



Universidade Federal de Pernambuco
Centro de Informática
Graduação em Engenharia da Computação
2014.2

Uma abordagem para binarização de documentos históricos manuscritos utilizando combinação de métodos global e local adaptativo

Trabalho de Graduação

Aluno: Diego Phoenix Alves Menor

Orientador: Carlos Alexandre Barros de Mello

Recife, 23 de Fevereiro de 2015

Sumário

Agradecimentos.....	3
Resumo	4
Capítulo 2 - Métodos de binarização.....	10
2.1 – Algoritmo de Otsu	11
2.2 – Algoritmo de Niblack.....	13
2.3 – Algoritmo de Sauvola	14
2.4 – Método de Su-Lu-Tan.....	15
2.5 - Método de Howe	16
2.6 – Método de R.F. Moghaddam <i>et al.</i>	16
2.7 - Método de Lelore e Bouchara	16
2.8 - Método de M.A. Ramírez-Ortegón <i>et al.</i>	17
Capítulo 3 - Algoritmo de Gatos <i>et al.</i>	18
3.1 - Estimativa de Fundo	19
3.2 - Normalização da Imagem	24
3.3 - Binarização global e pós-processamento	26
3.4 - Estimativa da largura de traço e cálculo do contraste para binarização local da imagem normalizada	29
3.5 - Combinação dos resultados de binarização global e local	33
Capítulo 4 - Resultados experimentais	37
4.1 - Medidas de avaliação	39
4.2 - Análise comparativa.....	42
Capítulo 5 – Conclusão	52
Referências	54

Agradecimentos

A minha mãe, pela coragem de ter enfrentado desafios em prol da minha educação. Ao meu pai, que sabiamente me mostrou a arte do ensino e me fez valorizar cada parte. A minha família, que foi minha base e meu porto seguro nos momentos de dificuldade. A minha esposa, que encarou tudo do meu lado, sem hesitar, e me deu o apoio que eu precisava pra alcançar meus objetivos. Aos amigos, que me encorajaram a seguir meus sonhos e nunca perder o brilho no olhar. Ao meu orientador, que me abriu portas, me ensinou quando eu não queria aprender e com paciência acreditou na minha competência. A todos os professores que honraram seu papel de educador, enxergando o aluno como personagem principal de todo o processo de ensino. E a todos os obstáculos que enfrentei, pois construíram a pessoa e o profissional que sou hoje, aluno ontem, hoje e sempre.

Resumo

Uma das etapas fundamentais e críticas em muitas técnicas de processamento de imagens é a binarização, que consiste em transformar uma imagem digital colorida ou em tons de cinza em uma imagem com dois tons (em geral, preto e branco) onde há uma separação do objeto de observação e do fundo da imagem. Em imagens de documentos, o objeto de observação é a letra (ou tinta), que é caracterizado pela cor preta na imagem binarizada. Por sua vez, o fundo é o papel, que caracteriza-se pela cor branca na imagem binarizada. Todavia, imagens de documentos antigos apresentam alguns desafios para o processo de binarização, geralmente ocasionados, no documento original, pela degradação do papel ou da tinta com o tempo ou por problemas de manuseio inadequado, como imagens amareladas e com deformações no documento (dobras). Imagens de documentos históricos manuscritos em particular apresentam uma série de problemas devido ao uso de canetas tinteiro ou de pena, como manchas que se espalham no papel com o tempo, iluminação não uniforme, letras enfraquecidas e vazamento de tinta de uma face para outra.

Embora muitos métodos de binarização tenham sido propostos, ainda não há um método que obtenha um resultado ótimo para todos os tipos de imagens de documentos históricos manuscritos e seus problemas. Alguns, porém, conseguem obter um resultado bem satisfatório e próximos ao ideal. O método de Gatos *et al.* [15] é um método de binarização destinado a realizar a binarização de imagens de documentos históricos manuscritos, tratando seus problemas específicos mencionados anteriormente. Esse método propõe a combinação das técnicas global de Otsu [3] e da técnica de binarização local adaptativa proposta por Niblack [4], visando obter as vantagens de ambas as abordagens.

O objetivo deste trabalho é implementar, analisar e comparar o desempenho do método de Gatos *et al.* com outros métodos do estado-da-arte, utilizando para isso um conjunto de imagens e medidas da competição DIBCO'13.

Capítulo 1 - Introdução

No processo de evolução humana, a comunicação através da escrita teve papel fundamental para o desenvolvimento da economia e das sociedades, desde sua formatação cuneiforme e silábica adotada na Mesopotâmia entre os sumérios e acádios, passando por adaptações através da escrita hieroglífica dos Egípcios. A escrita tornou possível a comunicação não presencial, permitindo que um conjunto de informações escritas num texto fossem guardados e repassados ao longo de diferentes lugares e gerações, quebrando de certa forma uma barreira de espaço e tempo trazida com a comunicação apenas oral. Assim, se podiam ter registros históricos, financeiros, conhecimentos arquivados em papiros e pergaminhos que poderiam ser enviados a outros lugares e consultados anos depois.

Nos tempos de hoje, apesar da grande disseminação dos meios digitais para armazenamento e tráfego de informações escritas, o papel continua sendo bastante utilizado. Porém, o papel apresenta uma série de problemas como material de preservação de informação: pode ser facilmente destruído, manchado, rasgado ou molhado, comprometendo parcialmente ou integralmente o conteúdo escrito naquele papel; com o tempo, ocorre um desgaste natural que pode levar ao "amarelamento" das folhas, enfraquecimento das letras e esfacelamento do papel, principalmente quando exposto a condições ambientais adversas, como luminosidade degradante e umidade do ar. Esses problemas podem ser precipitados no caso de documentos que são constantemente consultados e manuseados, principalmente de forma incorreta. Em algumas instituições e empresas, há também a questão do espaço físico alocado para armazenar uma grande quantidade de papéis e documentos, entre notas fiscais e contratos, o que requer esforços para que esses materiais fiquem organizados e bem condicionados.

Uma alternativa a esses problemas que acerbam o uso de papel como preservador de informação é a digitalização desses documentos. O processo de digitalização consiste em converter um

documento em formato físico de papel em um documento no formato digital. Um documento digitalizado é uma imagem em formato digital definido como k camadas de matrizes $m \times n$ (m linhas, n colunas). $k=1$ para imagens na escala de 256 tons de cinza; e $k=3$ para imagens no formato de 16 milhões de cores do sistema RGB, por exemplo, onde as matrizes correspondem aos tons vermelho (*Red*), verde (*Green*) e azul (*Blue*) da imagem, como definido em [1]. Essa imagem é formada por pixels, que são quadrados armazenando um valor correspondente a uma intensidade luminosa. Eles são representados por cada célula da matriz, que possui valores na escala apropriada (como por exemplo, 0 a 255 na escala de cinza do sistema RGB) para quantizar o nível de luminosidade do pixel correspondente. Quanto mais claro for o pixel, maior este valor, chegando ao nível 255 para a cor branca e 0 para cor preta no caso da escala de cinza no sistema de cores RGB.

Documentos digitalizados não sofrem, *a priori*, com os efeitos de degradação do tempo nem com os problemas de manuseio e condições ambientais adversas. Possuem ainda a vantagem de serem facilmente armazenados através de mídias como CDs ou DVDs, distribuídos e consultados pela internet ou redes locais de forma que não compromete a integridade do documento original, ainda com a vantagem do controle de acesso e maior segurança provido por esse meio. Apesar da digitalização também estar suscetível a problemas, como por exemplo, falhas na mídia digital onde está o arquivo ou erros decorrentes de uma falha no processo de digitalização, o qual pode comprometer a qualidade da imagem gerada, em geral, são problemas que podem ser sanados. O primeiro problema pode ser resolvido realizando cópias daquele documento digitalizado em várias mídias diferentes, replicando para que ele possa ser restaurado caso alguma das mídias seja danificada. No segundo caso, podemos aplicar técnicas de processamento de imagem para tentar melhorar a imagem digitalizada e retirar alguns defeitos trazidos no processo de digitalização, como sombras e dobraduras no documento original. Evita-se uma nova digitalização, dada a fragilidade que alguns documentos têm (como documentos históricos).

Uma das etapas cruciais e críticas de várias técnicas de processamento de imagem é a binarização ou limiarização, onde uma

imagem em tons de cinza (ou colorida) é convertida para uma imagem binária (apenas dois tons, em geral, preto e branco)(Figura 1). Em termos de documentos, é comum que os pixels brancos correspondam ao fundo da imagem original (*background*) e os pixels pretos correspondam ao objeto de observação (*foreground*). No caso de documentos, o *background* é o papel e o *foreground* é a tinta. Para imagens de documentos, essa etapa é necessária e fundamental num sistema de OCR (reconhecimento óptico de caracteres – *Optical Character Recognition*), pois permite mais facilmente que as letras sejam separadas do fundo para posterior reconhecimento dos caracteres [2]. Uma melhor qualidade de binarização da imagem, com poucos ou nenhum ruído, propicia uma maior precisão no reconhecimento de caracteres, pois ruídos na imagem binarizada podem ser confundidos com caracteres ou causar deformações em caracteres que dessa forma não serão reconhecidos.

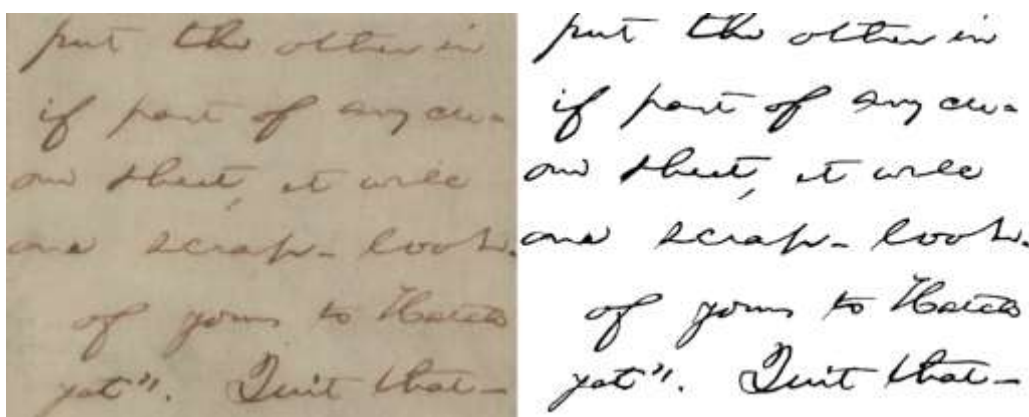


Figura 1– Resultado do processo de binarização: (esquerda) Imagem original e (direita) sua versão em escala de preto e branco

Na prática, um ponto de corte na escala de cores estabelece uma separação entre os tons mais claros, considerando-os como parte do papel (fundo), e os tons mais escuros, considerando-os como tinta. O problema principal da binarização é encontrar qual o ponto de corte ideal para determinada imagem.

Imagens de documentos históricos manuscritos particularmente apresentam uma série de problemas devido à degradação com o tempo, tanto do papel quanto da tinta, uso extenso e condições

ambientais adversas. Alguns dos principais problemas para binarização de imagens de documentos históricos são: degradação do papel (Figura 2.a), sombras ou manchas (Figura 2.b), iluminação não uniforme (Figura 2.c), deformação do documento (dobras) (Figura 2.d), vazamento da tinta de uma face para outra (efeito *bleed-through*) (Figura 2.e), degradação da letra e enfraquecimento do texto (*faint characters*) (Figura 2.f). A degradação do papel faz com que este assumam tons amarelados ou que surjam manchas que deixam regiões mais escuras. Nos dois casos, a presença da tinta pode gerar confusão na binarização. Caso a tinta não sofra degradação, ela pode ser confundida ao aparecer em uma região escurecida do papel. Se a tinta sofrer degradação e enfraquecer, ela pode tornar-se tão ou até mais clara que o papel. Essa falta de contraste entre tinta e fundo dificulta consideravelmente a binarização da imagem. Em alguns casos, a tinta do texto escrito em uma face acaba penetrando no papel e vazando para a outra face do papel, podendo fazer com que as letras de uma face sejam reconhecidas como letras da outra face. Há também os problemas devidos à forma incorreta de digitalização desses documentos, como em casos que uma iluminação não adequada incide no papel em um ângulo em que há reflexão indevida, causando clarões ou regiões mais escuras na imagem, caracterizando um caso de iluminação não uniforme da imagem.

Certos documentos mais espessos, com grande quantidade de páginas, apresentam dificuldade na digitalização por que as páginas não podem ser destacadas para que cada uma seja digitalizada individualmente. Isso acarreta uma imagem digitalizada com letras encurvadas (efeito conhecido como *warping*) e uma mancha na área da dobradura, dificultando também o processo de binarização.

Alguns outros problemas surgem relacionados à forma do conteúdo do documento. Em alguns casos, podemos encontrar tamanhos de objetos incompatíveis com a cena, no caso de manchetes em jornais, por exemplo, que geralmente possuem letras de tamanho destoante com relação a outros blocos de texto, dificultando a estimativa da largura de traço do texto em geral, um importante parâmetro para alguns algoritmos de binarização.

Motivado pelo desafio trazido pelas dificuldades acerca do processo de binarização de documentos históricos manuscritos, por problemas em geral decorrentes da degradação do papel com o tempo e pelo uso comum de caneta tinteiro e de pena para escrita, o que resulta em alguns problemas citados anteriormente como *bleed-through* e manchas, o objetivo deste trabalho é a apresentação, implementação e análise da técnica proposta por Gatos *et al.* [15], comparando seu desempenho com a de outros métodos conhecidos. Essa técnica propõe a combinação dos métodos de binarização global de Otsu [3] e do método de binarização local proposto por Niblack [4], visando alcançar um melhor desempenho geral.



Figura 2 - Documentos históricos manuscritos e seus problemas: (a) papel degradado; (b) manchas; (c) iluminação não-uniforme; (d) documento deformado; (e) vazamento de tinta entre faces (*bleed through*); (f) enfraquecimento das letras (*faint characters*)

Capítulo 2 - Métodos de binarização

Diversas técnicas de binarização já foram sugeridas, definindo basicamente 2 tipos de abordagem: global, onde o método define um único ponto de corte para toda a imagem, e o local, onde a imagem é dividida em regiões e a binarização é realizada considerando características de cada região (como média e desvio padrão, por exemplo). As técnicas de binarização local são adaptativas, levando em consideração mudanças de luminosidade, por exemplo, ao longo da imagem. Essas técnicas são, em geral, as que apresentam melhor resultado na binarização de documentos históricos. Entretanto, são as que demandam mais tempo e processamento para serem realizadas.

Neste capítulo, analisamos alguns dos principais métodos de binarização que servem de base para o estudo e análise comparativa com o método principal analisado neste trabalho.

Os métodos de Otsu[3] e Niblack[4] descritos a seguir, são combinados no método apresentado, através de seus resultados melhorados e adaptados ao contexto de documentos históricos manuscritos e seus problemas.

São também analisados os métodos que obtiveram as melhores pontuações na DIBCO'13[5]. As séries DIBCO e H-DIBCO (incluindo a DIBCO'13), são competições que, desde 2009, reúnem participantes de todo o mundo para revelar seus recentes avanços e avaliar os desempenhos de suas técnicas de binarização, assim como estabelecer medidas de avaliação de desempenho para serem aplicados na comparação entre resultados.

Embora haja muitos métodos propostos na literatura [6], um grande problema que dificulta a binarização e que o torna ainda um estudo em aberto é que diferentes técnicas de binarização têm desempenhos diferentes para diferentes conjuntos de imagens, não havendo ainda uma técnica de binarização geral que tenha um desempenho ótimo para todos os tipos de imagens e seus problemas específicos.

2.1 – Algoritmo de Otsu

Descrição

O método de binarização global proposto por Otsu[3] foi um dos primeiros a serem apresentados e ainda hoje é referência para outros métodos por obter resultados bastante satisfatórios, com pouco ruído de fundo, ao binarizar imagens que apresentam uma distribuição clara de classes distintas de intensidade para toda a imagem. Seu resultado, entretanto, apresenta-se ineficiente quando o método é aplicado a imagens de documentos históricos, que geralmente apresentam manchas e iluminação desigual. Esse método sugere a divisão de duas classes distintas na imagem, uma representando o texto ($C0$) e outra representando o fundo ($C1$). Ambas as classes tem suas próprias características estatísticas, como média, desvio padrão e variância.

Assim como todo método de binarização global, o método de Otsu estabelece um ponto de corte global em toda a imagem original $f(x,y)$ em escala de cinza onde os pixels da imagem resultado $g(x,y)$ são definidos da seguinte forma:

$$g(x,y) = \begin{cases} C0, & \text{se } f(x,y) < T \\ C1, & \text{caso contrário} \end{cases}$$

onde, para imagens, $C0$ é representado pelo valor 0 (preto), correspondente à letra, e $C1$ é representado pelo valor 1 (branco), correspondente ao papel.

Para encontrar o valor do ponto de corte T de forma não paramétrica, o método define outras três medidas (λ , κ e η), chamadas funções objetivo, que determinam a eficiência do método ao escolher um certo valor T , transformando o problema de escolha de T num problema de otimização, facilitando assim o cálculo.

O método se baseia na análise do histograma da imagem. Um histograma é uma função que, para cada valor de intensidade da escala (0 a 255 na escala de cinza), leva ao número de pixels na imagem que possuem aquele valor (frequência). Dividindo esse

número pelo total de pixels na imagem, é obtida a distribuição de probabilidades p_i de cada tom de cinza i . Suponha que um certo valor t divida o histograma em duas classes: C_0 , agrupando valores de 0 a t , e C_1 , que agrupa os valores de $t+1$ a 255. Define-se então as probabilidades $\omega_0(t)$ e $\omega_1(t)$, e níveis médios $\mu_0(t)$ e $\mu_1(t)$ das classes C_0 e C_1 , respectivamente, da seguinte forma:

$$\omega_0(t) = \sum_{i=0}^t p_i$$

$$\omega_1(t) = \sum_{i=t+1}^{255} p_i = 1 - \omega_0(t)$$

$$\mu_0(t) = \frac{\mu(t)}{\omega_0(t)}, \text{ onde } \mu(t) = \sum_{i=0}^t i \cdot p_i$$

$$\mu_1(t) = \frac{\sum_{i=t+1}^{255} i \cdot p_i}{\omega_1} = \frac{\mu_T - \mu(t)}{1 - \omega_0(t)}, \text{ onde } \mu_T = \sum_{i=0}^{255} i \cdot p_i$$

A medida mais fácil de obter das funções do objeto é a variância interclasse η , cuja maximização depende apenas de:

$$\begin{aligned} \sigma_B^2(t) &= \omega_0(t) * (\mu_0(t) - \mu_T) + \omega_1(t) * (\mu_1(t) - \mu_T) \\ &= \omega_0(t) * \omega_1(t) * (\mu_0 - \mu_1)^2 \end{aligned}$$

O melhor valor de t é aquele que maximiza η e conseqüentemente maximiza $\sigma_B^2(t)$.

$$\sigma_B^2(t^*) = \max(\sigma_B^2(t))$$

Esse valor (t^*) é então usado como ponto de corte na binarização do método de Otsu.

Parâmetros

Nenhum

Passos para realização do método

- 1) Encontre a distribuição de probabilidades a partir do histograma;
- 2) Para cada valor de intensidade t , variando de 0 a 255, separe os pixels em duas classes de acordo com o ponto de corte t ;
- 3) Encontre a probabilidade e o nível médio de cada classe;
- 4) Calcule $\sigma_B^2(t)$. Se esse for o máximo encontrado até então, atribua $t^* = t$.

2.2 – Algoritmo de Niblack

Descrição

O método de binarização local de Niblack [4] é um dos mais simples métodos locais, sendo utilizado como base em muitos outros métodos. Ele obtém bom resultado na captação dos pixels textuais, até mesmo aqueles pixels de letras esvaecidas, que em outros métodos são confundidos com o fundo. Porém o resultado apresenta muito ruído de fundo, classificando falsamente como letra muitos pixels do papel.

Esse método calcula para cada pixel i da imagem o valor da média μ e desvio padrão δ dos pixels situados na região retangular (janela) de tamanho $m \times n$ centrado no pixel i em questão. Então o pixel i é classificado entre fundo ou letra de acordo com o ponto de corte estabelecido pela equação:

$$T_{Niblack}(i) = \mu(i) + k \cdot \delta(i)$$

onde k é o parâmetro peso do desvio padrão. Se o valor de intensidade do pixel for menor que T , o pixel é considerado como letra. Caso contrário, é considerado como papel.

Os parâmetros de tamanho da janela e peso k moldam o resultado do algoritmo para uma dada imagem. Tamanhos de janela grandes demais podem acabar agrupando áreas com diferentes níveis

de luminosidade, enquanto que pequenas janelas não reúnem informação suficiente para a correta binarização.

Parâmetros

- As dimensões m e n da janela
- k - Peso do desvio padrão

Passos para realização do método

- 1) Para cada pixel i , calcule a média e desvio padrão dos pixels enquadrados pela janela de tamanho $m \times n$;
- 2) Calcule o ponto de corte para o pixel i através da equação anterior.
- 3) Se o valor de intensidade do pixel i for menor que o ponto de corte, atribui-se o valor 0 classificando-o como letra. Caso contrário, recebe o valor 1, classificando-o como papel.

2.3 – Algoritmo de Sauvola

O método de binarização de Sauvola [7] melhora o desempenho do algoritmo de Niblack, adaptando o parâmetro de peso do desvio padrão de acordo com a equação:

$$T_{Sauvola}(i) = \mu \cdot \left(1 - k \cdot \left(1 - \frac{\delta(i)}{R}\right)\right),$$

onde R é o novo parâmetro, denominado faixa dinâmica do desvio padrão. Dessa forma, Sauvola diminui a sensibilidade ao parâmetro de peso k , mas acaba se tornando sensível ao parâmetro de faixa.

O resultado da binarização utilizando o método de Sauvola em comparação ao do Niblack apresenta melhorias especialmente quando aplicado a imagens de documentos que possuem manchas e condições desfavoráveis de iluminação. Embora o resultado apresente menor ruído de fundo, a taxa de detecção de texto também diminui enquanto problemas de vazamento de tinta de uma face para outra permanece na maioria dos casos desse tipo.

Parâmetros

- Tamanho $m \times n$ da janela
- k - Peso do desvio padrão
- R - Faixa dinâmica do desvio padrão

Passos para realização do método

- 1) Para cada pixel i , calcule a média e desvio padrão dos pixels enquadrados pela janela de tamanho $m \times n$;
- 2) Calcule o ponto de corte para o pixel i através da equação anterior.
- 3) Se o valor de intensidade do pixel i for menor que o ponto de corte, atribui-se o valor 0 classificando-o como letra. Caso contrário, recebe o valor 1, classificando-o como papel.

2.4 – Método de Su-Lu-Tan

Esse método, desenvolvido por *Su et al.*, foi apresentado na DIBCO'13 alcançando o 1º lugar no *ranking* geral. Consiste basicamente num procedimento de 4 passos. Primeiro o contraste local e o gradiente local da imagem são combinados usando uma função exponencial com um fator adaptativo. Segundo, o contraste local é combinado com o mapa de contorno para extrair o contorno do conteúdo textual da imagem. Terceiro, a imagem do documento é binarizada por um ponto de corte local que é baseada no mapa de contorno e na estimativa da largura de traço. Por último, é aplicado um pós-processamento para melhorar a qualidade do resultado da binarização.

2.5 - Método de Howe

Esse método, desenvolvido por Nicholas Howe [8], foi apresentado na DIBCO'13, alcançando o 2º lugar geral. O método emprega um algoritmo base que minimiza uma função de energia global baseado no laplaciano da imagem, usando pares de pressão entre pixels vizinhos modulado pela localização das bordas de Canny. O método então automaticamente determina configurações promissoras dos pares de pressão e parâmetros de Canny, ao aplicar critérios de estabilidade aos resultados da binarização.

2.6 – Método de R.F. Moghaddam *et al.*

Esse método utiliza a mesma técnica da energia laplaciana usada por Howe no método 2.5, combinando-a com um framework descrito em [9]. Obteve o 3º lugar na DIBCO'13.

2.7 - Método de Lelore e Bouchara

Esse método é detalhado em [10]. Como primeiro passo, a imagem é redimensionada (2x), usando interpolação linear. Posteriormente, é aplicada uma detecção robusta de áreas contendo texto através da combinação do resultado de um algoritmo de agrupamento alcançado em torno da borda detectada com dois pontos de corte diferentes. A partir dessa junção, identificam-se áreas problemáticas onde ocorre alguma diferença entre as duas imagens. O resultado final da binarização é obtido recalculando a classe de pixels em torno dessas áreas problemáticas, utilizando uma vizinhança assimétrica.

Este método atingiu a 4ª colocação na DIBCO'13.

2.8 - Método de M.A. Ramírez-Ortegón *et al.*

Este método é composto por três módulos. O primeiro detecta a região de interesse (ROI) baseado na variância das intensidades de tons de cinza [11]. Depois, uma versão melhorada do método de transição é executada de acordo com o ROI encontrado [12]. Finalmente, a imagem binária é pós-processada por vários operadores para restaurar contornos [13] e remover artefatos [14].

Esse método obteve o 5º lugar na competição da DIBCO'13.

Capítulo 3 - Algoritmo de Gatos *et al.*

A técnica proposta por Gatos *et al.* combina os métodos de binarização global de Otsu e local de Niblack, utilizando técnicas no pré-processamento da imagem como *inpainting*, estimativa de fundo, normalização da imagem e cálculo do contraste para se obter um melhor resultado do processo de binarização. Essas técnicas buscam contornar os problemas encontrados na binarização de documentos históricos, como manchas, sombras e iluminação não uniforme do fundo, e em especial os problemas de degradação encontrados em documentos históricos manuscritos, como esvaecimento da letra (*faint characters*) e vazamento da tinta (*bleed-through*) causados pelo uso de canetas de pena ou tinteiro em papéis muito finos. Para estimativa de fundo, é utilizado um procedimento de *inpainting* que melhora o resultado da combinação das binarizações. O cálculo do contraste é realizado de forma global, executando combinações de componentes conectados para detecção mais completa dos caracteres esvaecidos quando possível. Adicionalmente, é realizado um pós-processamento em passos intermediários para remoção de pequenos ruídos que a estimativa de fundo e a normalização da imagem não conseguiram detectar e remover, evitando dessa forma a combinação com componentes ruidosos do fundo. Os estágios do algoritmo são mostrados na figura 3 a seguir:

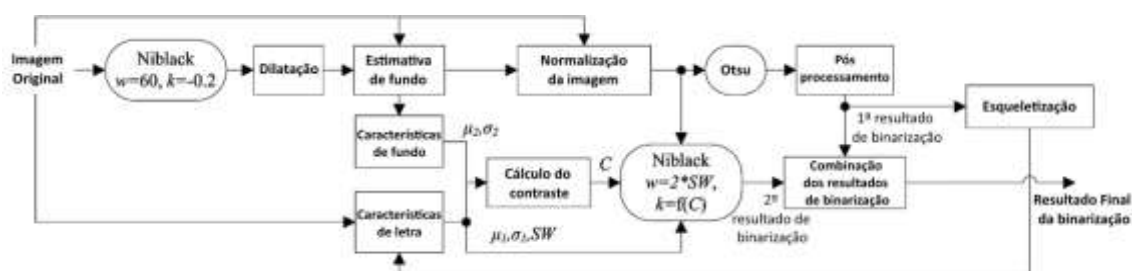


Figura 3 - Etapas do processo de binarização proposto por Gatos *et al.*

3.1 - Estimativa de Fundo

A primeira etapa do algoritmo de Gatos *et al.* é a estimativa de fundo (*background estimation* ou *background subtraction*), uma técnica para extração do que é considerado como fundo da imagem, separando-o do objeto de observação, que no nosso caso são as letras, do documento digitalizado [16]. Essa técnica, sucedida de uma normalização da imagem, é útil na tentativa de balancear a iluminação do fundo para aquelas imagens de documentos históricos digitalizados que sofrem de iluminação não uniforme, sombras e manchas. Essa técnica modifica a imagem transformando-a em uma imagem de distribuição bi-modal bem definida, que pode ser binarizada globalmente com resultados satisfatórios. No método proposto no algoritmo de Gatos *et al.*, a estimativa de fundo é realizada utilizando a técnica de *inpainting*, onde o objeto de observação da imagem é retirado, resultando numa imagem apenas com o fundo.

O processo de *inpainting* recebe como parâmetro a região ou elementos da imagem que se deseja extrair, chamado de máscara. São os pixels referentes às letras da imagem. Nesta etapa, é utilizado o resultado do método de binarização de Niblack, com parâmetros fixos $w=60$ e $k=-0,2$, como máscara para o *inpainting*. Esses parâmetros são usados inicialmente, pois conseguem lidar efetivamente com a maioria dos casos. Antes de ser usado como máscara para o *inpainting*, o inverso (o que é preto se transforma em branco e vice-versa) do resultado de Niblack é dilatado [1] com uma máscara 3x3, fazendo com que os pixels de letra dilatem para que pixels do fundo próximos a pixels de letras, que podem possuir intensidade próximas às intensidades das letras, sejam excluídos. A figura 4 mostra um exemplo do resultado dessa etapa.

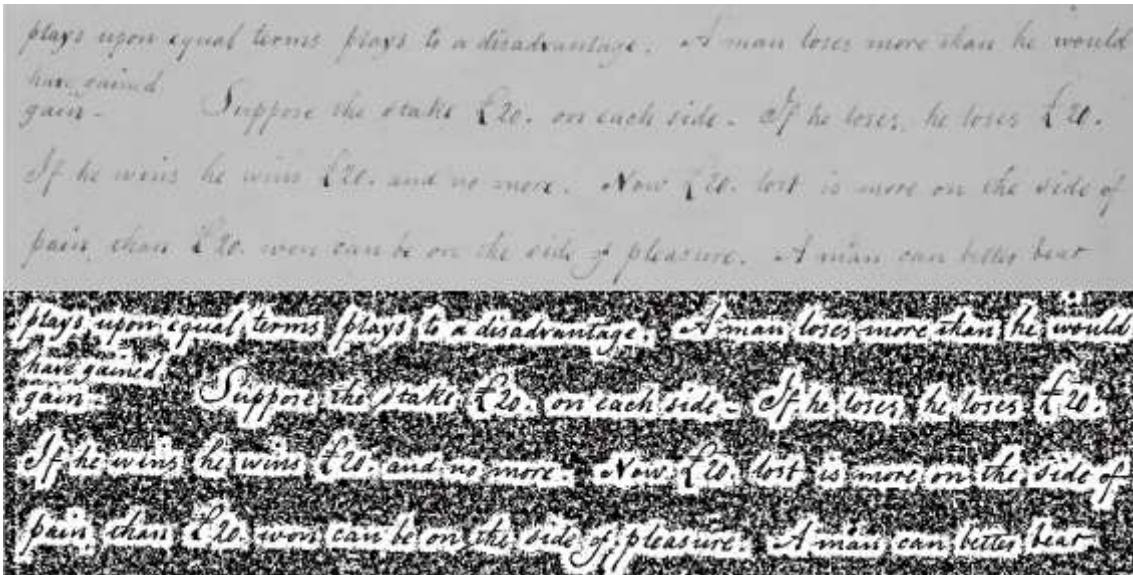


Figura 4—(topo) Imagem HW1 retirado do dataset da DIBCO'13; (embaixo) o resultado do processo de Niblack para $w=60$ e $k=-0,2$

Como proposta de *inpainting*, o algoritmo varre toda a imagem 5 vezes, onde as 4 primeiras são na seguinte ordem: da esquerda pra direita e de cima para baixo (EDCB), da esquerda pra direita e de baixo para cima (EDBC), da direita para esquerda e de cima para baixo (DECB) e por fim, da direita para esquerda e de baixo para cima (DEBC). Para cada varredura i , a imagem original é usada como entrada e obtemos uma imagem de saída $P_i(x, y)$. Para cada pixel preto da máscara (saída de Niblack dilatado), identificado como letra, calcula-se a média da intensidade dos pixels vizinhos na imagem original cujo correspondente na máscara seja branco, identificado como fundo, e atribui essa média ao pixel central em questão. Com isso, tenta-se substituir o valor daquele pixel, considerado letra, por um valor de fundo influenciado por seus vizinhos considerados fundo, apagando-o da imagem.

Na 5ª varredura, independente de que sentido se escolha, para cada posição da imagem é escolhido o mínimo dos $P_i(x, y)$, calculados anteriormente, como saída final do *inpainting* (*background*).

O pseudo-código que implementa esse algoritmo é mostrado a seguir:

$I(x, y)$: Imagem original

$M(x, y)$: Máscara de *inpainting*

$P_i(x, y)$: Resultado do *inpainting* da varredura i

$BG(x, y)$: Imagem final da estimativa do fundo

I_x, I_y : Largura e altura da imagem

$x_{começo}[4] = 0, 0, I_x, I_x$; $x_{fim}[4] = I_x, I_x, 0, 0$

$y_{começo}[4] = 0, I_y, 0, I_y$; $y_{fim}[4] = I_y, 0, I_y, 0$

para $i = 1 \rightarrow 4$, **faça**

$M = IM$

para $y = y_{começo}[i] \rightarrow y_{fim}[i]$, **faça**

para $x = x_{começo}[i] \rightarrow x_{fim}[i]$, **faça**

se $M(x, y) = 0$, **então**

$P_i(x, y) = \text{média}(I(x - 1, y) \cdot M(x - 1, y), I(x, y - 1) \cdot M(x, y - 1), I(x + 1, y) \cdot M(x + 1, y), I(x, y + 1) \cdot M(x, y + 1))$

$I(x, y) = P_i(x, y)$

$M(x, y) = 1$

fim

fim

fim

fim

para $y = y_{começo}[1] \rightarrow y_{fim}[1]$, **faça**

para $x = x_{começo}[1] \rightarrow x_{fim}[1]$, **faça**

$BG(x, y) = \min(P_i(x, y)), i = 1, \dots, 4$

fim

fim

retorne $BG(x, y)$

A escolha do mínimo valor dentre os valores de *inpainting* das 4 varreduras objetiva obter um melhor desempenho nas etapas posteriores de normalização e conseqüentemente de binarização global, baseando-se em resultados experimentais. Dessa forma, consegue-se obter uma menor interferência dos ruídos de fundo quando executarmos a binarização global de Otsu mais a frente, assim como consegue-se uma melhor separação do conteúdo textual e ruídos de bordas e manchas ao aplicar a binarização global. Resultados da aplicação da técnica de *inpainting* na estimativa de fundo da imagem são mostrados na Figura 5.



Figura 5 - Resultados do processo de *inpainting*: Imagens originais em (a), (c) e (f) e suas respectivas estimativas de fundo em (b), (d) e (g)

3.2 - Normalização da Imagem

Após a estimativa de fundo, é aplicado ao resultado deste o procedimento de normalização da imagem para balancear a iluminação do fundo, na tentativa de sanar os problemas trazidos por manchas, sombras e iluminação não-uniforme no resultado da binarização global de Otsu. A aplicação deste método ao resultado da estimativa de fundo transforma a imagem para que ela passe a ter uma distribuição bimodal bem definida, obtendo um bom resultado ao ser processado por um método de binarização global. Na prática, os pixels do fundo passam a ter tons mais semelhantes, o que permite ao processo de binarização global separar melhor os pixels do fundo dos pixels das letras, já que os pixels que antes eram de manchas e sombras agora não destoam tanto de pixels do fundo com tons mais uniformes, sendo assim também considerado fundo e não letra.

O método de normalização considera como imagem normalizada $F(x, y)$ a razão $\frac{I(x,y)}{BG(x,y)}$, da imagem original com a estimativa de fundo. Além disso, $F(x, y)$ é também normalizada para o intervalo entre os valores mínimo e o máximo da imagem original, utilizando a seguinte equação:

$$N(x, y) = [(I_{máx} - I_{mín}) * \frac{F(x, y) - F_{mín}}{(F_{máx} - F_{mín})} + I_{mín}]$$

onde $F(x, y) = \frac{I(x,y)+1}{BG(x,y)+1}$, a fim de que seja teoricamente correto e elimine a possibilidade de uma divisão por zero. $I_{mín}$, $I_{máx}$, $F_{mín}$ e $F_{máx}$ indicam os mínimos e máximos valores de $I(x, y)$ e $F(x, y)$ respectivamente. Exemplos de resultados da normalização são mostrados na Figura 6.

to give you a fair hearing and he wants you
not to make the changes suggested. Will
was proposing to change the wording a little
and try it again. Now he says that he

(a)

to give you a fair hearing and he wants you
not to make the changes suggested. Will
was proposing to change the wording a little
and try it again. Now he says that he

(b)

turn me of
the Star Saw
I buy much
y for the same
tas appear
had to command
the same you

(c)

turn me of
the Star Saw
I buy much
y for the same
tas appear
had to command
the same you

(d)

Figura 6 - Resultado da técnica de normalização: (a) e (c) são as imagens originais e (b) e (d) suas respectivas imagens normalizadas.

3.3 - Binarização global e pós-processamento

Após a imagem da estimativa de fundo ser normalizada, pode-se aplicar o método de binarização global de forma eficiente, obtendo-se resultados satisfatórios. Entretanto, para minimizar a contribuição de pequenos componentes ruidosos (pixels de fundo que falsamente foram identificados como letra) na computação das etapas a seguir, e diminuir as falsas combinações quando os resultados de Otsu e Niblack forem combinados, deve-se executar um pós-processamento para remover esses ruídos remanescentes do método de Otsu. Na proposta do artigo, o ruído é considerado como componentes conectados de altura menor que um valor h . Componente conectado de uma imagem é o conjunto de pixels da imagem que estão conectados entre si, através de uma 4-conectividade ou 8-conectividade (Figura 7), ou seja, dado um pixel p do componente conectado, existe pelo menos um caminho de p a todos os outros pixels do componente conectado [17]. Um componente conectado possui altura igual à distância do pixel mais ao topo da imagem (na linha $y_{\text{mín}}$) ao pixel mais baixo da imagem (na linha $y_{\text{máx}}$), ou seja, $y_{\text{máx}} - y_{\text{mín}}$.

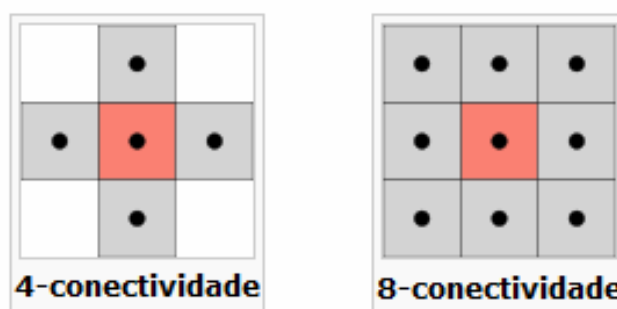


Figura 7 - Conectividade entre os pixels

Como dito antes, o objetivo dessa etapa é eliminar os componentes conectados que possuam tamanho menor que um valor h , calculado como o tamanho mínimo de um componente conectado que satisfaz a equação:

$$\sum_{j=1}^h \frac{RP^j}{RC^j} > 1$$

onde:

$$RP^j = \frac{\sum_{x=1}^{I_x} \sum_{y=1}^{I_y} O^j(x,y)}{3 \sum_{x=1}^{I_x} \sum_{y=1}^{I_y} O(x,y)}$$

é definido como a razão entre o número de pixels identificados como letra por Otsu e que pertencem a um componente conectado de altura j e o número total de pixels identificados como letra por Otsu.

$$RC^j = \frac{NC^j}{NC_o}$$

é definido como a razão entre o número NC^j de componentes conectados de altura j e o número total NC_o de componentes conectados de Otsu.

Assim, para achar o valor de h , iniciamos com valor $h = 1$ e o incrementamos até que o valor de RP^j seja maior que o de RC^j . Por fim, este valor de h é usado como limiar na eliminação de componentes ruidosos, permanecendo na imagem resultado do pós-processamento $OP(x, y)$ apenas os componentes conectados de altura maior ou igual a h , como critério mostrado na equação a seguir:

$$OP(x, y) = \bigcup_{i=1}^{NC_o} O_i(x, y), \forall i : H(i) \geq h,$$

onde $H(i)$ denota a altura do i -ésimo componente conectado $O_i(x, y)$ do resultado de Otsu. Exemplos do resultado de Otsu acompanhado do resultado do pós-processamento $OP(x, y)$ são apresentados na Figura 8.

second vacancy, will have no voice in discord with
the general applause.
The Salary of public ministers is not in the service of private
or Public Interest, but in the union of both.
The Affirmation of laws legal the prescriptions of
the Fruminate

(a)

to give you a fair hearing and he wants you
not to make the changes suggested. Will
was proposing to change the wording a little
and try it again. Now he says that he

(d)

second vacancy, will have no voice in discord with
the general applause.
The Salary of public ministers is not in the service of private
or Public Interest, but in the union of both.
The Affirmation of laws legal the prescriptions of
the Fruminate

(b)

to give you a fair hearing and he wants you
not to make the changes suggested. Will
was proposing to change the wording a little
and try it again. Now he says that he

(e)

second vacancy, will have no voice in discord with
the general applause.
The Salary of public ministers is not in the service of private
or Public Interest, but in the union of both.
The Affirmation of laws legal the prescriptions of
the Fruminate

(c)

to give you a fair hearing and he wants you
not to make the changes suggested. Will
was proposing to change the wording a little
and try it again. Now he says that he

(f)

Figura 8 - (a) e (d) imagens normalizadas; (b) e (e) resultado da binarização global de Otsu em (a) e (d) respectivamente; (c) e (f) pós-processamento para retirar ruídos de altura menor que 5 e 8 em (b) e (e) respectivamente

3.4 - Estimativa da largura de traço e cálculo do contraste para binarização local da imagem normalizada

Para a execução do algoritmo de binarização local de Niblack, o artigo propõe que sejam escolhidos os parâmetros de tamanho da janela w e o peso k baseados na estimativa da largura de traço SW e no cálculo do contraste C .

A estimativa da largura de traço (*stroke width*) busca estabelecer um valor médio da largura do traçado da letra em uma imagem de documento. Essa informação é utilizada para definir o valor de tamanho da janela w no método de Niblack. Quanto maior a largura de traço do documento, maior deve ser a janela que enquadrará os pixels desse traço. Para se calcular a largura de traço, foi utilizado o método descrito em [18], que usa o esqueleto do traço $S(x,y)$ e os pontos de contorno do traço. O esqueleto é obtido através da técnica de esqueletização de Lee e Chen [19] que reduz o conteúdo textual obtido da binarização de Otsu pós-processado à um traçado fino em formato de esqueleto que preserva a extensão e conectividade do conteúdo original. Os pontos de contorno, pixels que revestem as letras, são obtidos considerando todos os pixels *não-letra* da 8-vizinhança dos pixels mais externos das letras. De acordo com o método de Ntirogiannis *et al.*, para cada ponto pertencente ao esqueleto de um componente conectado j , é computado a largura de traço $S_j^{SW} = 2D + 1$, onde D é definido como a menor distância desse ponto a um ponto de contorno. Para cada componente conectado, considera-se então a maior largura de traço calculado anteriormente para cada ponto daquele componente conectado, $SW_j = \text{Max}(S_j^{SW})$. A estimativa da largura de traço final SW é calculada como a média da largura de traço de todos os componentes conectados.

O cálculo do contraste da imagem é determinado pela diferença entre a intensidade do *background* e do *foreground* [20]. O esqueleto $S(x, y)$ é também utilizado no método de Gatos *et al.* para o cálculo

do valor da média e do desvio padrão do *foreground*. Esses servirão como parâmetro para a realização do cálculo do contraste. O esqueleto é utilizado aqui, no lugar de todo o resultado da binarização, inclusive de pixels próximos a pixels das letras, pois pixels do esqueleto têm muito mais chance de realmente serem letra, evitando que pixels próximos do contorno do resultado da binarização, que podem ser pixels de transição ou até mesmo do fundo, tendencie a computação do cálculo do contraste. Calcula-se então a intensidade média $FG_{média}$ e desvio padrão FG_{std} dos pixels da imagem original que tem correspondente no esqueleto $S(x, y)$. Para o cálculo da média $BG'_{média}$ e desvio padrão BG'_{std} dos pixels do fundo, também necessários para o cálculo do contraste, é utilizado uma imagem de estimativa de fundo modificada, $BG'(x, y) = Média(P_i(x, y))$, onde os $P_i(x, y)$ são os resultados das varreduras calculadas anteriormente na etapa de estimativa de fundo. No método de Gatos *et al.*, o cálculo do contraste da imagem é realizada através do logaritmo da taxa de contraste, modificado para obter um melhor resultado quando aplicado em imagens de documentos históricos que apresentam degradações. O cálculo segue de acordo com a equação:

$$C = -50 \cdot \log_{10} \left(\frac{FG_{média} + FG_{std}}{BG'_{média} - BG'_{std}} \right)$$

O valor da intensidade média $FG_{média}$ dos pixels das letras é acrescido do desvio padrão FG_{std} no intuito de se aproximar da intensidade dos caracteres esvaecidos, enquanto que a intensidade média $BG'_{média}$ dos pixels do fundo é subtraído do desvio padrão BG'_{std} para se aproximar de variações de intensidade no fundo, tais como sombras e manchas. O valor constante de -50 é usado apenas para limitar o intervalo de valores finais entre 0 e 100: desconsiderando o caso em que todos os pixels correspondentes a letra tem intensidade 0 e todos os pixels de fundo tem intensidade 255, o que levaria ao logaritmo de zero inexistente, para o caso extremo em que o numerador é igual a 2,5 e o denominador é igual a 252,5, o logaritmo da razão é igual a 2 e o contraste C é igual a 100

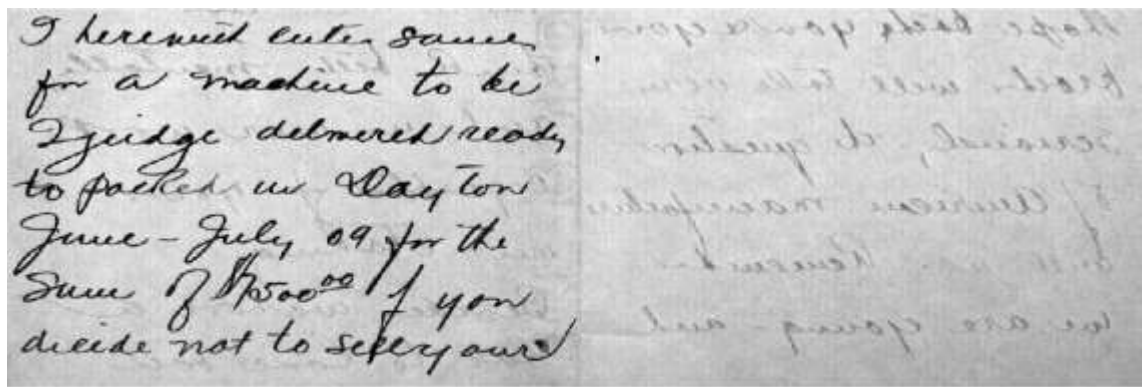
nesse caso. Entretanto, quanto mais próximos forem a intensidade do fundo e das letras, menor é o contraste C , pois $\lim_{x \rightarrow 1} \log_{10} x = 0$.

O algoritmo de binarização local de Niblack é então aplicado à imagem normalizada $N(x, y)$, obtida anteriormente e que já possui intensidade de fundo uniforme, utilizando os parâmetros de tamanho da janela w e peso k , de acordo com os valores encontrados de largura de traço SW e contraste C , como mostrado nas equações:

$$w = 2.SW$$

$$k = -0,2 - 0,1 \cdot \left\lfloor \left(\frac{C}{10} \right) \right\rfloor$$

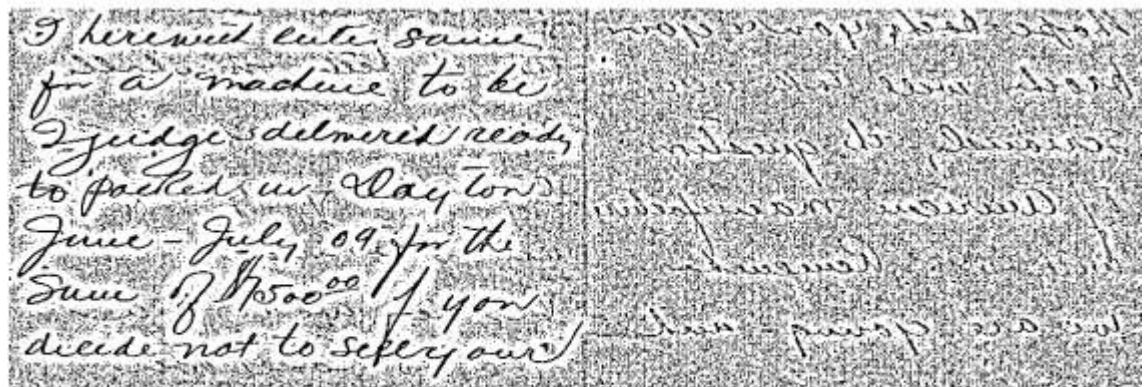
O parâmetro k , que por *default* é $-0,2$, é decrescido quanto maior for valor do contraste, para melhor detecção de escritas esvaecidas. Quando C tiver valor máximo de 100 , k tem valor igual a $-1,2$. O constante tamanho da janela do método de Niblack é apropriado para documentos manuscritos que contenham pouca variação de tamanho da largura de traço. Exemplo do resultado do método de binarização de Niblack utilizando parâmetros baseados em largura de traço e contraste como descrito anteriormente são mostrados na Figura 9.



(a)



(b)



(c)

Figura 9 - Resultados do método de Niblack: (a) Imagem original; (b) Niblack com $w=60$ e $k=-0,2$; (c) Niblack com $w=2.SW$ e $k(C)=-0,5$ usando $SW=14$ e $C=38,36$

3.5 - Combinação dos resultados de binarização global e local

O objetivo da combinação proposta no artigo é aproveitar os benefícios encontrados em ambos os tipos de binarização. Por um lado, o resultado $OP(x, y)$ da binarização global de Otsu seguido de pós-processamento possui baixo ruído de fundo porém falha em detectar partes esvanecidas dos caracteres. Do outro lado, o algoritmo de binarização local de Niblack, quando aplicado à imagem normalizada e utilizando parâmetros adequados, resulta numa imagem $NB_j(x, y)$ com muito ruído de fundo, porém é relativamente eficiente em detectar caracteres esvanecidos. Combinando os resultados de ambas as técnicas, busca-se alcançar um resultado com baixo ruído e eficiente em recuperar caracteres esvaecidos.

De acordo com o método de combinação proposto no artigo de Gatos *et al.*, o resultado $CO(x, y)$ mantém todos componentes conectados de Niblack que satisfazem a equação:

$$CO(x, y) = \bigcup_{j=1}^n NB_j(x, y), \forall j: d(j) = \textit{verdadeiro}$$

onde

$$d(j) = \begin{cases} \textit{verdadeiro}, & \text{se } 100 \cdot \frac{\sum_{x=1}^{I_x} \sum_{y=1}^{I_y} OP(x, y) \cdot NB_j(x, y)}{\sum_{x=1}^{I_x} \sum_{y=1}^{I_y} NB_j(x, y)} \geq C, \\ \textit{falso}, & \text{caso contrário} \end{cases}$$

e C é o valor do contraste encontrado anteriormente.

Dessa forma, só são mantidos os componentes conectados que possuem a maior parte dos seus pixels comuns aos pixels de letra correspondente em Otsu pós-processado, eliminando aqueles componentes que possuem taxa de combinação inferior à $\frac{C}{100}$ e

evitando assim grandes componentes conectados de Niblack que coincidem em apenas alguns pontos com Otsu. O contraste C é usado aqui como limiar por ter uma ligação próxima com o nível de ruído de fundo que permanece após a binarização de Otsu, provenientes de imagens que possuem *bleed-through* e, conseqüentemente, mais alto contraste. Quanto maior o contraste, mais ruídos permanecerão, mesmo após aplicar as etapas de estimativa de fundo e normalização, e mais criterioso deverá ser o algoritmo de combinação quanto aos componentes conectados de Niblack.

Por fim, o resultado $CO(x,y)$ é aperfeiçoado afim de sanar alguns componentes conectados deficientes retidos no processo anterior, devido a etapa de normalização da imagem e tamanho constante. O procedimento inclui todos os pixels do resultado de Otsu $O(x,y)$ antes de ser pós-processado que estejam na vizinhança 8-conectada dos pixels de $CO(x, y)$. A equação a seguir mostra como é realizado esse procedimento.

$$FB(x, y) = CO(x, y) \cup (O(x, y) * f(x, y))$$

onde

$$f(x, y) = \begin{cases} 1, se \sum_{m=x-1}^{x+1} \sum_{n=y-1}^{y+1} CO(m, n) > 0 \\ 0, caso contrário \end{cases}$$

A imagem $FB(x,y)$ é então o resultado final do método de binarização proposto por Gatos *et al.* O resultado dessa etapa final é mostrado na figura 10.

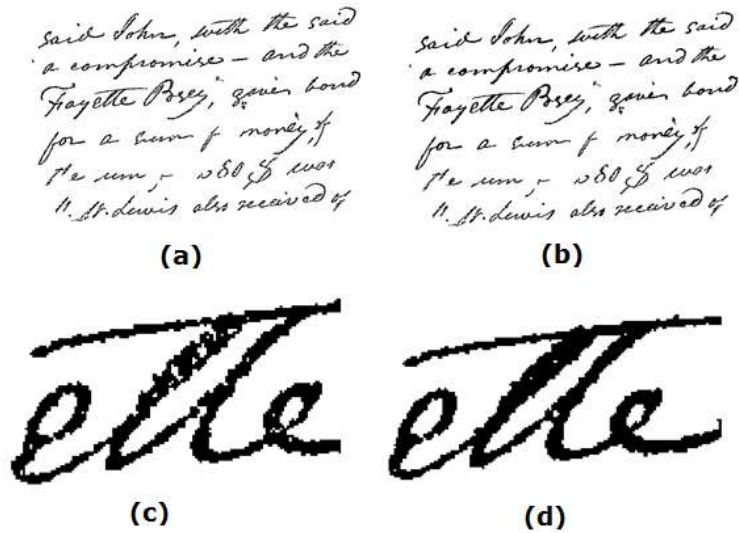


Figura 10 – Melhoramento da combinação: (a) e (c) imagem antes do melhoramento e zoom em (a) respectivamente; (b) e (d) imagem depois do melhoramento e zoom em (b) respectivamente.

Os resultados da binarização de Gatos *et al.* para cada etapa são mostrados nas figuras 11 e 12 a seguir.

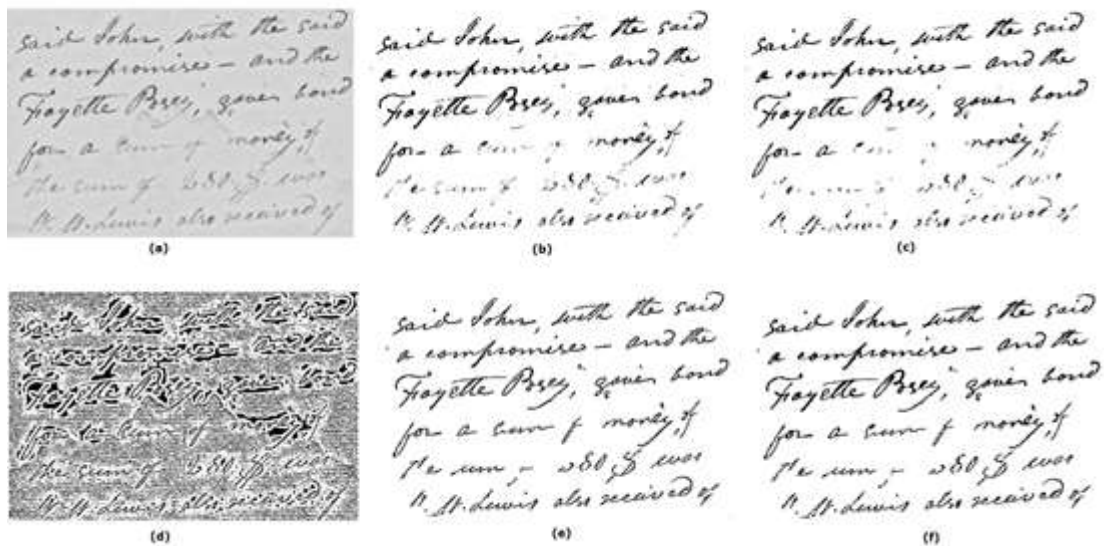


Figura 11 - Binarização da figura 5a: (a) Imagem normalizada da 5a; (b) Otsu de (a); (c) Pós-processamento de (b); (d) Niblack de (a) usando SW=7 e C=8,17; (e) Combinação de (c) com (d); (f) Pós-processamento de (e) e resultado final

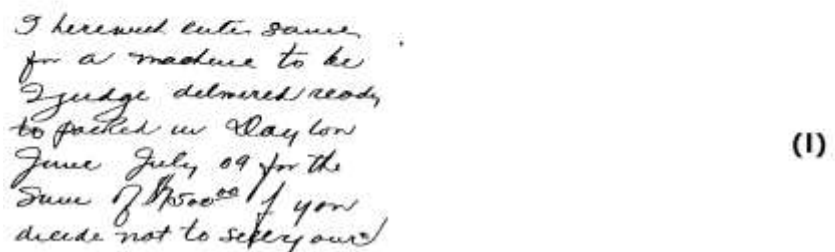
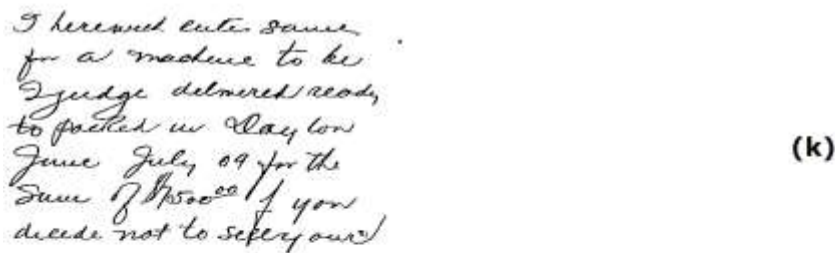
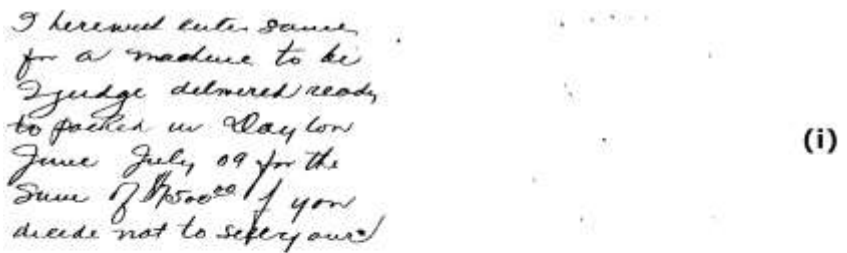
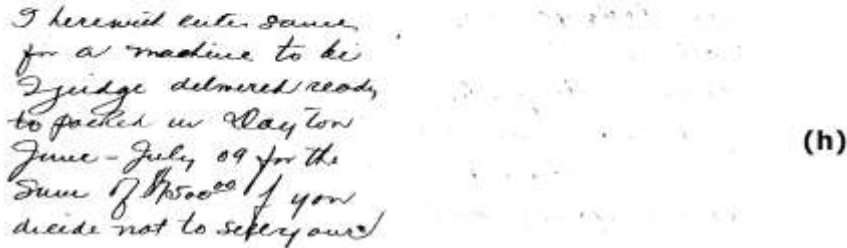
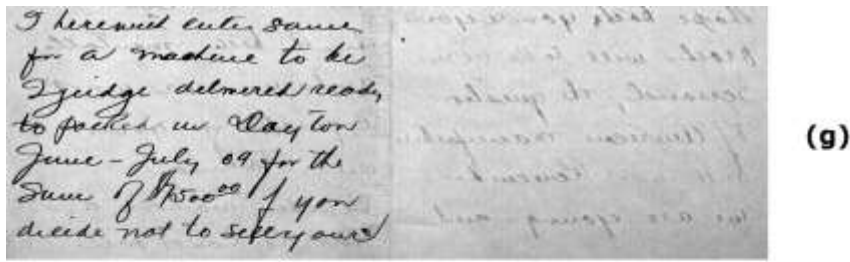


Figura 12 - - Binarização da figura 9a: (a) Imagem normalizada da 9a; (b) Otsu de (a); (c) Pós-processamento de (b); (d) Niblack de (a) usando SW=14 e C=39.36; (e) Combinação de (c) com (d); (f) Pós-processamento de (e) e resultado final

Capítulo 4 - Resultados experimentais

O algoritmo proposto por Gatos *et al.* e analisado neste trabalho foi aplicado em imagens de documentos históricos do *dataset* da DIBCO'13. Essas imagens possuem representações dos potenciais problemas que enfrenta o processo de binarização, sendo bastante útil para nossos experimentos aqui por abranger quase todo tipo de situação. Elas possuem dimensões e formatos diversos.

Os experimentos realizados são baseados na comparação entre o resultado do algoritmo e o resultado "ótimo", chamado *Ground Truth (GT)*, que é uma imagem binária ideal (perfeita separação das letras e do fundo) extraída da imagem original em tons de cinza. Essa comparação verifica os pixels que foram considerados letra e fundo do resultado da binarização verificando se coincidem no pixel correspondente no *Ground Truth* (Figura 13).

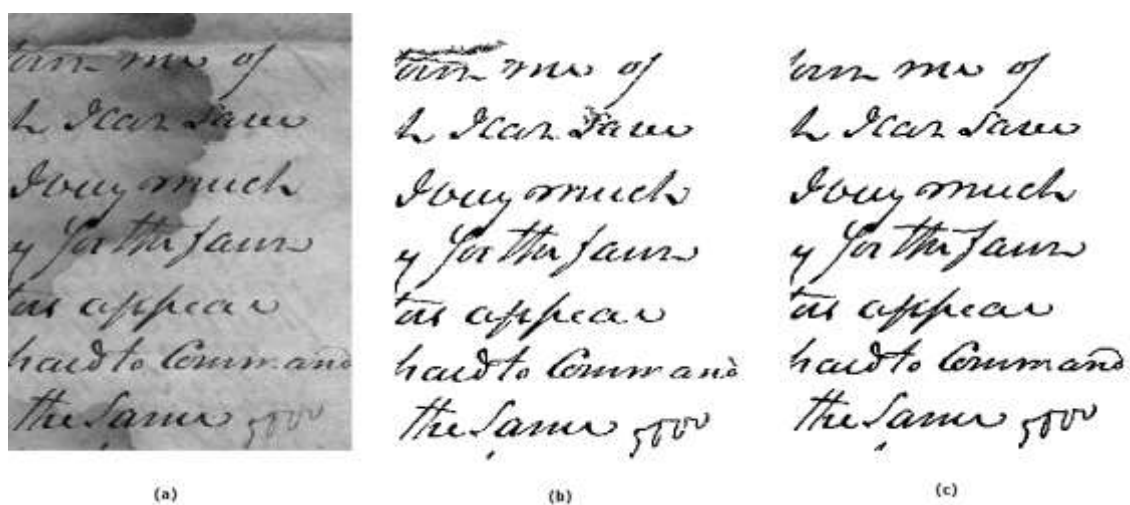


Figura 13 - (a) imagem original; (b) Resultado do método proposto; (c) *Ground Truth*

Para a análise apresentada neste capítulo, algumas das medidas que são apresentadas a seguir utilizam alguns conceitos:

True Positive (TP): taxa de pixels reconhecidos corretamente como texto

True Negative (TN): taxa de pixels reconhecidos corretamente como fundo

False Positive (FP): taxa de pixels reconhecidos erroneamente como texto

False Negative (FN): taxa de pixel reconhecidos erroneamente como fundo

Recall (ou *Sensitivity*): Representa a taxa de pixels corretamente detectados na imagem pelo total de pixels positivos, dado pela fórmula:

$$Recall = \frac{TP}{TP + FN} = \frac{TP}{TOTAL DE PIXELS LETRA}$$

Precision: Representa a taxa de acerto do texto presente na imagem binarizada, indicando a qualidade do resultado. Pode ser expressa pela fórmula:

$$Precision = \frac{TP}{TP + FP} = \frac{TP}{TOTAL DE PIXELS LETRA}$$

Um pixel reconhecido corretamente como texto é aquele que tem seu valor binarizado igual a 0, tanto no resultado da binarização quanto no *Ground Truth*. Assim também, um caso de *true negative* é aquele em que o pixel tem valor 1, tanto no resultado da binarização quanto no *Ground Truth*.

4.1 - Medidas de avaliação

Para realizar os estudos de desempenho do método proposto, algumas medidas foram usadas a fim de comparar os seus valores para indicar o quão bom foi o resultado de sua binarização. Existem vários tipos de medidas, porém adotamos aqui as mesmas usadas na DIBCO'13 [21].

4.1.1 - *F-Measure (FM) ou F-Score*

Representa a média harmônica das medidas *Precision* e *Recall*, e pode ser calculada da seguinte forma:

$$F - Measure = \frac{2.Recall.Precision}{Recall+Precision}$$

quantizado no intervalo entre 0 e 1.

Quanto maior seu valor, mais próximo do *Ground Truth* é a imagem binarizada.

4.1.2 – *Pseudo F-Measure (p-FM)*

Calculado da mesma forma que FM, porém utiliza o esqueleto do GT para calcular o *Pseudo Recall*, no lugar do *Recall*. Assim como F-Measure, quanto maior seu valor, melhor o desempenho da binarização.

4.1.3 – *Peak Signal to Noise Ratio (PSNR)*

A razão entre sinal de pico e ruído é calculada tomando como base o erro médio quadrático (MSE – *Mean Square Error*):

$$MSE = \frac{\sum_{x=1}^M \sum_{y=1}^N (I(x,y) - I'(x,y))^2}{M.N},$$

onde I é o resultado da binarização, I' é o GT e M e N são as dimensões da imagem. Assim, o PSNR é calculado como:

$$PSNR = 10 * \log \left(\frac{Cr^2}{MSE} \right)$$

sendo Cr a diferença entre a cor máxima e a mínima presentes na imagem.

Quanto maior seu valor, mais parecidas são as imagens do resultado da binarização com o GT.

4.1.4 – Negative Rate Metric (NRM)

Média entre a taxa de falso negativo NR_{FN} e a taxa de falso positivo NR_{FP}

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

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

Quanto menor seu valor, melhor o resultado da binarização.

4.1.5 – Misclassification Penalty Metric (MPM)

Define uma medida de penalidade para aquele pixel que está distante da borda do objeto do GT

$$MP_{FN} = \frac{\sum_{i=1}^{FN} d_{FN}^i}{D} \quad MP_{FP} = \frac{\sum_{j=1}^{FP} d_{FP}^j}{D},$$

onde $\sum_{i=1}^{FN} d_{FN}^i$ é a soma das distâncias dos pixels falso negativos até a borda mais próxima, $\sum_{j=1}^{FP} d_{FP}^j$ das distâncias dos pixels falso positivos até a borda mais próxima e D é a soma de todas as distâncias de todos os pixels. MPM é então definido como:

$$MPM = \frac{MP_{FN} + MP_{FP}}{2}$$

Quanto menor essa medida, melhor o método de binarização identifica as bordas do texto e maior a precisão do resultado da binarização.

4.1.6 – *Distance Reciprocal Distortion (DRD)*

Essa medida é usada para avaliar a distorção visual em imagens de documentos binarizados, tendo relação com a percepção visual do ser humano. Ela mede a distorção para todos os S pixels invertidos da seguinte forma:

$$DRD = \frac{\sum_{k=1}^S DRD_k}{NUBN},$$

onde $NUBN$ é o número de blocos 8x8 não-uniformes na imagem GT e DRD_k é a distorção k-ésimo pixel invertido definida como a soma dos pixels num bloco 5x5 de GT que diferem do pixel central invertido k na imagem do resultado da binarização e dada pela equação:

$$DRD_k = \sum_{i=-2}^2 \sum_{j=-2}^2 |GT_k(i, j) - B_k(x, y)| * W_{Nm}(i, j)$$

Assim como MPM e NRM, quanto menor o seu valor, melhor o desempenho da binarização.

4.2 - Análise comparativa

O resultado do algoritmo de Gatos *et al.* foi comparado com o resultado do método de binarização que obteve 1º (método de Su-Lu-Tan), 2º (método de Howe), 3º (método de R.F. Moghaddam *et al.*), 4º (método de Lelore e Bouchara) e 5º (método de M.A. Ramírez-Ortegón *et al.*) lugar do *ranking* final da DIBCO'13, de acordo com as medidas estabelecidas anteriormente. Esses resultados estão apresentados nas Tabelas 1 e 2.

A Tabela 1 mostra o somatório da posição no *ranking* parcial para cada imagem e cada método, de acordo com as medidas FM, F_{PS} , PSNR e DRD, utilizados na DIBCO'13. Por exemplo, o 1º colocado da DIBCO'13 obteve melhor pontuação nas medidas FM e F_{PS} , recebendo 1 ponto para cada medida, e segunda melhor pontuação nas medidas PSNR e DRD, recebendo 2 pontos para cada uma dessas medidas, totalizando 6 pontos para a imagem HW1 (1ª imagem de documento manuscrito da DIBCO'13). A Tabela 1 mostra também a média de cada medida para cada método. Na Tabela 2, a pontuação final é a soma das pontuações de cada imagem, mostrada na Tabela 1, para cada método. O *ranking* final é baseado então nessas pontuações em ordem crescente. Quanto menor a pontuação, melhor a posição do método no *ranking* final.

As Figuras 12 até 19 apresentam alguns exemplos de aplicação do método. As figuras mostram a imagem original (da DIBCO'13) no topo, seguido do respectivo *Ground Truth*, os resultados do 1º e 2º lugar da DIBCO 2013 (método de Su-Lu-Tan e método de Howe respectivamente), comparando-os com o resultado do método de Gatos *et al.*

	Imagens de documentos manuscritos da DIBCO'13							
	HW1	HW2	HW3	HW4	HW5	HW6	HW7	HW8
1º lugar da DIBCO'13	6	8	6	6	9	6	8	8
2º lugar da DIBCO'13	8	6	10	8	6	9	6	10
3º lugar da DIBCO'13	13	17	10	15	12	15	15	12
4º lugar da DIBCO'13	18	20	19	19	17	19	22	19
5º lugar da DIBCO'13	22	23	24	24	23	23	20	21
Gatos et al.	17	10	15	12	17	12	13	14

Tabela 1 - Pontuação de cada método para cada imagem

Rank	Método	Pontuação	FM	F _{PS}	PSNR	DRD
1º	1º lugar da DIBCO'13	57	92,70	93,19	21,29	3,18
2º	2º lugar da DIBCO'13	63	92,12	94,19	20,68	3,10
3º	3º lugar da DIBCO'13	109	91,81	92,67	20,68	4,02
4º	Gatos et al.	110	91,76	92,54	20,17	3,22
5º	4º lugar da DIBCO'13	153	91,69	92,16	20,54	3,59
6º	5º lugar da DIBCO'13	180	90,92	92,82	19,32	3,91

Tabela 2 - Ranking baseado no somatório da pontuação pra cada imagem.

plays upon equal terms plays to a disadvantage. A man loses more than he would have gained. Suppose the stake £20. on each side. If he loses, he loses £20. If he wins he wins £20. and no more. Now £20. lost is more on the side of pain, than £20. won can be on the side of pleasure. A man can better bear

(a)

plays upon equal terms plays to a disadvantage. A man loses more than he would have gained. Suppose the stake £20. on each side. If he loses, he loses £20. If he wins he wins £20. and no more. Now £20. lost is more on the side of pain, than £20. won can be on the side of pleasure. A man can better bear

(b)

plays upon equal terms plays to a disadvantage. A man loses more than he would have gained. Suppose the stake £20. on each side. If he loses, he loses £20. If he wins he wins £20. and no more. Now £20. lost is more on the side of pain, than £20. won can be on the side of pleasure. A man can better bear

(c)

plays upon equal terms plays to a disadvantage. A man loses more than he would have gained. Suppose the stake £20. on each side. If he loses, he loses £20. If he wins he wins £20. and no more. Now £20. lost is more on the side of pain, than £20. won can be on the side of pleasure. A man can better bear

(d)

plays upon equal terms plays to a disadvantage. A man loses more than he would have gained. Suppose the stake £20. on each side. If he loses, he loses £20. If he wins he wins £20. and no more. Now £20. lost is more on the side of pain, than £20. won can be on the side of pleasure. A man can better bear

(e)

Figura 14 - HW1 da DIBCO'13: (a) Imagem original; (b) GT de (a); (c) 1º lugar do ranking da DIBCO'13; (d) 2º lugar do ranking da DIBCO'13; (e) Gatos et al.

second vacancy, will have no voices in discord with
the general applause.
The Salus populi consists not in the sacrificing of private
or Public Interest, but in the union of both.
The Assassination of Caesar beget the prescriptions of
the Triumvirate

(a)

second vacancy, will have no voices in discord with—
the general applause.

The Salus populi consists not in the sacrificing of private
or Public Interest, but in the union of both.

(b)

The Assassination of Caesar beget the prescriptions of
the Triumvirate

second vacancy, will have no voices in discord with—
the general applause.

The Salus populi consists not in the sacrificing of private
or Public Interest, but in the union of both.

(c)

The Assassination of Caesar beget the prescriptions of
the Triumvirate

second vacancy, will have no voices in discord with—
the general applause.

The Salus populi consists not in the sacrificing of private
or Public Interest, but in the union of both.

(d)

The Assassination of Caesar beget the prescriptions of
the Triumvirate

second vacancy, will have no voices in discord with
the general applause

The Salus populi consists not in the sacrificing of private
or Public Interest, but in the union of both.

(e)

The Assassination of Caesar beget the prescriptions of
the Triumvirate

Figura 15 - HW2 da DIBCO'13: (a) Imagem original; (b) GT de (a); (c) 1º lugar do ranking da DIBCO'13; (d) 2º lugar do ranking da DIBCO'13; (e) Gatos et al.

Now if there be any meaning in words, all
as a Legislator prescribing or a professor explaining
this must signify, that the moral conduct of a
Law is such a Being as this. - (a)

Now if there be any meaning in words, all
as a Legislator prescribing or a professor explaining
this must signify, that the moral conduct of a
Law is such a Being as this. - (b)

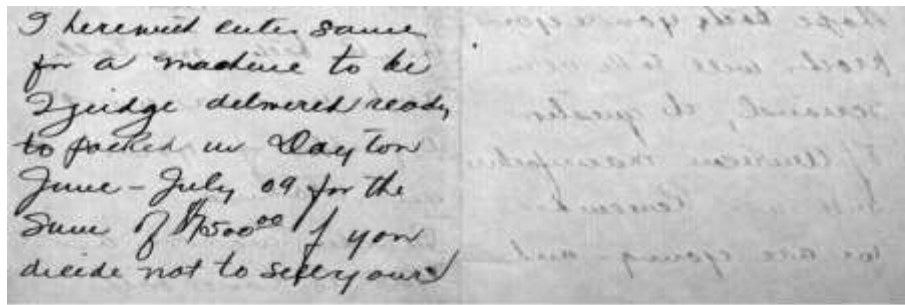
Now if there be any meaning in words, all
as a Legislator prescribing or a professor explaining
this must signify, that the moral conduct of a
Law is such a Being as this. - (c)

Now if there be any meaning in words, all
as a Legislator prescribing or a professor explaining
this must signify, that the moral conduct of a
Law is such a Being as this. - (d)

Now if there be any meaning in words, all
as a Legislator prescribing or a professor explaining
this must signify, that the moral conduct of a
Law is such a Being as this. - (e)

Figura 16 - HW3 da DIBCO'13: (a) Imagem original; (b) GT de (a); (c) 1º lugar do ranking da DIBCO'13; (d) 2º lugar do ranking da DIBCO'13; (e) Gatos et al.

I herewith enter same
for a machine to be
I judge delivered ready
to packed in Clayton
June - July 09 for the
sum of \$1500⁰⁰ if you
decide not to sefer our



(a)

I herewith enter same
for a machine to be
I judge delivered ready
to packed in Clayton
June - July 09 for the
sum of \$1500⁰⁰ if you
decide not to sefer our

(b)

I herewith enter same
for a machine to be
I judge delivered ready
to packed in Clayton
June - July 09 for the
sum of \$1500⁰⁰ if you
decide not to sefer our

(c)

I herewith enter same
for a machine to be
I judge delivered ready
to packed in Clayton
June - July 09 for the
sum of \$1500⁰⁰ if you
decide not to sefer our

(d)

I herewith enter same
for a machine to be
I judge delivered ready
to packed in Clayton
June July 09 for the
sum of \$1500⁰⁰ if you
decide not to sefer our

(e)

Figura 17 - HW4 da DIBCO'13: (a) Imagem original; (b) GT de (a); (c) 1º lugar do ranking da DIBCO'13; (d) 2º lugar do ranking da DIBCO'13; (e) Gatos et al.

Please send copy of this to W.W.
Pastor - Flint & Co. Broad Exchange
For enclosure to W. Wright
Enclosure
W. W. Schawte -
This mms. received while I was
finishing letter to you. Please return it
W. W.

(a)

Please send copy of this to W.W.
Pastor - Flint & Co. Broad Exchange
For enclosure to W. Wright
Enclosure
W. W. Schawte -
This mms. received while I was
finishing letter to you. Please return it
W. W.

(b)

Please send copy of this to W.W.
Pastor - Flint & Co. Broad Exchange
For enclosure to W. Wright
Enclosure
W. W. Schawte -
This mms. received while I was
finishing letter to you. Please return it
W. W.

(c)

Please send copy of this to W.W.
Pastor - Flint & Co. Broad Exchange
For enclosure to W. Wright
Enclosure
W. W. Schawte -
This mms. received while I was
finishing letter to you. Please return it
W. W.

(d)

Please send copy of this to W.W.
Pastor - Flint & Co. Broad Exchange
For enclosure to W. Wright
Enclosure
W. W. Schawte -
This mms. received while I was
finishing letter to you. Please return it
W. W.

(e)

Figura 18 - HW5 da DIBCO'13: (a) Imagem original; (b) GT de (a); (c) 1º lugar do ranking da DIBCO'13; (d) 2º lugar do ranking da DIBCO'13; (e) Gatos et al.

or four days.
My last was quite unkind but I
was impatient then. I thought strange
that I did not get letters from you
but I know it is on account of irreg-
ularity in the mails.
So here is a kiss for my harshness.
Nothing of much importance has oc-
curred since but I write.

(a)

or four days.
My last was quite unkind but I
was impatient then. I thought strange
that I did not get letters from you
but I know it is on account of irreg-
ularity in the mails.
So here is a kiss for my harshness.
Nothing of much importance has oc-
curred since but I write.

(b)

or four days.
My last was quite unkind but I
was impatient then. I thought strange
that I did not get letters from you
but I know it is on account of irreg-
ularity in the mails.
So here is a kiss for my harshness.
Nothing of much importance has oc-
curred since but I write.

(c)

or four days.
My last was quite unkind but I
was impatient then. I thought strange
that I did not get letters from you
but I know it is on account of irreg-
ularity in the mails.
So here is a kiss for my harshness.
Nothing of much importance has oc-
curred since but I write.

(d)

or four days.
My last was quite unkind but I
was impatient then. I thought strange
that I did not get letters from you
but I know it is on account of irreg-
ularity in the mails.
So here is a kiss for my harshness.
Nothing of much importance has oc-
curred since but I write.

(e)

Figura 19 - HW6 da DIBCO'13: (a) Imagem original; (b) GT de (a); (c) 1º lugar do ranking da DIBCO'13; (d) 2º lugar do ranking da DIBCO'13; (e) Gatos et al.



(a)



(b)



(c)



(d)



(e)

Figura 20 - HW7 da DIBCO'13: (a) Imagem original; (b) GT de (a); (c) 1º lugar do ranking da DIBCO'13; (d) 2º lugar do ranking da DIBCO'13; (e) Gatos et al.

that you stand in no need of any
organized boy to sustain you, or
to enlighten your Council, or to
give you information upon which
to form correct opinions of men
& things - I was very apprehensive

(a)

that you stand in no need of any
organized boy to sustain you, or
to enlighten your Council, or to
give you information upon which
to form correct opinions of men
& things - I was very apprehensive

(b)

that you stand in no need of any
organized boy to sustain you, or
to enlighten your Council, or to
give you information upon which
to form correct opinions of men
& things - I was very apprehensive

(c)

that you stand in no need of any
organized boy to sustain you, or
to enlighten your Council, or to
give you information upon which
to form correct opinions of men
& things - I was very apprehensive

(d)

that you stand in no need of any
organized boy to sustain you, or
to enlighten your Council, or to
give you information upon which
to form correct opinions of men
& things - I was very apprehensive

(e)

Figura 21 - HW8 da DIBCO'13: (a) Imagem original; (b) GT de (a); (c) 1º lugar do ranking da DIBCO'13; (d) 2º lugar do ranking da DIBCO'13; (e) Gatos et al.

Capítulo 5 – Conclusão

O método de Gatos *et al.* consegue detectar a maioria dos caracteres esvaecidos, enquanto consegue remover praticamente todo o *bleed-through*, sem deixar muito ruído de fundo ao aplicar a etapa de pós-processamento em Otsu. Consegue lidar bem com as imagens que apresentam manchas e baixo contraste, ao aplicar as técnicas de estimativa de *background* e normalização, fazendo com que a intensidade se torne uniforme em todo o fundo da imagem. Então, este método mostra-se bastante eficiente na aplicação em documentos históricos manuscritos por tratar com eficácia dos problemas que envolvem esse conjunto específico de imagens.

Baseado na Tabela 1, pode-se perceber que algoritmo de Gatos *et al.* alcança alto desempenho comparado aos métodos com melhores desempenhos na DIBCO'13, obtendo a 4ª posição na Tabela 2, porém muito próximo ao 3º colocado, com diferença de apenas 1 ponto. Em algumas imagens, consegue obter um resultado melhor que a maioria dos outros métodos, ao preservar grande parte das letras enfraquecidas, como na Figura 20.

Entretanto, pequenos elementos da imagem pertencentes ao conteúdo textual, como o ponto da letra "i", vírgulas e hifens, são descartados na etapa de pós-processamento por possuir um tamanho pequeno comparado às outras letras do texto, como pode ser visto na Figura 22b. O método também deixa a desejar quando aplicado a imagens que possuem deformidades, como dobras, identificando como letra as áreas escuras devido à deformidade, como mostrado na Figura 22d. Também não consegue lidar bem com variações bruscas de largura de traço, pois utiliza um único valor estimado para toda imagem.

I herewith enter same
for a machine to be
I judge delivered ready
to packed in Clayton
June July 09 for the
sum of \$1500⁰⁰ if you
decide not to sefer any

(a)

I herewith enter same
for a machine to be
I judge delivered ready
to packed in Clayton
June July 09 for the
sum of \$1500⁰⁰ if you
decide not to sefer any

(b)



(c)



(d)

Figura 22 - Problemas no método de Gatos *et al.*: (a) *Ground Truth* da imagem 15(a); (b) resultado da binarização de Gatos *et al.* na imagem 15(a); (c) imagem original com dobradura; (d) resultado da binarização de Gatos *et al.* sobre (c)

Para estudos futuros visando melhorar o desempenho desse método, pode ser interessante a utilização de outras técnicas de estimativa de *background*, como por exemplo baseado em máximos e mínimos locais. A etapa de pós-processamento de Otsu pode considerar tamanho (quantidade de pixels) do componente conectado ao invés de altura, para não perder pequenos elementos textuais da imagem. A etapa em que Niblack utiliza a largura do traço SW para definir a janela fixa pode ser melhorada, avaliando também o desvio padrão das larguras encontradas para cada componente conectado, adaptando melhor a janela fixa às mudanças de largura de traço das letras no texto, ou até mesmo utilizar uma janela variável adaptativa.

Referências

- [1] R.Gonzalez e R.Woods, *Digital Image Processing*, 3a Edição, Prentice Hall, 2007.
- [2] T.Chattopadhyay, V.R.Reddy e U. Garain, "Automatic Selection of Binarization Techniques for Robust OCR", *2013 International Conference on Document Analysis and Recognition (ICDAR'13)*, Washington DC, EUA, 2013.
- [3] N.Otsu, "A thresholding selection method from gray-level histogram", *IEEE Trans. Systems Man Cybernet.* 9 (1), 62–66, 1979.
- [4] W.Niblack, "An Introduction to Digital Image Processing". Prentice-Hall, Englewood Cliffs, NJ, pp. 115–116, 1986.
- [5] I.Pratikakis, B.Gatos e K.Ntirogiannis, "ICDAR 2013 Document Image Binarization Contest (DIBCO 2013)", *International Conference on Document Analysis and Recognition (ICDAR'13)*, Washington DC, EUA, pp. 1471-1476, 2013.
- [6] M.Sezgin e B.Sankur, "Survey over image thresholding techniques and quantitative performance evaluation", *J. Electronic Imaging* 13 (1), pp. 146-168, 2004.
- [7] J.Sauvola e M.Pietikainen, "Adaptive document image binarization". *Pattern Recognition* 33 (2), pp.225–236, 2000.
- [8] N.Howe, "Document binarization with automatic parameter tuning", *International Journal on Document Analysis and Recognition*, vol. 16, pp. 247–258, 2013.
- [9] R.Moghaddam, F.Moghaddam e M.Cheriet, "Unsupervised ensemble of experts (EoE) framework for automatic binarization of document images", *anais do ICDAR 2013*.

[10] T.Lelore e F.Bouchara, "FAIR: a Fast Algorithm for Document Image Restoration", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2013.

[11] M.A.Ramírez-Ortegón, E.Tapia, L.L.Ramírez-Ramírez, R.Rojas e E.Cuevas, "Transition pixel: A concept for binarization based on edge detection and gray-intensity histograms", *Pattern Recognition*, vol. 43, no. 4, pp. 1233–1243, 2010.

[12] M.A.Ramírez-Ortegón, E.Tapia, R.Rojas e E.Cuevas, "Transition thresholds and transition operators for binarization and edge detection", *Pattern Recognition*, vol. 43, no. 10, pp. 3243-3254, 2010.

[13] M.A.Ramírez-Ortegón, E.A.Duñez-Guzmán, R.Rojas e E.Cuevas, "Unsupervised measures for parameter selection of binarization algorithms", *Pattern Recognition*, vol. 44, no. 3, pp. 491–502, 2011.

[14] M.A.Ramírez-Ortegón, V.Märgner, E.Cuevas e R.Rojas, "An optimization for binarization methods by removing binary artifacts", *Pattern Recognition Letters*, vol. 43, no. 11, pp. 1299-1306, 2013.

[15] B.Gatos, K.Ntirogiannis e I.Pratikakis, "A Combined Approach for the Binarization of Handwritten Document Images", *Pattern Recognition Letters - Special Issue on Frontiers in Handwriting Processing*, vol. 35, no. 1, pp. 3-15, 2014.

[16] M.Piccardi, "Background subtraction techniques: a review", *IEEE International Conference on Systems, Man and Cybernetics* 4, pp. 3099–3104, 2014.

[17] M.Dillencourt, H.Samet e M.Tamminen. "A general approach to connected-component labeling for arbitrary image representations". *J. ACM*, 1992.

[18] K.Ntirogiannis, B.Gatos e I.Pratikakis, "A modified adaptive logical level binarization technique for historical document

images”, nos anais do Proc. 10th International Conference on Document Analyze Recognition, pp. 1171–1175, 2009.

[19] H.Lee e B.Chen, “Recognition of handwritten chinese characters via short line segments”, Pattern Recognition 25 (5), pp. 543–552, 1992.

[20] S.Klein, T.Carney, L.Barghout-Stein e C.Tyler, “Seven models of masking”. em Electronic Imaging'97, International Society for Optics and Photonics, pp. 13-24, 1997.

[21] K.Ntirogiannis, B.Gatos e I.Pratikakis, “Performance Evaluation Methodology for Historical Document Image Binarization”, IEEE Transaction on Image Processing, vol. 22, n.2, Fev. 2013