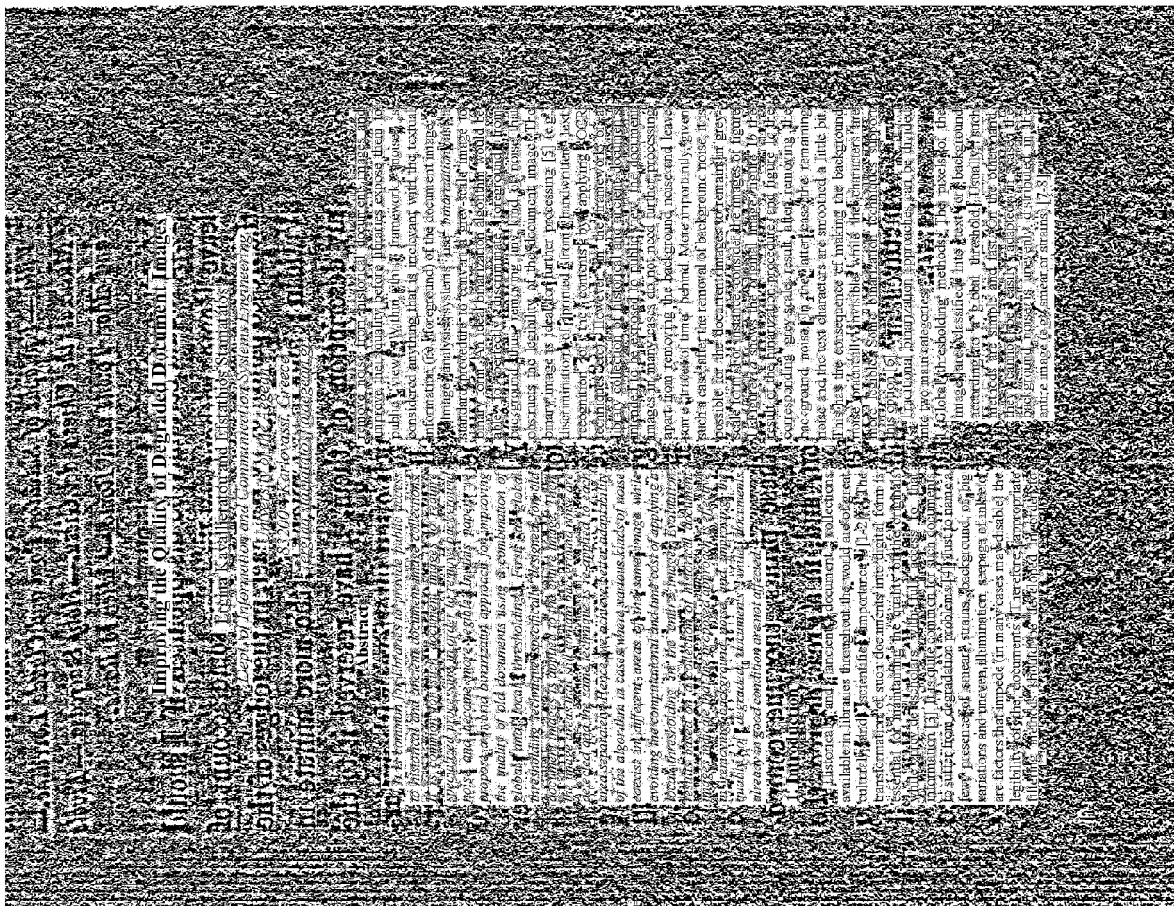
Image analysis systems use binarization as a standard procedure to convert a grey-scale image to binary form. An ideal binarization algorithm would be able to perfectly discriminate foreground from background, thus, removing any kind of noise that obstructs the legibility of the document image. The binary image is ideal for further processing [5] (e.g., discrimination of printed from handwritten text, recognition of the contents by applying OCR techniques etc.). However, in the framework of a library collection of historical and ancient documents intended to be exposed to public view, the document images in many cases do not need further processing apart from removing the background noise and leave some "traces of time" behind. More importantly, given such a case, after the removal of background noise, it is possible for the document images to remain in grey-scale form. For instance, consider the images of figure 1. Figure 1a shows the original image, figure 1b the result of the binarization procedure, and figure 1c the corresponding grey-scale result after removing the background noise. In the latter case the remaining noise and the text characters are smoothed a little bit. This has the consequence of making the background noise practically invisible while the characters are more legible. Some binarization techniques support this option [6].

Traditional binarization approaches can be divided into two main categories:

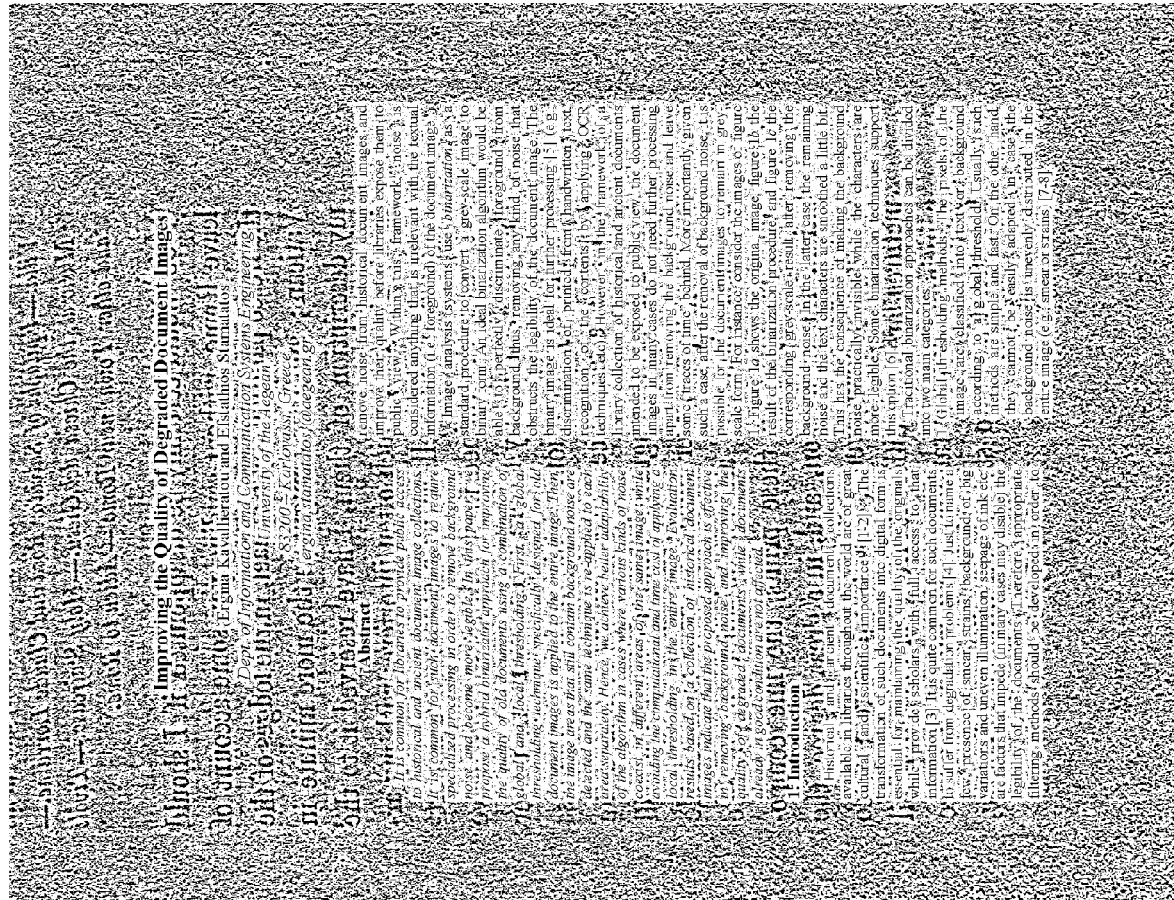
1. Global thresholding methods: The pixels of the image are classified into text or background according to a global threshold. Usually, such methods are simple and fast. On the other hand, they cannot be easily adapted in case the background noise is unevenly distributed in the entire image (e.g., smears or stains) [7-8].

(e) Complex hull.

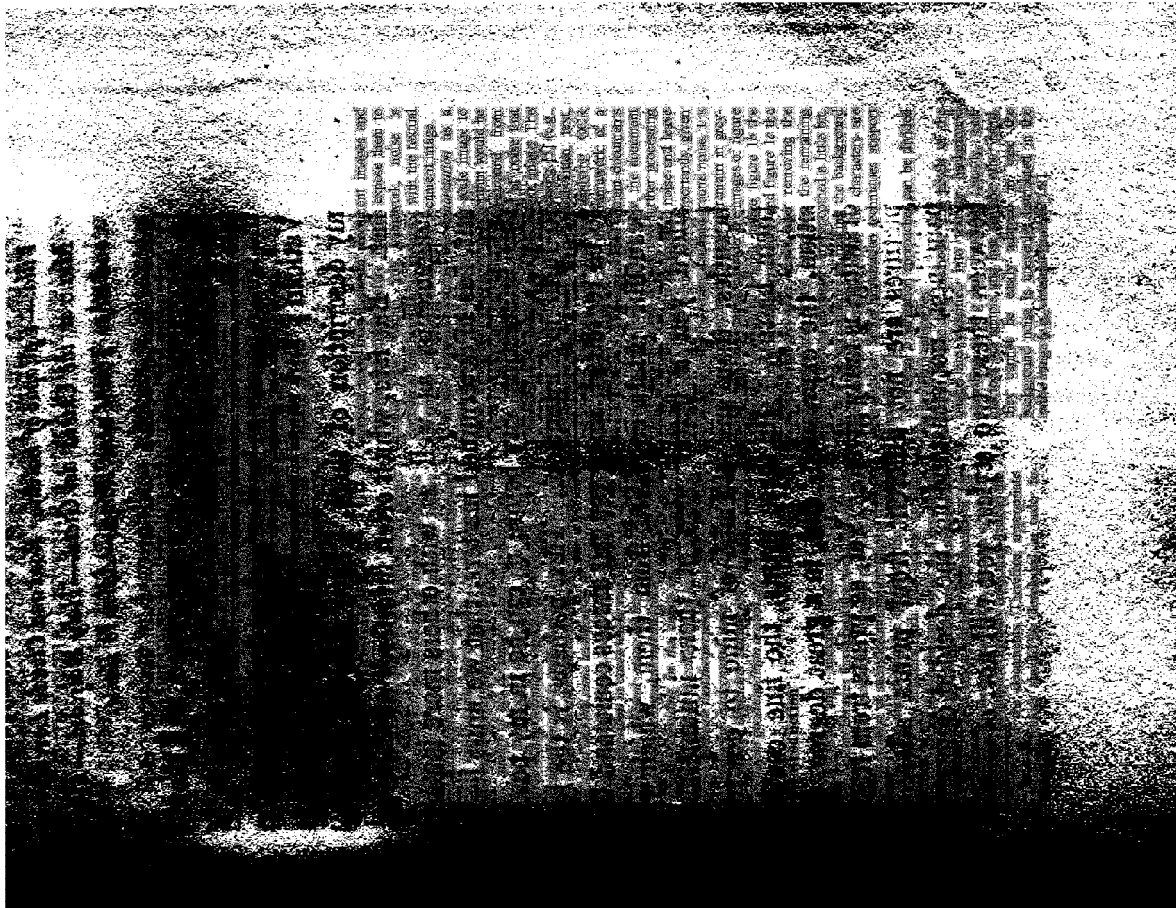
(d) Níveis lógicos.



(g) Sauvola.



(f) Niblack.



(i) C-means.

Improving the Quality of Degraded Document Images

Ergina Kavallieratou and Efsthios Stamatiatos
*Dept. of Information and Communication Systems Engineering,
 University of the Aegean
 83200 – Karlovassi, Greece
 {ergina.stamatiatos}@aegean.gr*

Abstract

It is common for libraries to provide public access to historical and ancient document image collections. It is common for such document images to require specialized processing in order to remove background noise and become more legible. In this paper, we propose a hybrid binarization approach for improving the quality of old documents using a combination of global and local thresholding. First, a global thresholding technique specifically designed for old document images is applied to the entire image. Then, the image areas that still contain background noise are detected and the same technique is re-applied to each area separately. Hence, we achieve better adaptability of the algorithm in cases where various kinds of noise coexist in different areas of the same image while avoiding the computational and time cost of applying a local thresholding in the entire image. Evaluation results based on a collection of historical document images indicate that the proposed approach is effective in removing background noise and improving the quality of degraded documents while documents already in good condition are not affected.

1. Introduction

Historical and ancient document collections available in libraries throughout the world are of great cultural and scientific importance [1-2]. The transformation of such documents into digital form is essential for maintaining the quality of the originals while provide scholars with full access to that information [3]. It is quite common for such documents to suffer from degradation problems [4]. Just to name a few, presence of smear stains, background of big variations and uneven illumination, seepage of ink etc. are factors that impede in many cases may disable the legibility of the documents. Therefore, appropriate filtering methods should be developed in order to

remove noise from historical document images and improve their quality before libraries expose them to public view. Within this framework, noise is considered anything that is irrelevant with the textual information (i.e. foreground) of the document image.

Image analysis systems use binarization as a standard procedure to convert a grey-scale image to binary form. An ideal binarization algorithm would be able to perfectly discriminate foreground from background, thus removing any kind of noise that obstructs the legibility of the document image. The binary image is then used for further processing [5] (e.g., discrimination of printed from handwritten text, recognition of the contents by applying OCR techniques etc.). However, in the framework of a library collection of historical and ancient documents intended to be exposed to public view, the document images in many cases do not need further processing apart from removing the background noise and leave some "traces of time" behind. More importantly, given such a case, after the removal of background noise, it is possible for the document images to remain in grey-scale form. For instance, consider the images of figure 1. Figure 1a shows the original image, figure 1b the result of the binarization procedure, and figure 1c the corresponding grey-scale result after removing the background noise. In the latter case the remaining noise and the text characters are smoothed a little bit. This has the consequence of making the background noise practically invisible while the characters are more legible. Some binarization techniques support this option [6].

Traditional binarization approaches can be divided into two main categories:

1. Global thresholding methods: The pixels of the image are classified into text or background according to a global threshold. Usually, such methods are simple and fast. On the other hand, they cannot be easily adapted in case the background noise is unevenly distributed in the entire image (e.g., smear or stains) [7-8].

(h) Brink.

Improving the Quality of Degraded Document Images

Erigna Kavalieratou and Efsthios Stamatiatos
Dept. of Information and Communication Systems Engineering,
University of the Aegean
83200 – Karlovassi, Greece
{erigna.stamatiatos}@aegean.gr

Abstract

It is common for libraries to provide public access to historical and ancient document image collections. It is common for such document images to require specialized processing in order to remove background noise and become more legible. In this paper, we propose a hybrid binarization approach for improving the quality of old documents using a combination of global and local thresholding. First, a global thresholding technique specifically designed for old document images is applied to the entire image. Then, the image areas that still contain background noise are detected and the same technique is re-applied to each area separately. Hence, we achieve better adaptability of the algorithm in cases where various kinds of noise coexist in different areas of the same image while avoiding the computational and time cost of applying a local thresholding in the entire image. Evaluation results based on a collection of historical document images indicate that the proposed approach is effective in removing background noise and improving the quality of degraded documents while documents already in good condition are not affected.

1. Introduction

Historical and ancient document collections available in libraries throughout the world are of great cultural and scientific importance [1-2]. The transformation of such documents into digital form is essential for maintaining the quality of the originals while provide scholars with full access to that information [3]. It is quite common for such documents to suffer from degradation problems [4]. Just to name a few, presence of smears, stains, background of big variations and uneven illumination, seepage of ink etc. are factors that impede (in many cases may disable) the legibility of the documents. Therefore, appropriate filtering methods should be developed in order to

remove noise from historical document images and improve their quality before libraries expose them to public view. Within this framework, noise is considered anything that is irrelevant with the textual information (i.e., foreground) of the document image.

Image analysis systems use *binarization* as a standard procedure to convert a grey-scale image to binary form. An ideal binarization algorithm would be able to perfectly discriminate foreground from background, thus, removing any kind of noise that obstructs the legibility of the document image. The binary image is ideal for further processing [5] (e.g., discrimination of printed from handwritten text, recognition of the contents by applying OCR techniques etc). However, in the framework of a library collection of historical and ancient documents intended to be exposed to public view, the document images in many cases do not need further processing apart from removing the background noise and leave some traces of noise behind. More importantly, given such a case, after the removal of background noise, it is possible for the document images to remain in grey-scale form. For instance, consider the images of figure 1. Figure 1a shows the original image, figure 1b the result of the binarization procedure, and figure 1c the corresponding grey-scale result after removing the background noise. In the latter case the remaining noise and the text characters are smoothed a little bit. This has the consequence of making the background noise practically invisible while the characters are more legible. Some binarization techniques support this option [6].

Traditional binarization approaches can be divided into two main categories:

1. Global thresholding methods: The pixels of the image are classified into text or background according to a global threshold. Usually, such methods are simple and fast. On the other hand, they cannot be easily adapted in case the background noise is unevenly distributed in the entire image (e.g., smears or stains) [7-8].

(1) Johanssen.

Improving the Quality of Degraded Document Images

Erigna Kavalieratou and Efsthios Stamatiatos
Dept. of Information and Communication Systems Engineering,
University of the Aegean
83200 – Karlovassi, Greece
{erigna.stamatiatos}@aegean.gr

Abstract

It is common for libraries to provide public access to historical and ancient document image collections. It is common for such document images to require specialized processing in order to remove background noise and become more legible. In this paper, we propose a hybrid binarization approach for improving the quality of old documents using a combination of global and local thresholding. First, a global thresholding technique specifically designed for old document images is applied to the entire image. Then, the image areas that still contain background noise are detected and the same technique is re-applied to each area separately. Hence, we achieve better adaptability of the algorithm in cases where various kinds of noise coexist in different areas of the same image while avoiding the computational and time cost of applying a local thresholding in the entire image. Evaluation results based on a collection of historical document images indicate that the proposed approach is effective in removing background noise and improving the quality of degraded documents while documents already in good condition are not affected.

1. Introduction

Historical and ancient document collections available in libraries throughout the world are of great cultural and scientific importance [1-2]. The transformation of such documents into digital form is essential for maintaining the quality of the originals while provide scholars with full access to that information [3]. It is quite common for such documents to suffer from degradation problems [4]. Just to name a few, presence of smears, stains, background of big variations and uneven illumination, seepage of ink etc. are factors that impede (in many cases may disable) the legibility of the documents. Therefore, appropriate filtering methods should be developed in order to

remove noise from historical document images and improve their quality before libraries expose them to public view. Within this framework, noise is considered anything that is irrelevant with the textual information (i.e., foreground) of the document image.

Image analysis systems use *binarization* as a standard procedure to convert a grey-scale image to binary form. An ideal binarization algorithm would be able to perfectly discriminate foreground from background, thus, removing any kind of noise that obstructs the legibility of the document image. The binary image is ideal for further processing [5] (e.g., discrimination of printed from handwritten text, recognition of the contents by applying OCR techniques etc). However, in the framework of a library collection of historical and ancient documents intended to be exposed to public view, the document images in many cases do not need further processing apart from removing the background noise and leave some traces of noise behind. More importantly, given such a case, after the removal of background noise, it is possible for the document images to remain in grey-scale form. For instance, consider the images of figure 1. Figure 1a shows the original image, figure 1b the result of the binarization procedure, and figure 1c the corresponding grey-scale result after removing the background noise. In the latter case the remaining noise and the text characters are smoothed a little bit. This has the consequence of making the background noise practically invisible while the characters are more legible. Some binarization techniques support this option [6].

Traditional binarization approaches can be divided into two main categories:

1. Global thresholding methods: The pixels of the image are classified into text or background according to a global threshold. Usually, such methods are simple and fast. On the other hand, they cannot be easily adapted in case the background noise is unevenly distributed in the entire image (e.g., smears or stains) [7-8].

(j) Fisher.

Improving the Quality of Degraded Document Images

Erigna Kavalieratou and Elestathios Stamatiatos

Dept. of Information and Communication Systems Engineering,
University of the Aegean
83200 – Karlovassi, Greece
(erigna.stamatiatos@aegean.gr)

Abstract

It is common for libraries to provide public access to historical and ancient document image collections. It is common for such document images to require specialized processing in order to remove background noise and become more legible. In this paper, we propose a hybrid binarization approach for improving the quality of old documents using a combination of global and local thresholding. First, a global thresholding technique specifically designed for old document images is applied to the entire image. Then, the image areas that still contain background noise are detected and the same technique is re-applied to each area separately. Hence, we achieve better adaptability of the algorithm in cases where various kinds of noise coexist in different areas of the same image while avoiding the computational and time cost of applying a local thresholding in the entire image. Evaluation results based on a collection of historical document images indicate that the proposed approach is effective in removing background noise and improving the quality of degraded documents while documents already in good condition are not affected.

1. Introduction

Historical and ancient document collections available in libraries throughout the world are of great cultural and scientific importance [1-2]. The transformation of such documents into digital form is essential for maintaining the quality of the originals while provide scholars with full access to that information [3]. It is quite common for such documents to suffer from degradation problems [4]. Just to name a few, presence of smears, stains, background of big variations and uneven illumination, seepage of ink etc. are factors that impede in many cases may disable the legibility of the documents. Therefore, appropriate filtering methods should be developed in order to

remove noise from historical document images and improve their quality before libraries expose them to public view. Within this framework, noise is considered anything that is irrelevant with the textual information (i.e. foreground) of the document image.

Image analysis systems use binarization as a standard procedure to convert a grey-scale image to binary form. An ideal binarization algorithm would be able to perfectly discriminate foreground from background, thus, removing any kind of noise that obscures the legibility of the document image. The binary image is ideal for further processing [5] (e.g., discrimination of printed from handwritten text, recognition of the contents by applying OCR techniques etc). However, in the framework of a library collection of historical and ancient documents intended to be exposed to public view, the document images in many cases do not need further processing apart from removing the background noise and leave some traces of time behind. More importantly, given such a case, after the removal of background noise, it is possible for the document images to remain in grey-scale form. For instance, consider the images of figure 1. Figure 1a shows the original image, figure 1b the result of the binarization procedure, and figure 1c the corresponding grey-scale result; after removing the background noise. In the latter case the remaining noise and the text characters are smoothed a little bit. This has the consequence of making the background noise practically invisible while the characters are more legible. Some binarization techniques support this option [6].

Traditional binarization approaches can be divided into two main categories:

1. Global thresholding methods: The pixels of the image are classified into text or background according to a global threshold. Usually, such methods are simple and fast. On the other hand, they cannot be easily adapted in case the background noise is unevenly distributed in the entire image (e.g., smears or stains) [7-8].

(n) Otsu.

Improving the Quality of Degraded Document Images

Erigna Kavalieratou and Elestathios Stamatiatos

Dept. of Information and Communication Systems Engineering,
University of the Aegean
83200 – Karlovassi, Greece
(erigna.stamatiatos@aegean.gr)

Abstract

It is common for libraries to provide public access to historical and ancient document image collections. It is common for such document images to require specialized processing in order to remove background noise and become more legible. In this paper, we propose a hybrid binarization approach for improving the quality of old documents using a combination of global and local thresholding. First, a global thresholding technique specifically designed for old document images is applied to the entire image. Then, the image areas that still contain background noise are detected and the same technique is re-applied to each area separately. Hence, we achieve better adaptability of the algorithm in cases where various kinds of noise coexist in different areas of the same image while avoiding the computational and time cost of applying a local thresholding in the entire image. Evaluation results based on a collection of historical document images indicate that the proposed approach is effective in removing background noise and improving the quality of degraded documents while documents already in good condition are not affected.

1. Introduction

Historical and ancient document collections available in libraries throughout the world are of great cultural and scientific importance [1-2]. The transformation of such documents into digital form is essential for maintaining the quality of the originals while provide scholars with full access to that information [3]. It is quite common for such documents to suffer from degradation problems [4]. Just to name a few, presence of smears, stains, background of big variations and uneven illumination, seepage of ink etc. are factors that impede in many cases may disable the legibility of the documents. Therefore, appropriate filtering methods should be developed in order to

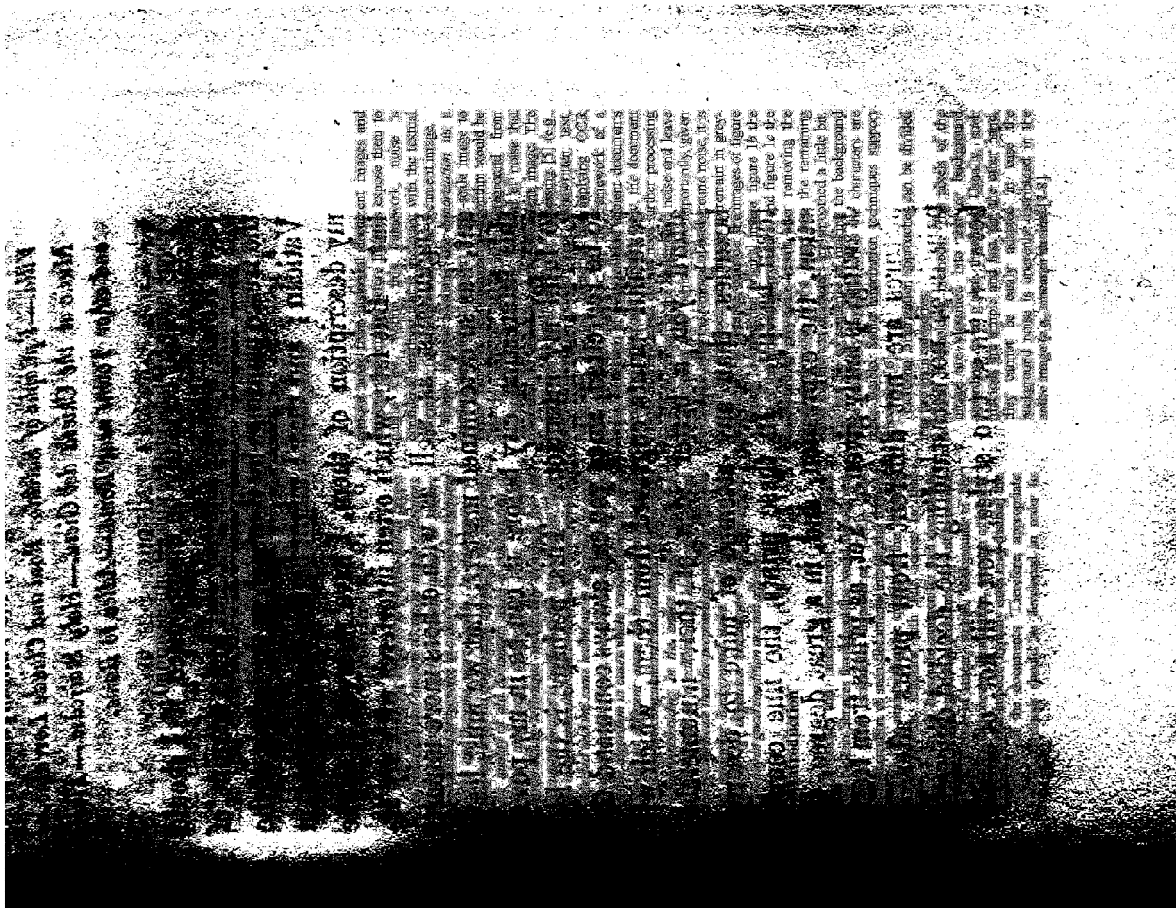
remove noise from historical document images and improve their quality before libraries expose them to public view. Within this framework, noise is considered anything that is irrelevant with the textual information (i.e. foreground) of the document image.

Image analysis systems use binarization as a standard procedure to convert a grey-scale image to binary form. An ideal binarization algorithm would be able to perfectly discriminate foreground from background, thus, removing any kind of noise that obscures the legibility of the document image. The binary image is ideal for further processing [5] (e.g., discrimination of printed from handwritten text, recognition of the contents by applying OCR techniques etc). However, in the framework of a library collection of historical and ancient documents intended to be exposed to public view, the document images in many cases do not need further processing apart from removing the background noise and leave some traces of time behind. More importantly, given such a case, after the removal of background noise, it is possible for the document images to remain in grey-scale form. For instance, consider the images of figure 1. Figure 1a shows the original image, figure 1b the result of the binarization procedure, and figure 1c the corresponding grey-scale result; after removing the background noise. In the latter case the remaining noise and the text characters are smoothed a little bit. This has the consequence of making the background noise practically invisible while the characters are more legible. Some binarization techniques support this option [6].

Traditional binarization approaches can be divided into two main categories:

1. Global thresholding methods: The pixels of the image are classified into text or background according to a global threshold. Usually, such methods are simple and fast. On the other hand, they cannot be easily adapted in case the background noise is unevenly distributed in the entire image (e.g., smears or stains) [7-8].

(m) Nível de cinza médio.



(p) Pun.

Improving the Quality of Degraded Document Images

Ergina Kavalieratou and Efsthios Stamatios

Dept. of Information and Communication Systems Engineering,

University of the Aegean

83200 - Karlovassi, Greece

ergina.stamatios@aegean.gr

Abstract

It is common for libraries to provide public access to historical and ancient document image collections. It is common for such document images to require specialized processing in order to remove background noise and become more legible. In this paper, we propose a hybrid binarization approach for improving the quality of old documents using a combination of global and local thresholding. First, a global thresholding technique specifically designed for old document images is applied to the entire image. Then, the image areas that still contain background noise are detected and the same technique is re-applied to each area separately. Hence, we achieve better adaptability of the algorithm in cases where various kinds of noise coexist in different areas of the same image while avoiding the computational and time cost of applying a local thresholding in the entire image. Evaluation results based on a collection of historical document images indicate that the proposed approach is effective in removing background noise and improving the quality of degraded documents while documents already in good condition are not affected.

1. Introduction

Historical and ancient document collections available in libraries throughout the world are of great cultural and scientific importance [1-2]. The transformation of such documents into digital form is essential for maintaining the quality of the originals while provide scholars with full access to that information [3]. It is quite common for such documents to suffer from degradation problems [4]. Just to name a few, presence of smears, stains, background of big variations and uneven illumination, seepage of ink etc. are factors that impede (in many cases may disable) the legibility of the documents. Therefore, appropriate filtering methods should be developed in order to

remove noise from historical document images and improve their quality before libraries expose them to public view. Within this framework, noise is considered anything that is irrelevant with the textual information (i.e. foreground) of the document image.

Image analysis systems use binarization as a standard procedure to convert a grey-scale image to binary form. An ideal binarization algorithm would be able to perfectly discriminate foreground from background, thus, removing any kind of noise that obscures the legibility of the document image. The binary image is then further processed [5] (e.g., discrimination of printed from handwritten text, recognition of the contents by applying OCR techniques etc.). However, in the framework of a library collection of historical and ancient documents intended to be exposed to public view, the documents in many cases do not need further processing apart from removing the background noise and leave some traces of noise behind. More importantly, given such a case, after the removal of background noise, it is possible for the document images to remain in grey-scale form. For instance, consider the images of figure 1. Figure 1a shows the original image, figure 1b the result of the binarization procedure, and figure 1c the corresponding grey-scale result after removing the background noise. In the latter case the remaining noise and the text characters are smoothed a little bit. This has the consequence of making the background noise practically invisible while the characters are more legible. Some binarization techniques support this option [6].

Traditional binarization approaches can be divided into two main categories: Global thresholding methods. The pixels of the image are classified into text or background according to a global threshold. Usually, such methods are simple and fast. On the other hand, they cannot be easily adapted in case the background noise is unevenly distributed in the entire image (e.g., smears or stains) [7-8].

(c) Porcentagem de preto.

Improving the Quality of Degraded Document Images

Erigna Kavallierou and Efsthios Stamatiou
*Dept. of Information and Communication Systems Engineering,
 University of the Aegean*
 83200 – Karlovassi, Greece
 {erigna.stamatiou}@aegean.gr

Abstract

It is common for libraries to provide public access to historical and ancient document image collections. It is common for such document images to require specialized processing in order to remove background noise and become more legible. In this paper, we propose a hybrid binarization approach for improving the quality of old documents using a combination of global and local thresholding. First, a global thresholding technique specifically designed for old document images is applied to the entire image. Then, the image areas that still contain background noise are detected and the same technique is re-applied to each area separately. Hence, we achieve better adaptability of the algorithm in cases where various kinds of noise coexist in different areas of the same image while avoiding the computational and time cost of applying a local thresholding in the entire image. Evaluation results based on a collection of historical document images indicate that the proposed approach is effective in removing background noise and improving the quality of degraded documents while documents already in good condition are not affected.

1. Introduction

Historical and ancient document collections available in libraries throughout the world are of great cultural and scientific importance [1-2]. The transformation of such documents into digital form is essential for maintaining the quality of the originals while provide scholars with full access to that information [3]. It is quite common for such documents to suffer from degradation problems [4]. Just to name a few, presence of smears, stains, background of big variations and uneven illumination, seepage of ink etc. are factors that impede (in many cases may disable) the legibility of the documents. Therefore, appropriate filtering methods should be developed in order to

remove noise from historical document images and improve their quality before libraries expose them to public view. Within this framework, noise is considered anything that is irrelevant with the textual information (i.e., foreground) of the document image.

Image analysis systems use *binarization* as a standard procedure to convert a grey-scale image to binary form. An ideal binarization algorithm would be able to perfectly discriminate foreground from background, thus, removing any kind of noise that obstructs the legibility of the document image. The binary image is ideal for further processing [5] (e.g., discrimination of printed from handwritten text, recognition of the contents by applying OCR techniques etc). However, in the framework of a library collection of historical and ancient documents intended to be exposed to public view, the document images in many cases do not need further processing apart from removing the background noise and leave some traces of time behind. More importantly, given such a case, after the removal of background noise, it is possible for the document images to remain in grey-scale form. For instance, consider the images of figure 1. Figure 1a shows the original image, figure 1b the result of the binarization procedure, and figure 1c the corresponding grey-scale result after removing the background noise. In the latter case the remaining noise and the text characters are smoothed a little bit. This has the consequence of making the background noise practically invisible while the characters are more legible. Some binarization techniques support this option [6].

Traditional binarization approaches can be divided into two main categories:
 1. Global thresholding methods: The pixels of the image are classified into text or background according to a global threshold. Usually, such methods are simple and fast. On the other hand, they cannot be easily adapted in case the background noise is unevenly distributed in the entire image (e.g., smears or stains) [7-8].

(r) Yager.

Improving the Quality of Degraded Document Images

Erigna Kavallierou and Efsthios Stamatiou
*Dept. of Information and Communication Systems Engineering,
 University of the Aegean*
 83200 – Karlovassi, Greece
 {erigna.stamatiou}@aegean.gr

Abstract

It is common for libraries to provide public access to historical and ancient document image collections. It is common for such document images to require specialized processing in order to remove background noise and become more legible. In this paper, we propose a hybrid binarization approach for improving the quality of old documents using a combination of global and local thresholding. First, a global thresholding technique specifically designed for old document images is applied to the entire image. Then, the image areas that still contain background noise are detected and the same technique is re-applied to each area separately. Hence, we achieve better adaptability of the algorithm in cases where various kinds of noise coexist in different areas of the same image while avoiding the computational and time cost of applying a local thresholding in the entire image. Evaluation results based on a collection of historical document images indicate that the proposed approach is effective in removing background noise and improving the quality of degraded documents while documents already in good condition are not affected.

1. Introduction

Historical and ancient document collections available in libraries throughout the world are of great cultural and scientific importance [1-2]. The transformation of such documents into digital form is essential for maintaining the quality of the originals while provide scholars with full access to that information [3]. It is quite common for such documents to suffer from degradation problems [4]. Just to name a few, presence of smears, stains, background of big variations and uneven illumination, seepage of ink etc. are factors that impede (in many cases may disable) the legibility of the documents. Therefore, appropriate filtering methods should be developed in order to

(q) Wulu.

Improving the Quality of Degraded Document Images

Erigna Kavalieratos and Ektrahios Stamatiadis

Dept. of Information and Communication Systems Engineering,

University of the Aegean

83200 - Karlovassi, Greece

{erigna.stamatiadis}@aegean.gr

Abstract

It is common for libraries to provide public access to historical and ancient document image collections. It is common for such document images to require specialized processing in order to remove background noise and become more legible. In this paper, we propose a hybrid binarization approach for improving the quality of old documents using a combination of global and local thresholding. First, a global thresholding technique specifically designed for old document images is applied to the entire image. Then, the image areas that still contain background noise are detected and the same technique is re-applied to each area separately. Hence, we achieve better adaptability of the algorithm in cases where various kinds of noise coexist in different areas of the same image while avoiding the computational and time cost of applying a local thresholding in the entire image. Evaluation results based on a collection of historical document images indicate that the proposed approach is effective in removing background noise and improving the quality of degraded documents while documents already in good condition are not affected.

1. Introduction

Historical and ancient document collections available in libraries throughout the world are of great cultural and scientific importance [1-2]. The transformation of such documents into digital form is essential for maintaining the quality of the originals while provide scholars with full access to that information [3]. It is quite common for such documents to suffer from degradation problems [4]. Just to name a few, presence of smear, stains, background of big variations and uneven illumination, seepage of ink etc. are factors that impede (in many cases may disable) the legibility of the documents. Therefore, appropriate filtering methods should be developed in order to

remove noise from historical document images and improve their quality before libraries expose them to public view. Within this framework, noise is considered anything that is irrelevant with the textual information (i.e., foreground) of the document image.

Image analysis systems use binarization as a standard procedure to convert a grey-scale image to binary form. An ideal binarization algorithm would be able to perfectly discriminate foreground from background, thus, removing any kind of noise that obstructs the legibility of the document image. The binary image is ideal for further processing [5] (e.g., discrimination of printed from handwritten text, recognition of the contents by applying OCR techniques etc). However, in the framework of a library collection of historical and ancient documents intended to be exposed to public view, the document images do not need further processing apart from removing the background noise and leave some "traces of time" behind. More importantly, given such a case, after the removal of background noise, it is possible for the document images to remain in grey-scale form. For instance, consider the images of figure 1. Figure 1a shows the original image, figure 1b the result of the binarization procedure, and figure 1c the corresponding grey-scale result after removing the background noise. In the latter case the remaining noise and the text characters are smoothed a little bit. This has the consequence of making the background noise practically invisible while the characters are more legible. Some binarization techniques support this option [6].

Traditional binarization approaches can be divided into two main categories:

1. Global thresholding methods: The pixels of the image are classified into text or background according to a global threshold. Usually, such methods are simple and fast. On the other hand, they cannot be easily adapted in case the background noise is unevenly distributed in the entire image (e.g., smear or stains) [7-8].

(s) Ye-Danielsson.

Figura 5.4: Resultados da aplicação dos algoritmos de binarização.

5.4.1 Resultados do Teste IV

	Total	True-Pos.	False-Pos.	True-Neg.	False-Neg.	F-Measure	NRM	MPM	MSE	PSNR
Máximos e mínimos (Fig. 5.4c)	3,852	210.645	31.663	3.495.974	18	93,006	0,005	0,161	0,008	68,850
Níveis lógicos (Fig. 5.4d)	3,999	210.662	133	3.527.504	1	99,968	0,000	0,006	0,000	92,587
Complex hull (Fig. 5.4e)	3,435	210.663	182.825	3.344.812	0	69,739	0,026	55,840	0,049	61,237
Niblack (Fig. 5.4f)	2,787	210.661	774.562	2.753.075	2	35,231	0,110	161,005	0,207	54,967
Sauvola (Fig. 5.4g)	2,388	210.663	1.354.914	2.172.723	0	23,720	0,192	269,667	0,362	52,538
Brink (Fig. 5.4h)	3,999	210.663	152	3.527.485	0	99,964	0,000	0,011	0,000	92,039
C-means (Fig. 5.4i)	1,965	210.663	2.047.779	1.479.858	0	17,064	0,290	261,526	0,548	50,745
Fisher (Fig. 5.4j)	3,999	210.662	133	3.527.504	1	99,968	0,000	0,008	0,000	92,587
Johanssen (Fig. 5.4l)	3,999	210.647	107	3.527.530	16	99,971	0,000	0,006	0,000	92,959
Mean (Fig. 5.4m)	3,883	187.278	15	3.527.622	23.385	94,120	0,056	0,004	0,006	70,165
Otsu (Fig. 5.4n)	3,999	210.662	133	3.527.504	1	99,968	0,000	0,008	0,000	92,587
Porcentagem de preto (Fig. 5.4o)	3,459	210.663	170.223	3.357.414	0	71,224	0,024	53,306	0,046	61,547
Pun (Fig. 5.4p)	2,249	210.663	1.576.902	1.950.735	0	21,085	0,224	201,863	0,422	51,880
Wulu (Fig. 5.4q)	3,909	192.585	25	3.527.612	18.078	95,511	0,043	0,004	0,005	71,280
Yager (Fig. 5.4r)	3,999	210.662	133	3.527.504	1	99,968	0,000	0,008	0,000	92,587
Ye-Danielsson (Fig. 5.4s)	3,529	210.663	137.541	3.390.096	0	75,389	0,019	45,723	0,037	62,473

Tabela IV: Demonstração dos resultados obtidos a partir da comparação das imagens binarizadas produzidas pelos algoritmos com a imagem *ground truth* correspondente à original.

De acordo com a Tabela IV, várias técnicas, dentre elas a de Níveis lógicos, conseguiram obter índices quase ideais em pelo menos duas categorias, com resultados satisfatórios em outros índices.

6. CONCLUSÃO

Com esse estudo, pode-se concluir que a análise local das informações de uma imagem contribui de maneira significativa para o processo de limiarização da mesma. Foi possível perceber, através dos resultados demonstrados, que as técnicas estudadas neste trabalho apresentaram desempenho superior em relação a outras técnicas que fazem uso de estratégias globais de limiarização. Porém, não se pode generalizar esta afirmação, já que os resultados dependem do tipo de imagem que está sendo submetida ao processo de binarização.

É possível então perceber que, até o momento, não existe uma técnica de limiarização definitiva, e sim uma técnica mais adequada ao propósito sugerido. Ainda não é possível, contudo, identificar, a priori, qual técnica está mais apta a efetuar uma binarização eficiente para uma dada imagem, senão testando uma após a outra, como neste trabalho. Este, possivelmente, é um problema igualmente desafiador ao de se encontrar a técnica definitiva.

É fato que, apesar dos esforços desenvolvidos até então, ainda há muito a ser pesquisado na área de limiarização de imagens de documentos. Seja por influência de ruídos de *background*, seja por não uniformidade do texto presente na imagem, larguras diferentes ou intensidades diferentes, os algoritmos presentes na literatura ainda são sensíveis a tais elementos, que acabam por comprometer o resultado final da limiarização, principalmente por ainda utilizarem, em demasia, uso de elementos globais na construção da imagem binarizada.

6.1 CONTRIBUIÇÕES

Neste trabalho, buscamos oferecer uma alternativa para melhoria da qualidade de imagens de documentos através do uso da operação de binarização. Utilizando como base de teste imagens de documentos com características distintas entre si, testamos e implementamos dois algoritmos desenvolvidos especialmente para limiarização de imagens de texto e comparamos os seus resultados com outros algoritmos clássicos de binarização de imagens, apresentando seus resultados de forma visual e técnica.

6.2 TRABALHOS FUTUROS

Esta seção sugere pontos de investigação para possíveis melhorias no desempenho das técnicas estudadas neste trabalho, identificando problemas encontrados durante a implementação e possíveis causas para tal comportamento.

6.2.1 SUGESTÕES SOBRE A TÉCNICA DE LIMIAÇÃO POR MÁXIMOS E MÍNIMOS LOCAIS

A técnica de limiarização por máximos e mínimos se mostrou muito eficiente para limiarização de imagens de documentos datilografados, porém apresentou desempenho insatisfatório quando a imagem possui textos de tamanhos diferentes, como ilustrado neste trabalho. Isso se deve à dependência do algoritmo sobre a estimativa da largura do texto durante o processamento da imagem. Estudos posteriores podem encontrar uma maneira mais adequada de se estimar a largura, tratando-a como um item local, e não global como a técnica sugere.

Outra possibilidade de melhoria seria a agregação de alguma técnica de pós-processamento para remoção de ruídos conhecidos na literatura, como Sal e Pimenta, por exemplo, que frequentemente apareciam nas imagens binarizadas. Técnicas de análise através de componentes conectados, presentes imagens criadas a partir de vários limiares diferentes podem ajudar a melhorar o resultado visual da binarização.

6.2.2 SUGESTÕES SOBRE A TÉCNICA DE LIMIAÇÃO POR ANÁLISE DE NÍVEIS LÓGICOS

A técnica de limiarização por níveis lógicos provou ser mais eficiente do que a de máximos e mínimos locais, sendo inclusive mais eficiente do que mais da metade dos algoritmos testados neste trabalho. Porém, faz uso de uma carga de processamento muito alta, o que dificulta sua aplicação em um ambiente diferente do acadêmico. Um estudo sobre melhorias de desempenho computacional, uso de estruturas de dados mais eficientes, ou mesmo a simplificação de alguns critérios no algoritmo poderiam melhorar

consideravelmente a velocidade do processamento, diminuindo a carga computacional envolvida.

A técnica, assim como o algoritmo de máximos e mínimos, também apresenta uma dependência da correta estimativa da largura do texto, apesar desta não ser tão forte. Novamente, um método que utilize uma estimativa local de largura de texto pode contribuir para a melhoria na eficácia da binarização resultante.

7. REFERÊNCIAS

- [1] PEDRINI, Hélio; SCHWARTZ, William Robson. *Análise de Imagens Digitais* (1ª edição). São Paulo, Thomson Learning, 2008.
- [2] GONZALEZ, Rafael C; WOODS, Richard E. *Processamento Digital de Imagens* (3ª edição). São Paulo, Pearson Prentice Hall, 2010.
- [3] GOMES, Jonas; VELHO, Luiz. *Computação gráfica: Imagem* (2ª edição). Rio de Janeiro, Instituto de Matemática Pura e Aplicada (IMPA), 2002.
- [4] KATZ, S.W; BRINK, A.D. “*Segmentation of Chromosome Images*”, IEEE, 1993, pp 85-90.
- [5] JAWAHAR C. V; BISWAS, P. K; RAY, K. “Investigations On Fuzzy Thresholding Based On Fuzzy Clustering”, *Pattern Recognition*, Vol. 30, No. 10, pp. 1605-1613, 1997.
- [6] JOHANNSEN G; BILLE, J. “*A Threshold Selection Method using Information Measures*”, *Proceedings, 6th Int. Conf. Pattern Recognition*, Munich, Germany, pp.140-143, 1982.
- [7] LI, L; GONG, J; CHEN, W. “Gray-level Image thresholding based on Fisher linear projection of two-dimensional histogram”, *Pattern Recognition*, Vol 30, No 5, pp 743-749 , 1997.
- [8] GLASBYE, C. A. “*An Analysis of Histogram-Based Thresholding Algorithm*”, *CVGIP: Graphical Models and Image Processing*, v.55, No 6, 1993, pp 53 2-537.
- [9] YAGER, R. R. “*On the Measures of Fuzziness and Negation.Part.1: Membership in the Unit Interval*”, *Int. Journal of Gen. Sys*, v.5, pp 221-229, 1979.
- [10] PARKER J. R. “*Algorithms for Image Processing and Computer Vision*”, Editor John Wiley and Sons, pp 139-148, 1996.

- [11] PAL, N. R; PAL, S. K. “Entropic thresholding”, *Signal Processing* 16, pp. 97-108, 1989.
- [12] KAPUR, J. N; SAHOO, P. K; WONG, K. C. “A New method for gray-level picture thresholding using the entropy of the histogram”, *CVGIP*, 29, pp. 273-285, 1985.
- [13] PUN, T. “A new method for gray level picture thresholding using the entropy of the histogram”, *Signal Processing* 2, pp. 223 – 237, 1980.
- [14] PUN, T. “Entropic thresholding: a new approach”, *CVGIP*, 16, pp. 210 – 239, 1981.
- [15] WU, Lu; MA, Songde; LU, Hanqing. “An Effective Entropic thresholding for Ultrasonic Images”, *IEEE*, pp. 1552-1554, 1998.
- [16] OTSU, N. “A *Threshold Selection Method from Gray-level Histograms*”, *IEEE Transactions on Systems, Man and Cybernetics*, v. SMC 9, no 1, pp.62-66, 1979.
- [17] SAUVOLA, J; PIETAKSINEN, M. “Adaptive document image binarization” *Pattern Recognition*. 33, 225–236 ~2000
- [18] NIBLACK, W. “An Introduction to Digital Image Processing”. Prentice Hall, Englewood Cliffs, 1986.
- [19] SEZGIN, M; SANKUR, B. “Survey over image thresholding techniques and quantitative performance evaluation”, *Journal of Electronic Imaging* 13(1), 146–165, 2004.
- [20] HE, J; DO, Q. D. M; DOWNTON, KIM, A. C., J. H. “A Comparison of Binarization Methods for Historical Archive Documents”, *IEEE*, pp. 538 - 542 Vol. 1, 2005
- [21] SU, Bolan; LU, Shijan; TAN, Chew Lim. “Binarization of Historical Document Images Using the Local Maximum and Minimum”. *Workshop on Document Analysis and Systems*, Boston, EUA, 2010.

- [22] SU, Bolan; LU, Shijan; TAN, Chew Lim. "Document Image Binarization Using Background Estimation and Stroke Edges". International Journal on Document Analysis and Recognition, vol. 13, pp. 303-314, 2010.
- [23] ZHAO, Aiguo; KAMEL, Mohamed. "Extraction of Binary Character/Graphics Images from Grayscale Document Images". Graphical Models and Image Processing, vol. 55, No. 3, pp. 203-217, 1993.
- [24] YAN, Hong; YANG, Yibing; "An adaptative logical method for binarization of degraded document images". Pattern Recognition, vol. 33, pp. 787-807, 2000.
- [25] GATOS, B; NTIROGIANNIS, K; PRATIKAKIS, I. "ICDAR 2009 Document Image Binarization Contest". 10th International Conference on Document Analysis and Recognition, 2009
- [26] MELLO, Carlos A. B; NEVES, Renata F. P. "A Local Thresholding Algorithm for Images of Handwritten Historical Documents".