UNIVERSIDADE FEDERAL DE PERNAMBUCO
CENTRO DE TECNOLOGIA E GEOCIÊNCIAS
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA ELÉTRICA

DISSERTAÇÃO DE MESTRADO

PHOTODOC-
UM AMBIENTE PARA PROCESSAMENTO DE IMAGENS DE DOCUMENTOS ADQUIRIDAS POR CÂMERAS DIGITAIS PORTÁTEIS

Gabriel de França Pereira e Silva

UNIVERSIDADE FEDERAL DE PERNAMBUCO
CENTRO DE TECNOLOGIA E GEOCIÊNCIAS
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA ELÉTRICA

PHOTODOC - UM AMBIENTE PARA PROCESSAMENTO DE IMAGENS DE DOCUMENTOS ADQUIRIDAS POR CÂMERAS DIGITAIS PORTÁTEIS

por

Gabriel de França Pereira e Silva

Dissertação submetida ao Programa de Pós-Graduação em Engenharia Elétrica da Universidade Federal de Pernambuco como parte dos requisitos para obtenção do grau de Mestre em Engenharia Elétrica.

Orientador: Prof. Rafael Dueire Lins, Ph.D.

Recife, Agosto de 2009.

S586p **Silva, Gabriel de França Pereira e**

Photodoc – um ambiente para processamento de imagens de documentos adquiridas por câmeras digitais portáteis / Gabriel de França Pereira e Silva. – Recife: O Autor, 2009.

xii, 120 f.; il., gráfs., tabs.

Dissertação (Mestrado) – Universidade Federal de Pernambuco. CTG. Programa de Pós-Graduação em Engenharia Elétrica, 2009.

Inclui Referências Bibliográficas e Apêndices.

1. Engenharia Elétrica. 2. Câmeras Digitais. 3. Análise de Imagens de Documentos. 4. Correção de Distorção. 5. OCR. I. Título.

UFPE

621.3 CDD (22.ed.) **BCTG/2010-044**

Universidade Federal de Pernambuco

Pós-Graduação em Engenharia Elétrica

PARECER DA COMISSÃO EXAMINADORA DE DEFESA DE DISSERTAÇÃO DO MESTRADO ACADÊMICO DE

GABRIEL DE FRANÇA PEREIRA E SILVA

TÍTULO

"PHOTODOC – UM AMBIENTE PARA PROCESSAMENTO DE IMAGENS DE DOCUMENTOS ADQUIRIDAS POR CÂMERAS DIGITAIS PORTÁTEIS"

A comissão examinadora composta pelos professores: RAFAEL DUEIRE LINS, DES/UFPE, VALDEMAR CARDOSO DA ROCHA JÚNIOR, DES/UFPE e GEORGE DARMITON DA CUNHA CAVALCANTI, CIN/UFPE sob a presidência do primeiro, consideram o candidato **GABRIEL DE FRANÇA PEREIRA E SILVA APROVADO**.

Recife, 24 de agosto de 2009.

EDUARDO FONTANA
Coordenador do PPGEE

RAFAEL DUEIRE LINS
Orientador e Membro Titular Interno

GEORGE DARMITON DA CUNHA CAVALCANTI
Membro Titular Externo

VALDEMAR CARDOSO DA ROCHA JÚNIOR
Membro Titular Interno

Agradecimentos

Primeiramente agradeço aos meus pais, Jose e Libânia. Vocês me deram todo o carinho e apoio que um filho poderia pedir, serei eternamente grato nunca esquecendo os exemplos de caráter que vocês são.

A Renata meu grande amor pela paciência e apoio neste percurso.

Ao professor Rafael Dueire Lins, meus sinceros agradecimentos, por ter aceitado ser meu orientador e por todas as lições e ensinamentos passados.

À banca examinadora composta pelos professores Rafael Dueire Lins, George Darmiton da Cunha Cavalcanti e Valdemar Cardoso da Rocha Junior pelas contribuições no desenvolvimento desta dissertação através de suas sugestões.

Aos Professores Hélio Magalhães de Oliveira, Ricardo Menezes Campello de Souza, e Cecílio José Lins Pimentel pelo exemplo de profissionais que são.

Aos meus amigos irmãos que me acompanham desde os tempos do Santa Maria: Carlos, Breno, Rodrigo e Marcio, com os quais já vão mais de 18 anos de alegrias, sou muito grato.

Ao meu irmão Zinaldo, companheiro de projetos, viagens, meu sincero agradecimento.

Aos mestres André Ricardson, João Marcelo e Márcio Lima companheiros de pesquisa, pela ajuda e sugestões.

Agradeço aos meus avós, tios, primos e outros familiares, por sempre desejarem o melhor para mim.

Agradeço aos amigos da pós-graduação que compartilharam as dificuldades nas disciplinas e no desenvolvimento desta dissertação. Em especial: Daniel Oliveira, Daniel Simões, Eduarda Simões, Gilson Jerônimo, Brenno Miro, Roberto Sotero, Giovanna Angelis, Juliano Bandeira, Caio Sousa e Elda Lizandra.

Gabriel de França P. e Silva
Universidade Federal de Pernambuco
21 de agosto de 2009

Resumo da Dissertação apresentada à UFPE como parte dos requisitos necessários para a obtenção do grau de Mestre em Engenharia Elétrica.

PHOTODOC - UM AMBIENTE PARA PROCESSAMENTO DE IMAGENS DE DOCUMENTOS ADQUIRIDAS POR CÂMERAS DIGITAIS PORTÁTEIS

Gabriel de França Pereira e Silva

Agosto/2009

Orientador: Prof. Rafael Dueire Lins, PhD.
Área de Concentração: Telecomunicações/Processamento de Imagens/Engenharia de Documentos.
Palavras-chaves: câmeras digitais, análise de imagens de documentos, correção de distorção, OCR.
Número de páginas: 116

O uso de câmeras digitais portáteis tornou-se uma alternativa viável para a aquisição de imagens de documentos devido a seu baixo custo, portabilidade e evolução. Essa aplicação não originalmente prevista vem tornando-se corriqueira. Tais dispositivos estão atualmente embarcadas em muitos dos equipamentos portáteis, como telefones celulares, palm tops, e outros dispositivos eletrônicos fáceis de utilizar e transportar, usados diariamente por milhares de pessoas. Devido a este novo uso dado às câmeras digitais, são necessários novos algoritmos capazes de processar essas imagens, melhorando-as e assim possibilitando um acesso regular e eficaz da informação nelas contidas. Uma vez que essas imagens podem apresentar problemas, como perda de foco, distorções de perspectiva, distorções causadas pela lente, bem como bordas mais complexas do que as encontradas em documentos escaneados. Esta dissertação analisa alguns dos fatores influentes na qualidade de documentos digitalizados através do uso de câmeras digitais e apresenta o ambiente PhotoDoc. Esse ambiente foi concebido de forma a ser amigável ao usuário, o qual a partir da interface gráfica do PhotoDoc poderá automaticamente: remover bordas, corrigir perspectiva, buscar por imagens de documentos fotografados, realçá-las, binarizá-las e transcrever estas imagens com o auxílio de ferramentas de OCR (Optical character recognition). Escolheu-se desenvolver o ambiente proposto nesta dissertação na forma de plugin do ImageJ, por se tratar de um software que apresenta uma série de outras funcionalidades para o processamento de imagens. Esse conjunto único de características torna o PhotoDoc uma solução pioneira, pois não existe ferramenta acadêmica ou comercial com tal funcionalidades até a presente data.

Abstract of the Dissertation presented to the UFPE as part of necessary requirements for the title of Master in Electrical Engineering.

PHOTODOC - AN ENVIRONMENT FOR PROCESSING DOCUMENT IMAGES ACQUIRED BY PORTABLE DIGITAL CAMERAS

Gabriel de França Pereira e Silva

August/2009

Supervisor: Prof. Rafael Dueire Lins, PhD.
Concentration Area: Telecommunications / Image Processing / Documents Engineering.
Keywords: digital cameras, documents image analysis, distortion correction, OCR.
Number of Pages: 116

The use of portable digital cameras has become a viable alternative to the purchase of document images due to its low cost, portability and evolution. This application is not originally planned to become commonplace. Such devices are now embedded in many mobile devices such as cell phones, palm tops, and other electronic devices easier to use and carry, used daily by thousands of people. Due to this new use given to digital, requiring new algorithms capable of processing these images, improving them and thus allowing a regular and effective access to the information contained therein. Since these images may have problems such as loss of focus, distortions of perspective distortion caused by the lens and the edges are more complex than those found in scanned documents. This thesis examines some of the factors influencing the quality of scanned documents through the use of digital cameras and displays the environment PhotoDoc. This environment was designed to be User-friendly, which from the graphical user interface can automatically PhotoDoc: remove borders, correct perspective, searching for documents images photographed, highlight them, binary code and transcribe them with these images of tools OCR (Optical character recognition). He chose to develop the environment proposed in this dissertation in the form of plugin for ImageJ, because it is a software that shows a number of other features for image processing. These unique characteristics make PhotoDoc a pioneering solution, because there is no academic or business tool with such features to date.

Sumário

Lista de Figuras

Lista de Tabelas

Capítulo 1
Introdução

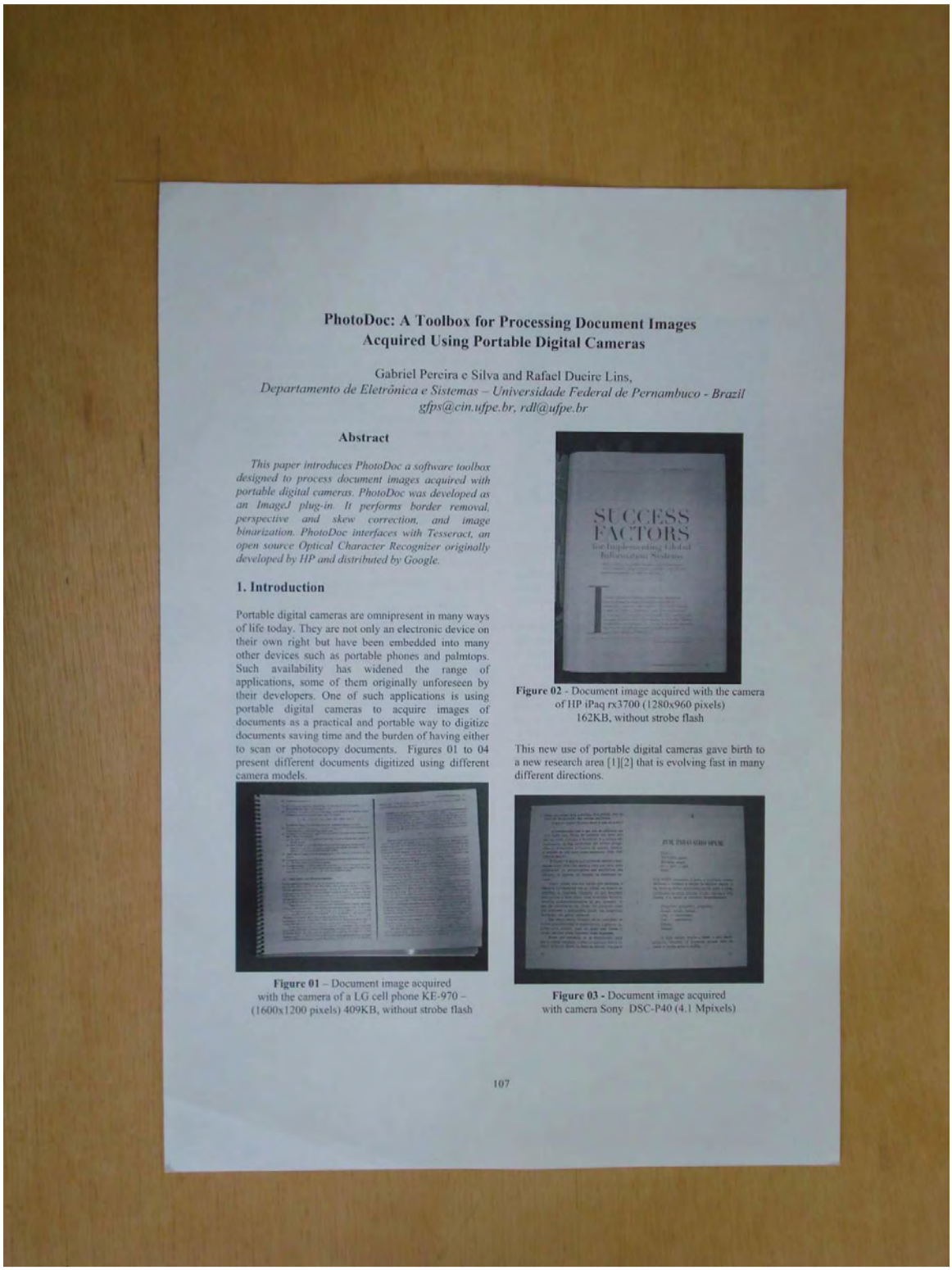
Até a segunda metade do século XX, o papel havia se tornado o principal meio de armazenamento e publicação de conhecimento e informações da humanidade. Apesar da importância do papel, ele é difícil de preservar e armazenar, posto que muitos documentos demandem grande quantidade de espaço físico, além de se degradarem com o tempo. Os motivos pelos quais o papel ainda ocupa uma função essencial deve-se principalmente ao fato de ser barato, portátil, fácil de anotar e, principalmente, de ler [1].

Com o advento e disseminação do computador, surgiram formas mais eficientes de armazenamento, preservação, e facilidades para consultas a documentos legados, através do uso de formato digital. Surgiram então os *scanners*, tornando possível a transposição de documentos manuscritos ou impressos para imagens em formato digital. Documentos digitalizados são facilmente editáveis e indexáveis, possibilitando ainda um baixo custo de armazenamento e transmissão via redes de dados. Por volta dos anos 80 surgiu o conceito de escritório sem papel (*Paperless Office*), segundo o qual o papel seria abolido das empresas, embora atualmente as predições apontem que a coexistência do papel com outras tecnologias seja mais provável [1][2].

A partir da chegada dos dispositivos para a digitalização de documentos, diversas pesquisas estão sendo realizadas para melhorar o processo de captura de imagens, fazendo-as ficarem mais nítidas e reduzindo o espaço para armazenamento. A popularização das câmeras digitais portáteis, sobre tudo pelo fato das mesmas estarem cada vez mais presentes em outros dispositivos eletrônicos, juntamente com a evolução da tecnologia ocasionando a melhora da qualidade das imagens geradas por esses dispositivos. Entretanto, em virtude de câmeras digitais não terem sido projetadas inicialmente para digitalização de documentos, sua aplicação nesse âmbito torna-se limitada, fazendo-se necessária a realização de pesquisas para adequar estas imagens às aplicações comuns a documentos digitalizados, como reconhecimento de caracteres. Esta atitude fez surgir uma nova área de pesquisa, que avança para a solução dos diferentes problemas característicos da aquisição de imagens por estes dispositivos.

O estudo para este tipo de imagem assemelha-se aos realizados em imagens adquiridas por *scanners*, mas apresenta características e problemáticas próprias. Embora cada problema tenha dificuldades específicas, todos estes visam produzir imagens de melhor qualidade. Um exemplo típico de imagem de documento é ilustrado na Figura 1.1 a qual foi fotografada à mão livre e sem suporte mecânico. É possível observar que a imagem apresentada na Figura 1.1, é legível para a maioria das pessoas, ainda que a presença de borda em volta do documento reduza efetivamente a resolução dos caracteres, uma vez que parte da área útil é utilizada para a borda da imagem que não

incorpora nenhuma informação. Ainda é possível perceber outras distorções, como distorção de perspectiva, iluminação irregular, e curvatura nas extremidades do documento.



PhotoDoc: A Toolbox for Processing Document Images Acquired Using Portable Digital Cameras

Gabriel Pereira e Silva and Rafael Dueire Lins,
Departamento de Eletrônica e Sistemas – Universidade Federal de Pernambuco - Brazil
gfps@cin.ufpe.br, rdl@ufpe.br

Abstract

This paper introduces PhotoDoc a software toolbox designed to process document images acquired with portable digital cameras. PhotoDoc was developed as an ImageJ plug-in. It performs border removal, perspective and skew correction, and image binarization. PhotoDoc interfaces with Tesseract, an open source Optical Character Recognizer originally developed by HP and distributed by Google.

1. Introduction

Portable digital cameras are omnipresent in many ways of life today. They are not only an electronic device on their own right but have been embedded into many other devices such as portable phones and palmtops. Such availability has widened the range of applications, some of them originally unforeseen by their developers. One of such applications is using portable digital cameras to acquire images of documents as a practical and portable way to digitize documents saving time and the burden of having either to scan or photocopy documents. Figures 01 to 04 present different documents digitized using different camera models.

Figure 01 – Document image acquired with the camera of a LG cell phone KE-970 – (1600x1200 pixels) 409KB, without strobe flash

Figure 02 - Document image acquired with the camera of HP iPaq rx3700 (1280x960 pixels) 162KB, without strobe flash

This new use of portable digital cameras gave birth to a new research area [1][2] that is evolving fast in many different directions.

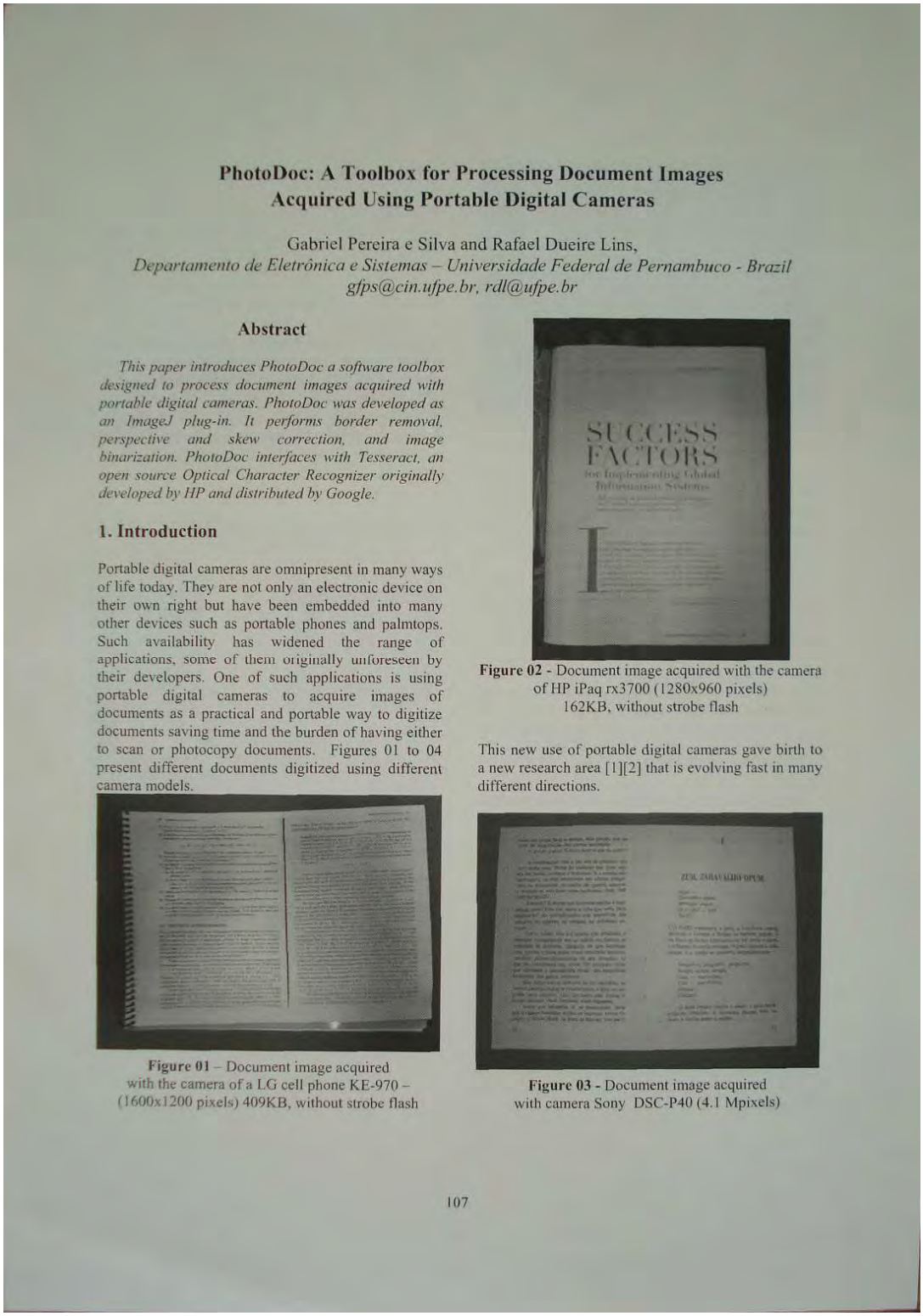
Figure 03 - Document image acquired with camera Sony DSC-P40 (4.1 Mpixels)

107

Figura 1.1 - Ilustração de documento fotografado à mão livre.

A Figura 1.2 apresenta o documento da Figura 1.1 processado pelo PhotoDoc, tendo sua borda removida e a perspectiva e inclinação corrigidas. Esse processamento proporcionou uma redução do tamanho da imagem de 503KB para 322KB, ambas utilizando o formato de arquivo *JPEG* com a mesma taxa de compressão. Outro fator decisivo que viabiliza tal processamento se

deve ao fato da impressão referente à Figura 1.2 consumir aproximadamente 50% menos recursos de *toner* que a Figura 1.1, considerando-se impressão em preto-e-branco, uma vez que o número de pixels escuros foi reduzido nessa proporção.



PhotoDoc: A Toolbox for Processing Document Images Acquired Using Portable Digital Cameras

Gabriel Pereira e Silva and Rafael Dueire Lins,
Departamento de Eletrônica e Sistemas – Universidade Federal de Pernambuco - Brazil
gfps@cin.ufpe.br, rdl@ufpe.br

Abstract

This paper introduces PhotoDoc a software toolbox designed to process document images acquired with portable digital cameras. PhotoDoc was developed as an ImageJ plug-in. It performs border removal, perspective and skew correction, and image binarization. PhotoDoc interfaces with Tesseract, an open source Optical Character Recognizer originally developed by HP and distributed by Google.

1. Introduction

Portable digital cameras are omnipresent in many ways of life today. They are not only an electronic device on their own right but have been embedded into many other devices such as portable phones and palmtops. Such availability has widened the range of applications, some of them originally unforeseen by their developers. One of such applications is using portable digital cameras to acquire images of documents as a practical and portable way to digitize documents saving time and the burden of having either to scan or photocopy documents. Figures 01 to 04 present different documents digitized using different camera models.

Figure 01 – Document image acquired with the camera of a LG cell phone KE-970 – (1600x1200 pixels) 409KB, without strobe flash

Figure 02 - Document image acquired with the camera of HP iPaq rx3700 (1280x960 pixels) 162KB, without strobe flash

This new use of portable digital cameras gave birth to a new research area [1][2] that is evolving fast in many different directions.

Figure 03 - Document image acquired with camera Sony DSC-P40 (4.1 Mpixels)

107

Figura 1.2 - Documento da Figura 1.1 processado pelo PhotoDoc.

No caso de documentos impressos é essencial que os caracteres das imagens dos documentos se mantenham legíveis, de forma a permitir a transcrição automática do formato de imagem para o de

texto editável. O reconhecimento de caracteres impressos em documentos ou até mesmo manuscritos em letras de forma, obtidos através de dispositivos de digitalização com alta resolução e ambiente controlado, é visto como um problema onde é possível obter uma excelente transcrição por meio de ferramentas comerciais de OCR. Já questões envolvendo documentos adquiridos com baixa resolução ou com alto grau de interferência externa, como é o caso da maioria das câmeras digitais, possuem uma transcrição problemática, e são consideradas temas de muito interesse em pesquisas.

É notável a quantidade de variáveis que influenciam o processo de obtenção de imagens digitais de boa qualidade, como: parâmetros de digitalização (brilho, contraste, resolução, número de cores, etc), inclinação do documento, distorções geométricas e de perspectiva, perda de foco, entre outras [78]. Com base nessas variáveis, pesquisadores desenvolveram modelos de degradação de documentos [77], os que serviram de base para o desenvolvimento desta dissertação.

1.1 Motivação

As imagens de documentos em papel são quase inevitavelmente degradadas no decurso de impressão, fotocópias e digitalização. A perda de qualidade, mesmo quando parece insignificante para olhos humanos, pode ser responsável por uma queda abrupta nos sistemas OCR. Essa fragilidade dos sistemas OCR, quando confrontado por uma imagem de baixa qualidade ou com alto grau de interferência (ruídos), é bem conhecida [76]. A precisão dos algoritmos de reconhecimento existentes cai abruptamente quando se degrada a qualidade de imagem mesmo ligeiramente [76] [77]. A definição da qualidade das imagens é comumente determinada de forma subjetiva pelos operadores responsáveis por manipulá-las, o que torna o processo dispendioso, falho e não padronizado.

Definir e estruturar uma solução para os problemas apresentados nos parágrafos anteriores a esta seção é essencial, já que o uso de câmeras digitais portáteis para digitalização de documentos apresenta um crescimento extraordinário, demandando por soluções robustas de fácil uso, o que torna o PhotoDoc um ambiente único, para o tratamento de imagens fotografadas de documentos.

1.2 Objetivos

Nesta dissertação tem-se por objetivo o desenvolvimento de algoritmos que busquem a melhoria das imagens de documentos fotografados [50], buscando uma melhor transcrição automática destas imagens por ferramenta de OCR. Esses algoritmos estão reunidos em um ambiente chamado PhotoDoc [51], que busca prover uma considerável melhoria na qualidade desses documentos, bem como abrir caminho para a solução de outros fatores indesejados causados durante a digitalização de documentos por câmeras digitais.

1.3 Metodologia

A metodologia adotada nesta dissertação é composta por três etapas bem definidas: classificação, correção de distorções e comparação. Sendo as duas primeiras etapas implementadas no ambiente PhotoDoc.

A etapa de classificação consiste em agrupar as imagens quanto a seu domínio de classe e dispositivo utilizado para sua digitalização, possibilitando um melhor pré-processamento das imagens.

Já a etapa de correção é a responsável por melhor automaticamente as imagens de documentos fotografados a partir de quatro passos:

- Detecção das bordas dos documentos, que consiste em detectar a fronteira entre o conteúdo e os objetos externos;
- Correção de perspectiva, que é uma transformação geométrica a fim de tornar a imagem como se estivesse vendo-a frontalmente;
- Realce das imagens, que consiste em diminuir o efeito negativo da iluminação sobre os documentos, deixando os caracteres mais visíveis;
- Binarização dos documentos, através da conversão de imagens coloridas em monocromáticas.

Por último, a etapa de comparação que é responsável pela medição quantitativa dos resultados.

1.4 Organização do Trabalho

Seguindo a metodologia adotada e juntamente com experimentos realizados no decorrer desta dissertação observou-se o melhor fluxo de execução dos algoritmos desenvolvidos. A Figura 1.1 ilustra este fluxo.

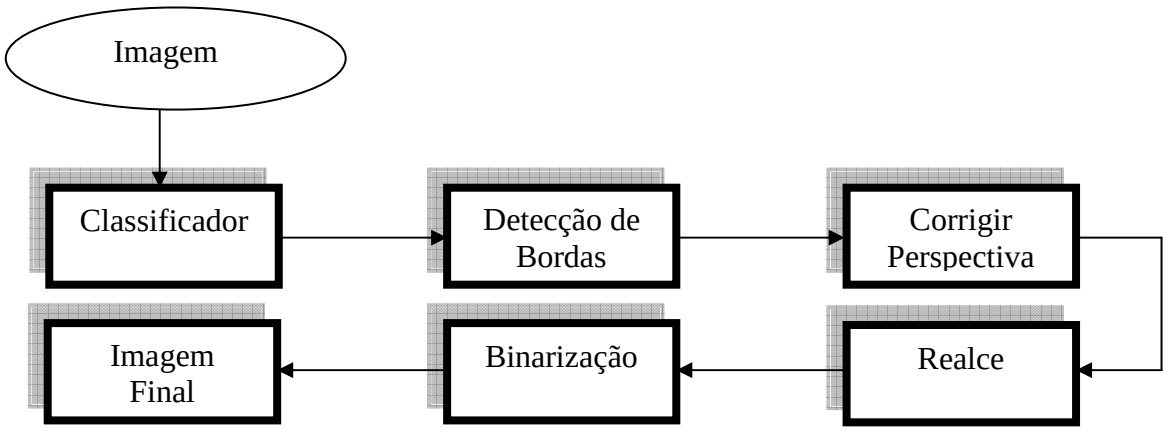


Figura 1.4 - Fluxograma de execução do PhotoDoc.

Para melhor entendimento optou-se organizar este trabalho seguindo a ordem de execução apresentada na Figura 1.4. Esta dissertação está dividida em dez capítulos e três anexos, sendo o primeiro capítulo esta introdução.

O capítulo 2 apresenta os princípios dos dispositivos de digitalização mais usuais, citando as vantagens, desvantagens e os problemas que normalmente ocorrem durante a digitalização de documentos e alguns conceitos para melhor compreensão desta dissertação.

O capítulo 3 apresenta as características e singularidades presentes nas imagens de documentos adquiridas por intermédio de câmeras digitais portáteis.

No capítulo 4 é apresentado um novo algoritmo para classificação automática de imagens.

O capítulo 5 discute a detecção das bordas de documentos adquiridos por câmeras digitais portáteis.

O capítulo 6 apresenta um algoritmo para correção da distorção de perspectiva.

O capítulo 7 apresenta um novo algoritmo de realce para documentos fotografados.

No capítulo 8 é realizado um estudo sobre a precisão de algoritmos de binarização aplicado sobre imagens de documentos adquiridas por meio de câmeras digitais portáteis.

O capítulo 9 apresenta uma análise dos resultados obtidos pelo processamento do PhotoDoc sobre as imagens de documentos .

O anexo A traz o manual do PhotoDoc, enquanto o anexo B apresenta a lista das publicações obtidas no desenvolvimento desta dissertação juntamente com uma cópia de cada um dos trabalhos. Por fim, o anexo C inclui um DVD com:

- Versão *pdf* desta dissertação;
- Software de instalação do ImageJ;
- *Pluging* PhotoDoc na versão executável;
- Manual do PhotoDoc na versão *pdf*;
- Imagens de teste utilizadas nesta dissertação.

Capítulo 2 Dispositivos para digitalização de documentos

Este capítulo introduz os princípios dos dispositivos de digitalização mais usuais, juntamente com suas vantagens, desvantagens e os problemas que normalmente ocorrem durante a digitalização de documentos. Para esta dissertação os dispositivos analisados foram os *scanners* de mesa e as câmeras fotográficas digitais portáteis tendo como foco os ruídos provenientes da digitalização.

2.1 Digitalização de documentos por *scanners*

Basicamente, *scanners* de mesa são compostos por um anteparo (local onde fica posicionado o papel), lâmpada, espelhos, lentes, filtros e um sensor, que pode ser uma matriz CCD (*charge-coupled device*) ou um sensor de contato CIS (*contact image sensor*), responsável por captar as imagens. No processo de captura, o documento é posto sobre uma superfície de vidro, e a tampa do escaner é fechada. Uma lâmpada é utilizada para iluminar o documento, enquanto o mecanismo completo (espelhos, lentes, filtros e sensores) se move, passando por toda a superfície de vidro. A imagem do documento é refletida por espelhos, cujo último espelho reflete a imagem em lentes, esta lente foca a imagem através de um filtro no sensor. Tal processo é ilustrado na Figura 2.1.

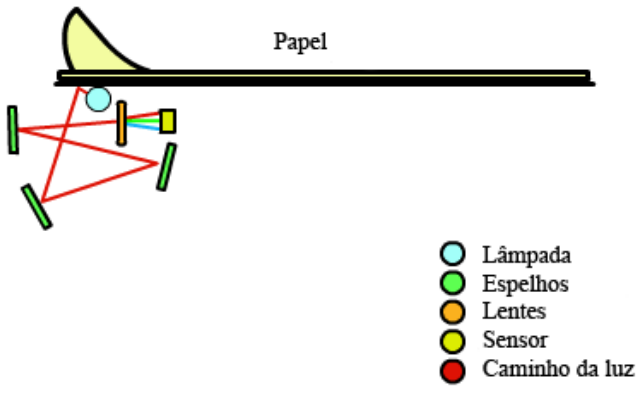


Figura 2.1 - Processo de digitalização por *scanners*.

Os *scanners* digitais têm sido os dispositivos predominantes na digitalização de documentos na última década.

Tipicamente os *scanners* são encontrados em quatro formas:

• *Scanners* de mesa é o tipo comumente utilizado, encontrados muitas vezes associado a impressoras. Neste caso o usuário alimenta o *scanner* com um documento por vez;

• *Scanners* com alimentação automática são semelhantes aos *scanners* de mesa, exceto que nesse caso o usuário pode inserir uma pilha de documentos cabendo ao escaner mover os documentos um a um;

• *Scanners* de mão utilizam basicamente a mesma tecnologia dos *scanners* de mesa, mas depende do usuário para mover o mecanismo de digitalização sobre o documento. Tipicamente não produz imagens de boa qualidade, entretanto podem ser úteis na captura rápida de textos;

• *Scanners* de linha de produção utilizam mecanismos mais complexos para aquisição das imagens, buscando melhor qualidade além de velocidade. Normalmente são utilizados para digitalização de grandes acervos bibliográficos.

2.2 Digitalização de documentos por câmeras fotográficas digitais

Assim como câmeras fotográficas convencionais, as do tipo digitais também podem ser modeladas como uma câmera escura, que é representada por uma caixa com um pequeno orifício onde a imagem é projetada de cabeça para baixo no lado oposto ao orifício. A Figura 2.2 ilustra esse modelo, que também é conhecido como *pinhole*, esse modelo é popular devido à sua fácil modelagem matemática, tido como o mais simples possível [13][16].

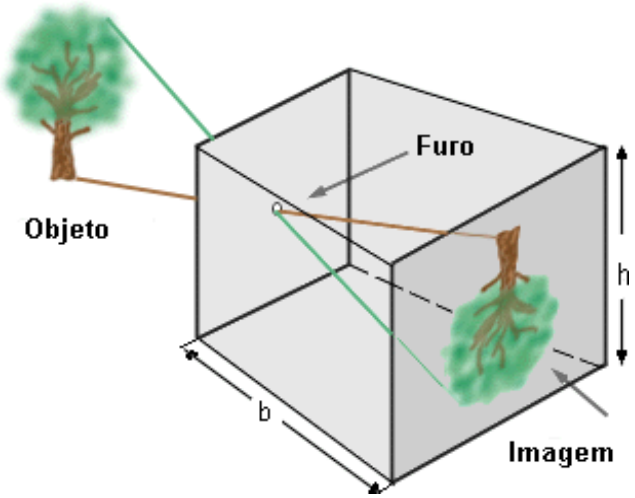


Figura 2.2 – Modelo de câmera escura.

Porém, as câmeras fotográficas digitais ao invés de armazenar a luz em um filme, fazem de maneira eletrônica. Após a captura da imagem, uma rotina de programação é executada para transformar os impulsos elétricos em um arquivo de imagem. Tal processo é possível através do uso de um ou mais sensores, responsáveis pela conversão de luz em impulsos elétricos.

Nesta dissertação, escolheu-se detalhar as tecnologias de sensores de capturas que estão em vigor atualmente, os mecanismos que usam esses sensores para formar a imagem final e o mecanismo de focalização da imagem.

2.2.1 Sensores de captura

O sensor de captura é o componente responsável pela digitalização da imagem. Neles as imagens são projetadas e transformadas em sinais digitais. Os sensores mais comumente utilizados são os CCD (*charge-coupled device*) e CMOS (*complementary metal-oxide-semiconductor*).

A tecnologia dos sensores CCD surgiu no final década de 60, emprega basicamente silício e é constituído por uma superfície onde é projetada a luz, nesse processo há a captura dos pontos da imagem. Depois da exposição, os estímulos captados são transmitidos para uma unidade que os converte em sinais digitais. Devido a essa arquitetura os CCD funcionam através da leitura/escrita seqüencial uniforme.

Já a tecnologia empregada nos sensores CMOS é a mesma utilizada na fabricação da maioria dos componentes eletrônicos atuais, o que facilita a integração entre o sensor e os demais componentes eletrônicos de uma câmera digital. Essa tecnologia também permite que cada ponto seja acessado de forma independente e a leitura é feita por linhas da imagem.

As principais vantagens da utilização dos sensores CMOS em relação ao CCD são: menor custo de produção, baixo consumo energético e menor dimensão. Como desvantagens do CMOS em relação ao CCD destacam-se melhor relação sinal-ruído (SNR) e dificuldades ao capturar imagens em ambiente com pouca iluminação.

Em geral, as câmeras de telefones celulares usam a tecnologia CMOS, uma vez que esses dispositivos possuem limitações quanto ao tamanho e consumo energético. Já as câmeras dedicadas (profissionais e amadoras) utilizam um dos dois tipos de sensores, esta escolha depende do tipo de aplicação que a câmera é voltada (ambiente com pouca iluminação, SNR, etc).

2.2.2 Mecanismos de captura de imagens

Uma das dificuldades na aquisição de imagens por câmeras fotográficas digitais ocorre no processo de armazenamento das cores. Os sensores captam a luz numa determinada faixa de freqüência. No caso das câmeras digitais, a luz visível é capturada com sensores nas faixas de vermelho, verde e azul. O que diferencia um mecanismo de captura de outro é o arranjo dos sensores, a quantidade de vezes que eles são expostos, o tempo de exposição e como a imagem final é construída.

São três os métodos mais utilizados:

- No primeiro método são utilizados diferentes sensores para as cores vermelho, verde e azul (sistema RGB). Tal processo é eficiente, entretanto as câmeras que o utilizam possuem alto custo;
- No segundo método um separador de cores envia uma cor por vez para o sensor, porém a imagem precisa estar parada por um período maior, inviabilizando sua utilização na maioria das aplicações;
- No terceiro método utiliza-se apenas um sensor, onde cada ponto da imagem no sensor recebe uma cor e posteriormente são realizadas interpolações para definição da cor final do

ponto na imagem. Tal processo pode ser observado na Figura 2.3. A distribuição mais comum dessas cores ao longo da imagem é o filtro de Bayer, ilustrado na Figura 2.4. Este é o método utilizado na maioria das câmeras fotográficas digitais.

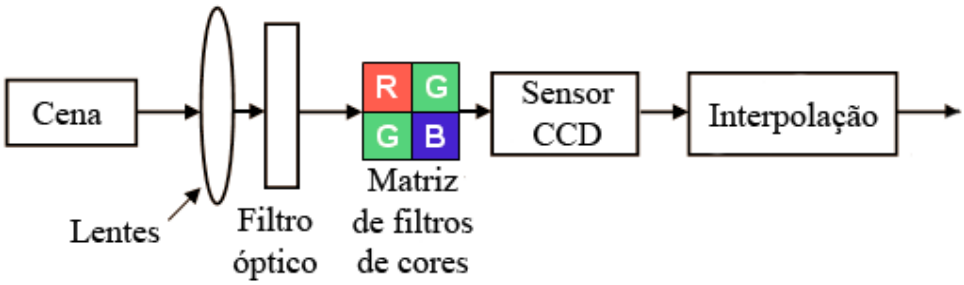


Figura 2.3 - A formação das imagens em câmeras fotográficas digitais portáteis.

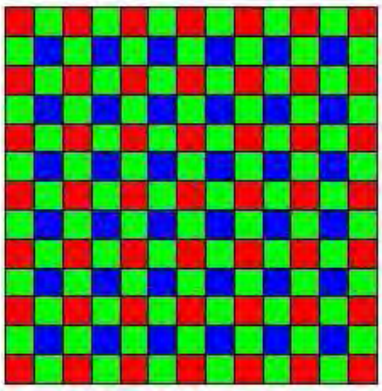


Figura 2.4 - Filtro de Bayer.

A digitalização das imagens obtidas a partir do terceiro método de captura que usa a interpolação das cores ao longo do filtro de Bayer pode resultar em imagens com algumas falhas, tais como o *aliasing (*transição não-suave entre cores), texto borrado devido à atenuação das altas freqüências, o que leva a redução da qualidade e aumento nos erros de reconhecimento de caracteres [9] [49].

2.2.3 Mecanismos de Foco

O foco é um mecanismo presentes nas câmeras fotográficas que tem como objetivo enfatizar uma dada área da cena a ser fotografada. Existem três tipos de mecanismos de foco: manual, fixo e automático.

- No foco manual, o ajuste é realizado pelo fotógrafo, ajustando a lente da câmera sem a necessidade de mover a câmera como um todo para focalizar uma parte da cena. Este mecanismo é comum em câmeras analógicas e câmeras digitais de uso dedicado;
- Já no foco fixo, não há ajuste, para enfatizar parte de uma cena o fotógrafo deve mover a câmera como um todo até uma distância em que enquadre a área de interesse. O mecanismo é comum em câmeras digitais de uso não dedicado (câmeras de celulares);
- Para o foco automático as principais formas de focalização são a ativa, passiva e preditiva.
    - A focalização ativa faz uso de sensores (ultra-som ou infravermelho) para calcular a distância do “alvo”. Através da velocidade da onda é possível aferir essa distância.

Uma limitação deste tipo de focalização ocorre quando há meios translúcidos diferente do ar (vidro, plástico, etc.) ou alguma outra forte fonte emissora de onda (luzes, velas, etc.) entre o objeto a ser fotografado e a câmera, fazendo com que o mecanismo erre a focalização [75].

- Já a focalização passiva consiste em analisar a luz que incide sobre a câmera e ajustar a posição da lente através dessa análise. Essa análise subdivide-se em duas: ajuste por contraste e detecção por fase. Mais detalhes em [75].
  - No ajuste por contraste calcula-se o nível de contraste à medida que se move a lente. A posição que obter maior contraste será utilizada para capturar a imagem;
  - Já na detecção por fase, a idéia é projetar duas imagens com origem no mesmo ponto e calcular a distância necessária para alinhar essas imagens, que devem ser simétricas na imagem focalizada. Isso é possível colocando dois sensores lineares em conjunto com lentes separadoras atrás do plano do dispositivo de captura, esses sensores irão capturar dois raios de luz de um mesmo ponto. Com o sinal obtido é possível calcular o quanto é necessário ajustar a lente para focalização. Esse processo é extremamente rápido, no entanto é caro, pois são necessários componentes adicionais (lentes e sensores lineares).
- A focalização preditiva foi desenvolvida pela Nikon [58] para tirar seqüências de fotos em movimento (usadas em câmeras profissionais). À medida que o objeto é focalizado, o sistema faz uso de informações do instante ($t'=t-1$) para prever a focalização no instante t, desta forma diminuindo o intervalo de tempo entre focalizar e tirar foto.

2.2.4 Desafios encontrados na digitalização

Atualmente grande parte dos softwares de processamento de imagens foi concebida a fim de filtrar documentos escaneados. Esses softwares podem ser usados para o processamento de imagens fotografadas produzindo bons resultados para documentos com baixos níveis de ruído, porém utilizá-los exige algum nível de treinamento dos usuários. São vários os ruídos presentes em imagens de documentos. Segundo [78] podemos classificar esses ruídos em quatro tipos:

- Ruído físico é quando há danos à integridade física e a legibilidade da informação original de um documento. O ruído físico pode ainda ser dividido em duas subcategorias [98], como interna e externa;
- Ruído de digitalização é o ruído introduzido pelo processo de digitalização. Vários problemas podem ser agrupados neste caso, tais como: resolução insuficiente, palheta inadequada, rotação e erro de orientação, distorção da curvatura da lente e distorções geométricas;

- Ruído de filtragem é a manipulações impróprias do arquivo digital podem degradar a informação existente na versão digital do documento (em vez de melhorá-la). A introdução de cores não originalmente presentes no documento, devido à manipulação aritmética, ou por *overflow* é um exemplo deste tipo de ruído;
- Ruído de armazenamento / transmissão – surge devido a algoritmos de armazenagem ou rede de transmissão com perdas. Artefato *JPEG [98]* é um exemplo típico deste tipo de interferências indesejáveis.

Para esta dissertação será dado enfoque aos ruídos de digitalização, sobretudo aos mais comuns encontrados em imagens de documentos adquiridos por meio de câmeras digitais portáteis. A Tabela 2.1 apresenta alguns desses ruídos e faz uma breve comparação entre *scanners* e câmeras digitais.

	Scanners	Câmeras Digitais
Resolução	100-600 dpi	50-600 dpi
Superfície	Plana	Arbitrária
Distorção	Mínima	Perspectiva e óptica
Iluminação	Adequada	Difícil de controlar
Plano de fundo	Conhecido	Complexo
Velocidade	Lenta	Rápida
Foco	Fixo	Variável
Portabilidade	Ruim	Boa
Usabilidade	Baixa	Alta
Tamanho	Grande	Compacto

Tabela 2.1 - Comparação entre *scanners* e câmeras digitais.

De acordo com a Tabela 2.1, podem-se listar os principais desafios encontrados na digitalização de imagens que incluem:

- Iluminação inadequada: câmeras digitais portáteis têm menos controle das condições de iluminação do que *scanners*, devido principalmente ao ambiente físico (sombras, reflexões) e respostas inadequadas dos dispositivos. Tais complicações ocorrem mesmo quando a iluminação ambiente é controlada. Os flashes das câmeras digitais portáteis foram desenvolvidos para iluminar o alvo da fotografia em distâncias superiores a um metro e meio, enquanto a fotografia de documentos exige maior proximidade da câmera devido à busca por maior resolução dos caracteres o que faz a distância média ser inferior a quarenta centímetros, induzindo a ocorrência de áreas com luminosidade irregular;
- Distorção de perspectiva: ocorre devido à utilização à mão livre das câmeras digitais portáteis sem o auxílio de suporte mecânico, o que faz com que o plano do documento não seja paralelo ao plano da câmera. Como resultado, os caracteres mais distantes da base da imagem possuem menor resolução e as linhas paralelas verticais não mais são preservadas. As ferramentas comerciais de OCR são sensíveis mesmo a pequenas distorções de perspectiva [26][49] o que já é suficiente para causar problemas significantes

no reconhecimento dos caracteres. Para *scanners* em geral, documentos são perfeitamente alinhados com a superfície de vidro, não apresentando distorções de perspectiva, exceto os casos em que não é possível devido ao volume do documento (Na Figura 2.5 pode-se observar esta distorção);

- Distorções de lente (efeito esférico): as lentes produzem curvaturas do centro da imagem para as bordas, fazendo com que à medida que o documento aproxime-se da lente, torne a distorção maior. Este tipo de distorção causa dificuldade na segmentação da imagem, devido ao fato de o documento não mais representar um quadrilátero;
- Planos de fundo complexos: freqüentemente o suporte mecânico sobre o qual está o documento também é capturado na imagem com o documento fotografado. Se o documento não tiver contorno regular, isto o tornará difícil de segmentar (processo pelo qual uma imagem é subdividida em suas regiões ou objetos constituintes), mesmo se uma parte do plano de fundo apresentado na imagem for uniforme;
- Zoom e foco: muitos dispositivos digitais são projetados para operar em uma grande variedade de distâncias, tornando o foco um fator significante. As pequenas distâncias utilizadas para captação de imagens de documentos podem apresentar foco inadequado, mesmo com mínimas mudanças de perspectiva. O método de captação de cores pela câmera também pode atenuar as freqüências altas, causando a perda de foco [9];

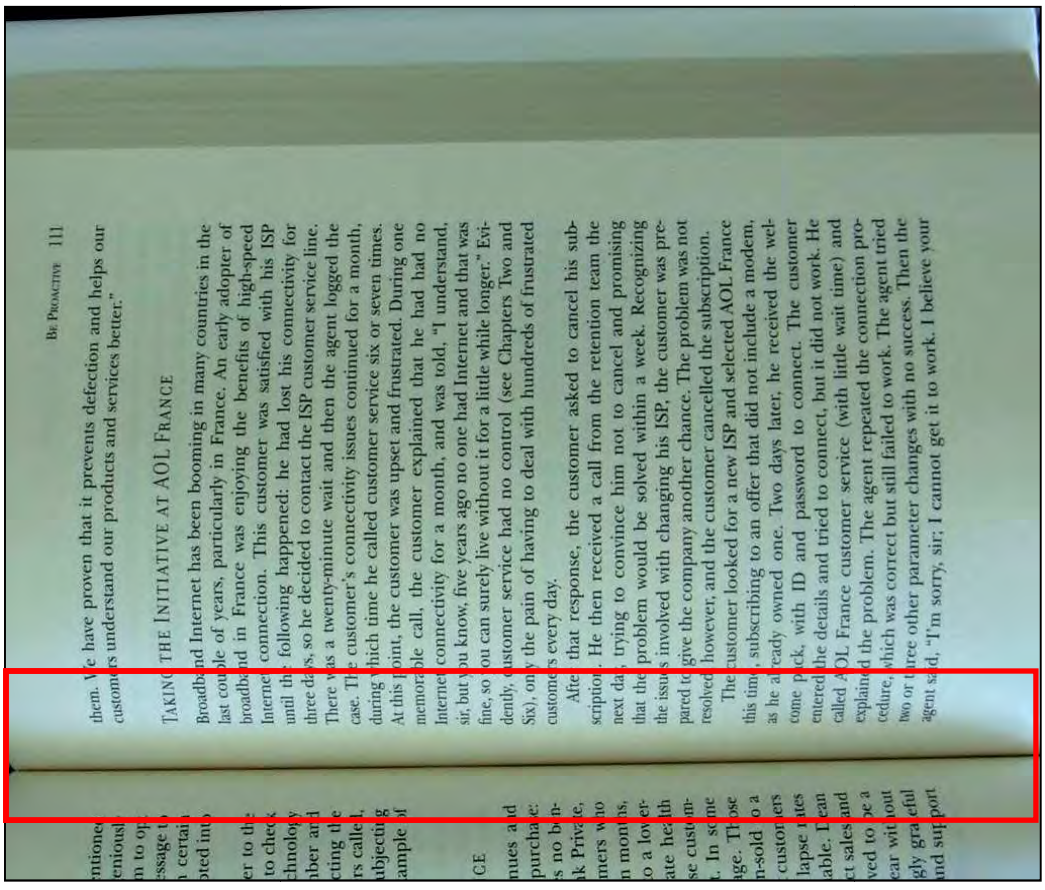


Figura 2.5 - Documento digitalizado através de escaner apresentando distorção geométrica.

Na Figura 2.6, é apresentado um documento fotografado com uma câmera digital Sony, modelo DSC-S40, a 4.1 *Mpixels* (2348 x 1684 pixels) com flash e no formato *JPEG* (perda ≈ 1%), apresentando distorções de perspectiva, curvatura e plano de fundo complexo.

A Figura 2.7 evidencia o mesmo documento digitalizado através de escaner da marca Hewlett-Packard, modelo 5300c, *true-color*, a 100dpi, com 1169 x 850 pixels, em formato bitmap sem compressão, apresentando pequena inclinação, erro comum na digitalização e facilmente corrigível, enquanto livre dos problemas citados sobre a Figura 2.6.

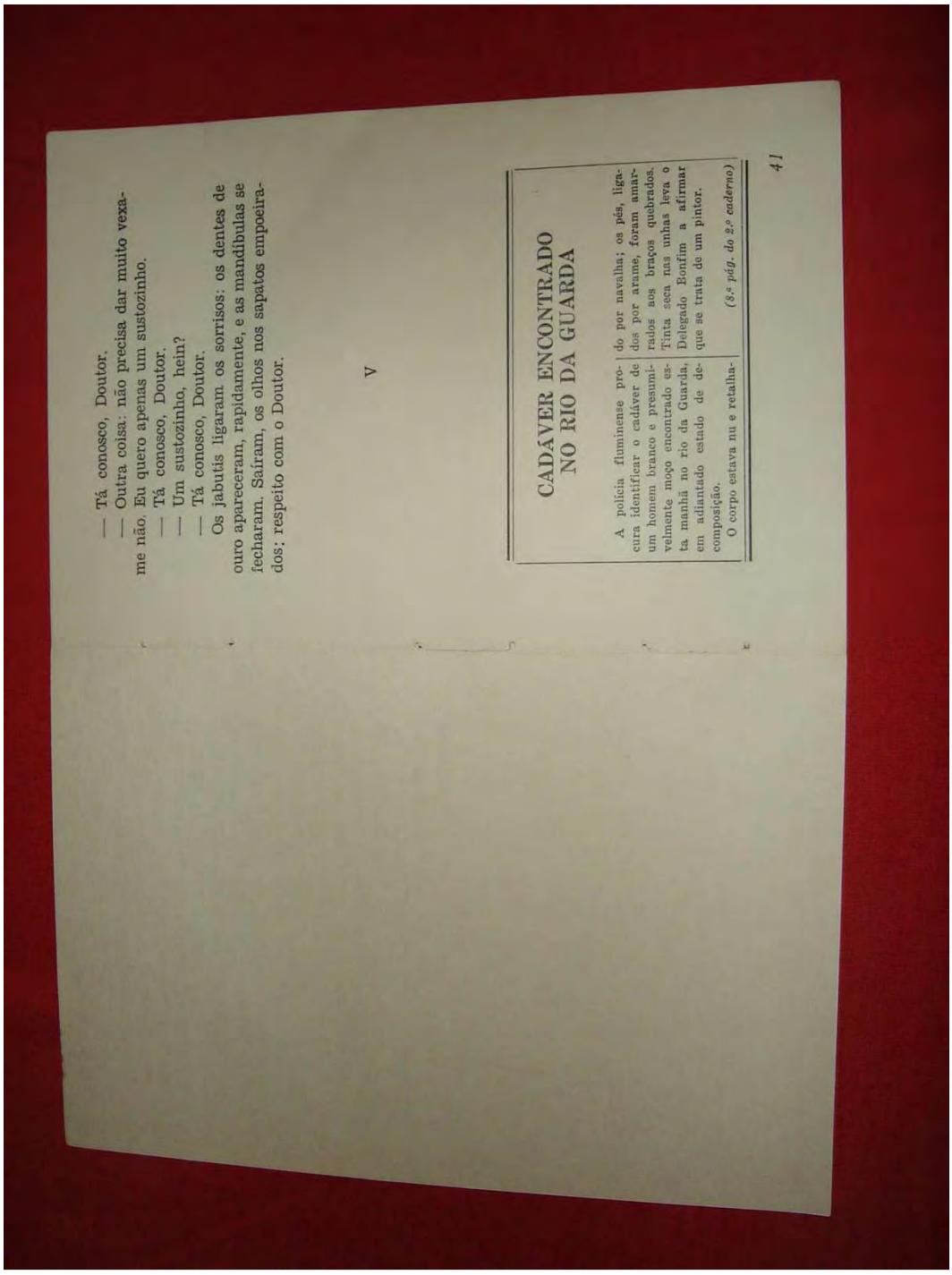


Figura 2.6 - Documento digitalizado através de câmera fotográfica digital com o uso de flash.

— Tá conosco, Doutor.
— Outra coisa: não precisa dar muito vexame não. Eu quero apenas um sustozinho.
— Tá conosco, Doutor.
— Um sustozinho, hein?
— Tá conosco, Doutor.

Os jabutis ligaram os sorrisos: os dentes de ouro apareceram, rapidamente, e as mandíbulas se fecharam. Saíram, os olhos nos sapatos empoeirados: respeito com o Doutor.

V

CADÁVER ENCONTRADO NO RIO DA GUARDA

A polícia fluminense procura identificar o cadáver de um homem branco e presumìvelmente moço encontrado esta manhã no rio da Guarda, em adiantado estado de decomposição.

O corpo estava nu e retalhado por navalha; os pés, ligados por arame, foram amarrados aos braços quebrados. Tinta seca nas unhas leva o Delegado Bonfim a afirmar que se trata de um pintor.

(8.ª pág. do 2.º caderno)

41

Figura 2.7 - Documento da Figura 2.6 digitalizado através de escaner HP 5300c, a 100dpi em true color.

Capítulo 3
Características dos Documentos Fotografados

As características mais comumente presentes em imagens fotográficas listadas na subseção 2.2.4 do Capítulo dois, encontram-se bem fundamentadas na literatura [13][17]. São tais desafios que a ferramenta PhotoDoc [51] busca tratar e solucionar.

Os algoritmos de processamento de imagens escaneadas podem ser úteis para imagens fotografadas, porém exigem a intervenção humana para fornecer parâmetros que possibilitem uma execução adequada. Um exemplo clássico é o tratamento das bordas dos documentos. Para a remoção dessas bordas é necessário indicar pontos de controle para cada imagem e aplicar um algoritmo de recorte, tornando o processo dispendioso e impreciso.

Portanto, fez parte desta dissertação analisar as características encontradas em documentos fotografados e utilizá-las para compor os algoritmos presentes no PhotoDoc, o qual é capaz de remover os ruídos característicos desse tipo de imagem, tais como: bordas, distorções de perspectiva e iluminação não uniforme. Para a pesquisa descrita neste trabalho foi criado um banco de imagens contendo mais de três mil fotografias, com resoluções de 3.2, 4.1, 5.1 e 7.2 *Mpixels* no formato *JPEG [98]*, juntamente com as imagens escaneadas a 100, 200 e 300 *dpis*, nos formatos *TIFF [*98], *PNG [98][*98] *e JPEG[98]*. Os documentos utilizados para compor o banco de imagem foram fotografados em formato *true color*, contendo páginas de livros, listas telefônicas, folhetos, apostilas, anais de conferencias e certificados, fotografados com e sem a ajuda de suporte mecânico e com uso de flash e sem o uso dele. A Figura 3.1 ilustra o planetário (suporte mecânico desenvolvido na UFPE) que busca sistematizar e modelar a distorção de perspectiva bem como a variação da distância entre a câmera digital portátil e os documentos utilizados neste estudo, que constam no apêndice C desta dissertação.

Dentre os documentos utilizados havia imagens com diferentes cores de fundo e *layout*. Os papéis constituintes de tais documentos eram translúcidos, apresentando em alguns casos fraca interferência frente-verso[30][31]. Os ambientes para digitalização destes documentos também foram variados (plano de fundo, iluminação ambiente, etc.), conforme se pode observar nas Figuras 3.2 e 3.3. Essas figuras ainda apresentam os histogramas R, G e B dessas imagens. Através dos histogramas, observa-se a grande variação na composição das imagens e a grande variação de cores nas bordas e nas figuras pode vir a fazer parte das imagens, o que torna impraticável a segmentação das bordas através da análise dos histogramas.

A análise das características do banco de imagens foi realizada através da observação das componentes R, G e B, observando a distribuição de cada uma dessas componentes, ao longo dos documentos e principalmente nos contornos destes. Verificaram-se também as características típicas

de figuras, textos, papéis e planos de fundo, assim como as variações de brilho em ambientes distintos.

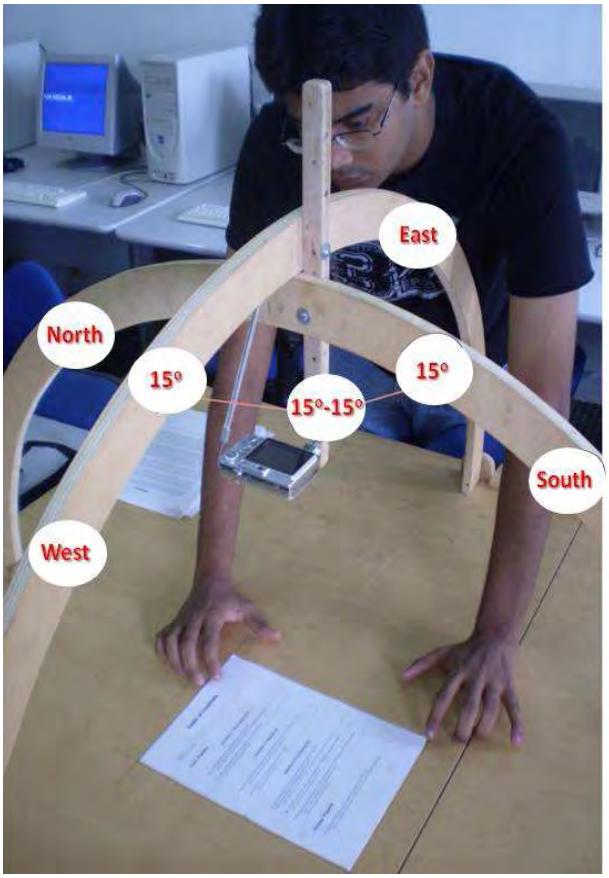
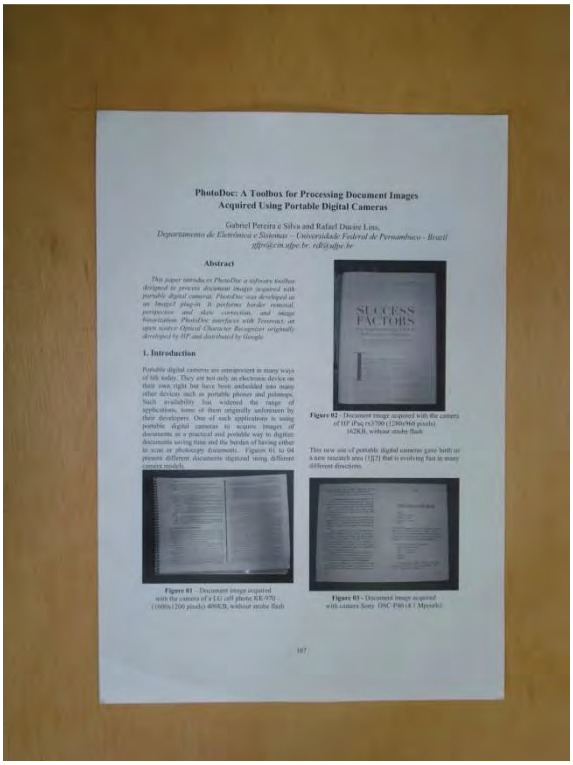
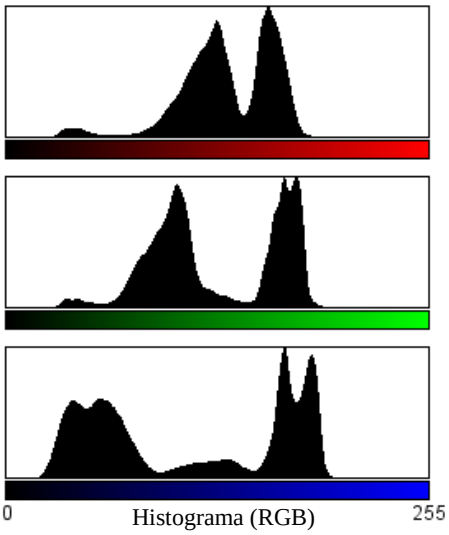


Figura 3.1 - Imagem do planetário.


PhotoDoc: A Toolbox for Processing Document Images Acquired Using Portable Digital Cameras



(Imagem (218,75KB JPEG) adquirida com o uso do planetário)

Figura 3.2 - Exemplos de documentos com variação do plano de fundo e tipo do documento.

(Imagem (304,57KB JPEG) adquirida à mão livre)

Figura 3.3 - Exemplos de documentos com variação do plano de fundo e tipo do documento.

As principais características observadas e identificadas como influentes no desenvolvimento dos algoritmos são descritas nas subseções seguintes.

3.1 O efeito da iluminação sobre os documentos

A grande maioria das câmeras digitais foi projetada para trabalhar com imagens complexas, em ambientes tridimensionais e com grande quantidade de cores. A distância entre essas câmeras e o objeto a ser fotografado é geralmente superior a um metro de distância, como é o caso de fotos de famílias, de paisagens, etc. Valendo-se dessa distância usual os dispositivos que emitem *flashs* em câmeras fotográficas foram projetados de forma a apresentar esse *flash* disperso, podendo ser considerado desprezível, ou tornar a fotografia visível em caso de ambientes de baixa iluminação. Ao utilizar-se de uma câmera digital para aquisição de documentos, usuários tende a aproximar a câmera do documento a fim de obter uma melhor resolução e eliminar o plano de fundo no qual o documento se encontra. Dessa forma, a distância entre a câmera e o documento costuma variar entre 40 cm e 60 cm. A essa distância, o flash causa iluminação irregular no documento, ou seja, ele não apresenta o efeito dispersivo, causando maior brilho em determinadas regiões do documento. O distúrbio causado pela distribuição irregular do brilho limita o uso dos algoritmos tradicionais de segmentação (detecção de bordas e binarização).

Pode-se constatar que em fotos com flash o brilho próximo à posição do dispositivo que emite o *flash* tende a ser maior e independente da iluminação do ambiente, diminuindo gradualmente ao afastar-se desse ponto, conforme pode ser observado na Figura 3.4. Em imagens obtidas sem o uso

de flash o ambiente exerce grande influência na composição da imagem, fazendo com que a região com maior brilho varie de uma imagem para outra.

Figura 3.4 - Região de maior atuação do flash.

Outro efeito indesejável que pode ser causado pela distribuição não uniforme de brilho faz com que a cor do papel do documento aproxime-se da cor da fonte, o que degrada o desempenho dos algoritmos de binarização aplicados sobre estas imagens. Tal efeito é evidente ao se observar a Figura 3.5, onde a cor destacada na parte central foi retirada do papel (R = 111, G = 138, B = 147) e a cor destacada na parte inferior foi retirada da fonte do documento (R = 114, G = 140, B = 141). Além disso, a cor da borda pode estar muito próxima da cor do papel o que limita a aplicação automática de algoritmos de segmentação de bordas, conforme ilustra a Figura 3.6. A diferença entre o fundo do documento e o plano de fundo é pequena. Ainda é possível observar que o contorno do documento não está bem definido.

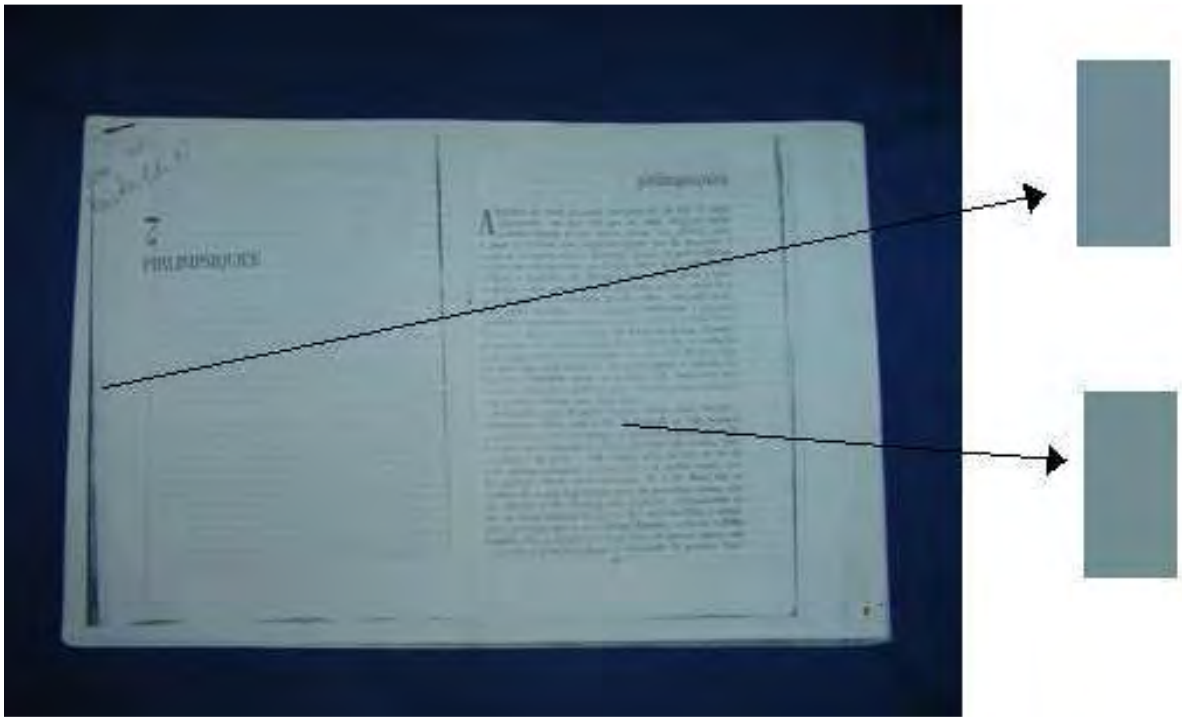
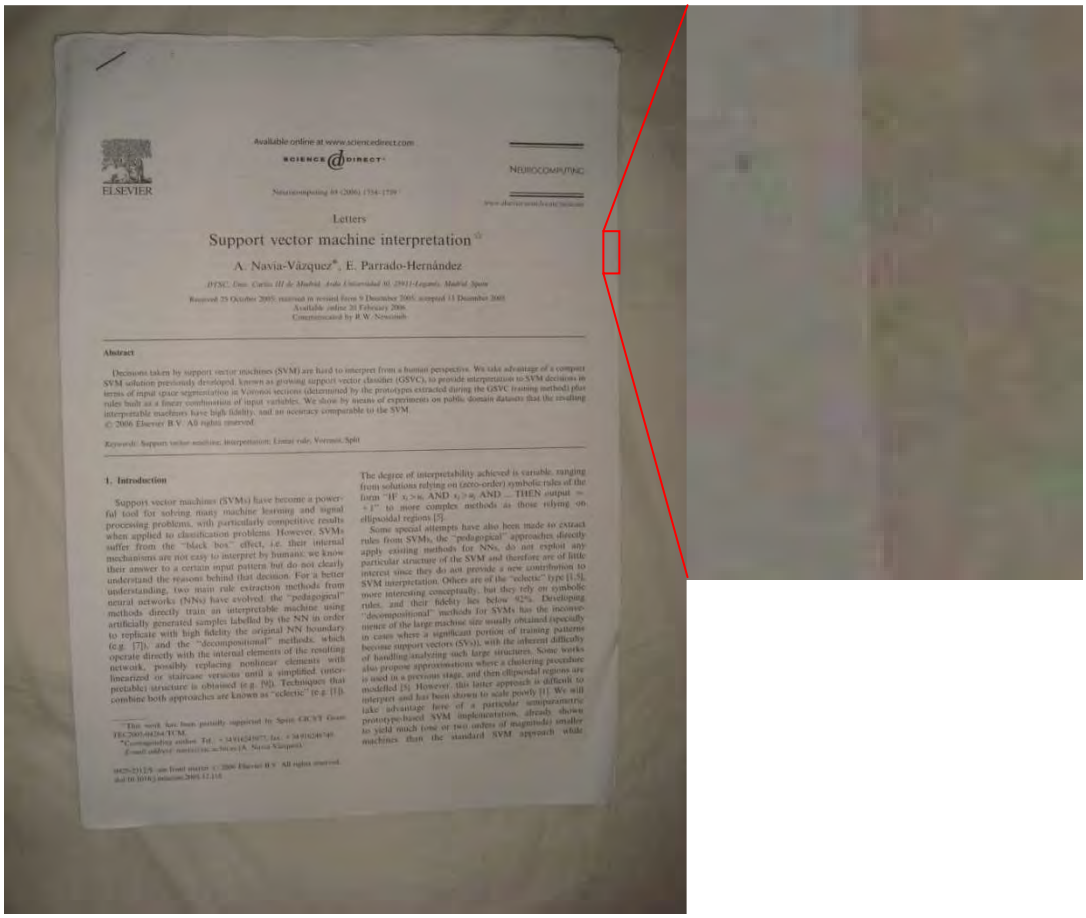


Figura 3.5 - Cores de papel e fonte com valores próximos.



Available online at www.sciencedirect.com
SCIENCE DIRECT
NEUROCOMPUTING
ELSEVIER

Letters

Support vector machine interpretation☆

A. Navia-Vázquez*, E. Parrado-Hernández

Abstract

1. Introduction

Figura 3.6 - Papel e borda com valores próximos de cores.

3.2 Os contornos em documentos fotografados

Em processamento de imagens pode-se definir contornos como o limite de um objeto e seu surgimento é associado a mudanças abruptas entre pixels consecutivos da imagem. No caso de imagens de documentos, há contornos em volta dos objetos que compõem o conteúdo (texto, tabelas figuras etc.) bem como entre o papel e o plano de fundo onde está fixado o documento. Para a remoção de bordas é comum a utilização da identificação do contorno do papel para a definição da área da imagem a ser recortada.

Grande parte das pesquisas na área de análise e detecção de contornos foi desenvolvida estudando-se imagens em tons de cinza [25]. Porém, a conversão para esse espaço de cores implica na diminuição da quantidade de informação efetiva, devido ao fenômeno conhecido como metamerismo. Esse fenômeno leva cores (*true color*), visualmente diferentes, a possuírem valores muito próximos.

Além disso, as imagens estudadas nessas pesquisas foram provenientes de *scanners* onde a captação de cores é realizada de forma mais adequada o que oferece maior intensidade dos contornos, enquanto nas imagens adquiridas por câmeras digitais a intensidade dos contornos tende a ser dependente do foco, além de estarem sujeitas aos efeitos descritos no Capítulo 2.

3.3 Conteúdo dos documentos

Documentos em geral tendem a possuir mais de um tipo de objeto que formam seu conteúdo, tais como: texto, tabelas, desenhos, figuras, gráficos, entre outros. Assim, os algoritmos para remoção automática de bordas devem ser eficientes ao ponto de identificar com certo grau de exatidão os objetos que compõem o conteúdo do documento e diferenciá-los da borda da imagem. Dessa forma é possível segmentar a borda sem afetar a informação presente no documento.

Observando-se inúmeros documentos digitalizados tanto por *scanners* quanto por câmeras digitais (processados pelo PhotoDoc), pode-se constatar o conteúdo dominante é o papel do documento, e que, excetuando-se os casos em que há imagens dentro do documento, a informação textual ocupa cerca de 3% a 5% da área da imagem referente ao documento. Tal informação é de grande utilidade já que a partir dela é possível estimar a cor media do papel do documento e assim segmentar o documento do seu plano de fundo.

Capítulo 4
Classificação automática de imagens

Classificação funcional de imagem consiste em separar diferentes tipos de imagens em classes que permitam otimizar seu processamento para leitura ou outra tarefa específica final. É uma importante área de pesquisa não só acadêmica como também para a indústria multimídia, tendo uma infinidade de aplicações. Por meio desta classificação podem-se tratar as imagens para obter uma melhor impressão ou até mesmo ajustar os parâmetros de uma câmera fotográfica digital (foco, intensidade do flash, etc.) durante o processo de captura de acordo com o "alvo" a ser fotografado (paisagens, documentos, faces, etc.) diminuindo os efeitos dos ruídos de digitalização em documentos, tais como os discutidos no Capítulo 2.

Este Capítulo apresenta recentes pesquisas sobre o assunto, juntamente com um método proposto para a classificação de imagens[80][79].

4.1 O estudo da classificação de imagens

O estudo sobre a classificação funcional de imagens foi iniciado no início da década de 80 [82][83], sobretudo devido à popularização dos *scanners* e tendo por objetivo agrupar imagens semelhantes de uma mesma massa de dados. Esses agrupamentos teriam a função de facilitar a pesquisa e recuperação de informações de grandes bases de imagens. Inicialmente uma imagem era tomada e depois comparada com as demais imagens da base, a idéia básica é tentar organizar essas imagens utilizando algumas características em comum [83][84]. As mesmas características observadas são usadas para analisar a imagem que irá servir de entrada para pesquisa e as contidas na massa de dados, ou seja, em vez de verificar imagem por imagem no conjunto de dados, o processo de recuperação tenta combinar as propriedades de uma imagem exemplo, com as diferentes imagens na base de dados para formar agrupamentos. Isto reduz drasticamente o espaço de busca, tornando o processo de recuperação mais eficiente. Diversas características são utilizadas para classificação de imagens, tais como análise do histograma [85][86] e informações sobre a semântica das imagens [87]. As imagens que possuem temas similares são mais susceptíveis a possuírem propriedades que são comuns entre si. Por outro lado, as imagens cujos temas são completamente descorrelacionados tendem a exibir propriedades muito diferentes.

Na presente dissertação optou-se em estudar três classes de imagens (Fotos, logotipos e documentos). A Figura 4.1 ilustra exemplos típicos de representantes das três classes de interesse aqui estudadas.

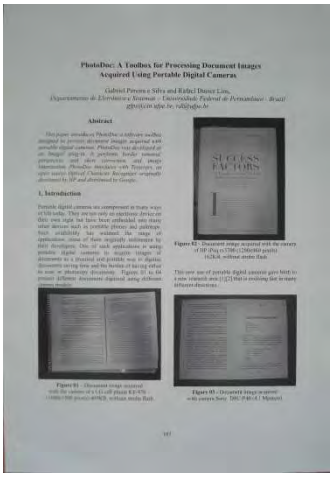


Foto Logo Documento

Figura 4.1 - Exemplo de imagens das três classes de interesse.

Inicialmente esta etapa da pesquisa foi voltada para tornar a impressão dessas imagens mais eficientes e de melhor qualidade visual, atendendo aos anseios da empresa HP (Hewlett-Packard). Em particular, documento, fotos e imagens de logotipo exigem tratamentos específicos para otimizar o seu aspecto quando copiados ou impressos. Documentos (texto, tabelas), por exemplo, exigem "afinamentos" que degradam a aparência das fotos e dos logotipos. Por sua vez, fotos podem ser inseridas em imagens de documentos ou logotipos, e esses por sua vez podem estar presentes nas fotos, o que torna necessário identificá-las e segmentá-las nesses documentos compostos.

Posteriormente visualizou-se a importância de classificar as imagens de documentos quanto aos dispositivos usados para a digitalização. Este novo enfoque é de particular importância para a ferramenta comerciais de OCR que poderão ajustar seus parâmetros ou mesmo solicitar o pré-processamento das imagens fazendo uma chamada a ferramenta PhotoDoc [51]. A Figura 4.2 ilustra as duas etapas da pesquisa.

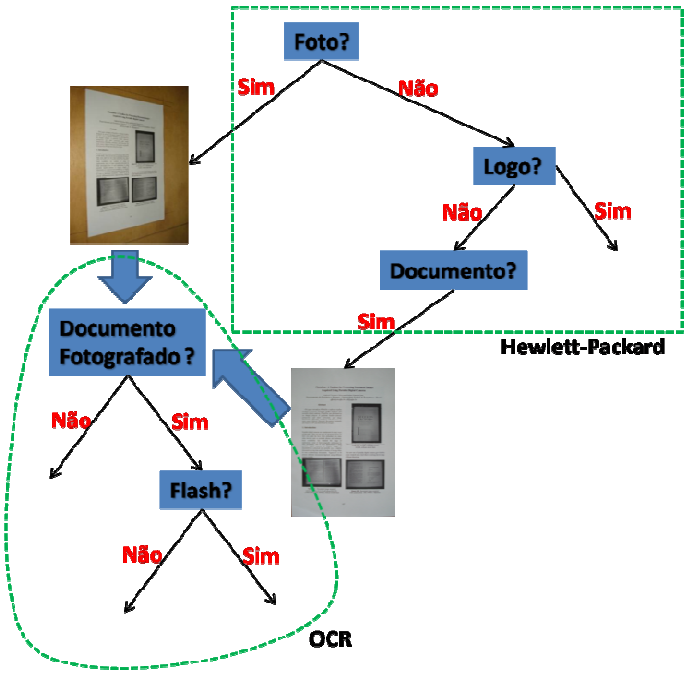


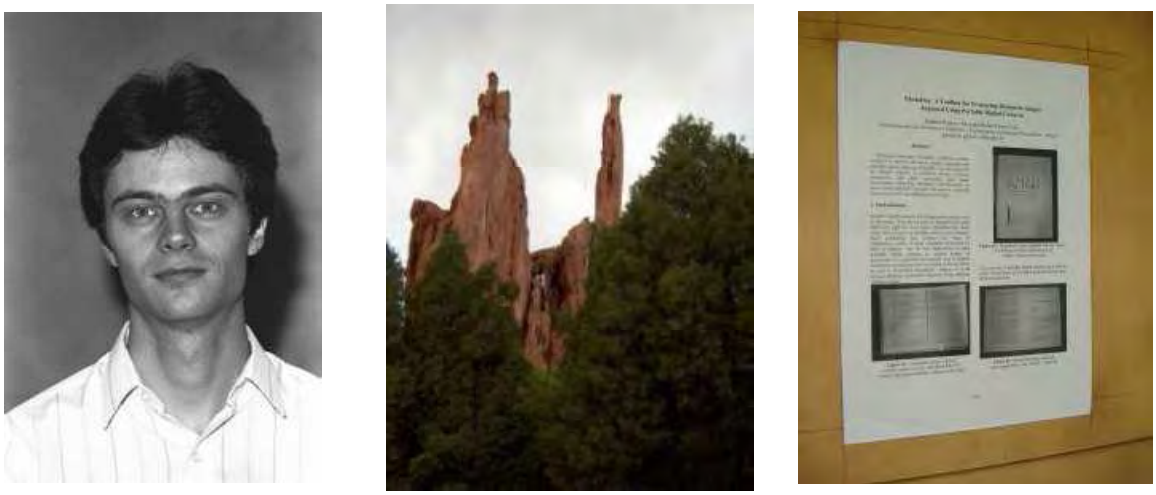
Figura 4.2 – As etapas da pesquisa de classificação de imagens.

4.2 Base de dados

O ponto de partida desta pesquisa foi a criação de um banco de dados representativo das classes de imagens de interesse. As imagens da base de dados foram rotuladas uma a uma manualmente, entretanto algumas vezes, a "mesma" imagem aparece nesta base com formatos diferentes, por exemplo, uma imagem pode aparecer nos formatos *JPG, TIFF, BMP* e, tal como as suas características (palheta de cor, tamanho) podem sofre alterações causadas pela mudança de um formato para outro. Já as imagens que não pertencem a nenhuma das classes ilustradas na Figura 4.1 serão classificadas como “Não Sei”. Por questões contratuais (UFPE – Hewlett-Packard) a base utilizada neste estudo não pode ser incluída no DVD anexo a esta dissertação.

A classe de agrupamento de fotografias abrangeu vários tipos de fotos (pessoas, paisagens, objetos e documentos). A maioria das fotografias está no formato *true color* embora houvesse imagens em tons de cinza. Essas fotos também variam de resolução desde imagens VGA (480x640 *pixels*) a 7.2 *Mpixels* (3072x2304). Essas imagens foram coletadas a partir de álbuns de família e obtidas na Internet.

No Capítulo 1 desta dissertação foi visto que existe um aumento expressivo no uso de câmeras digitais portáteis para digitalização de documentos, tais imagens foram incluídas neste estudo, trazendo um nível extra de dificuldade: por exemplo, um documento adquirido com uma câmera é classificado como uma fotografia ou um documento? A resposta a esta questão não é simples e pode ser um enigma até mesmo para um observador desapercebido. O critério adotado neste estudo foi: se a imagem englobar apenas o documento ela é classificada como "documento", se partes do entorno estão incluídos, é classificado como "foto". Pode-se notar essa diferenciação ao observar as imagens do documento ilustradas na parte direita das Figuras 4.1 e 4.3. No primeiro caso a imagem foi processada pelo PhotoDoc [51] enquanto no segundo não existiu tal processamento, classificadas por “documento” e “foto”, respectivamente.



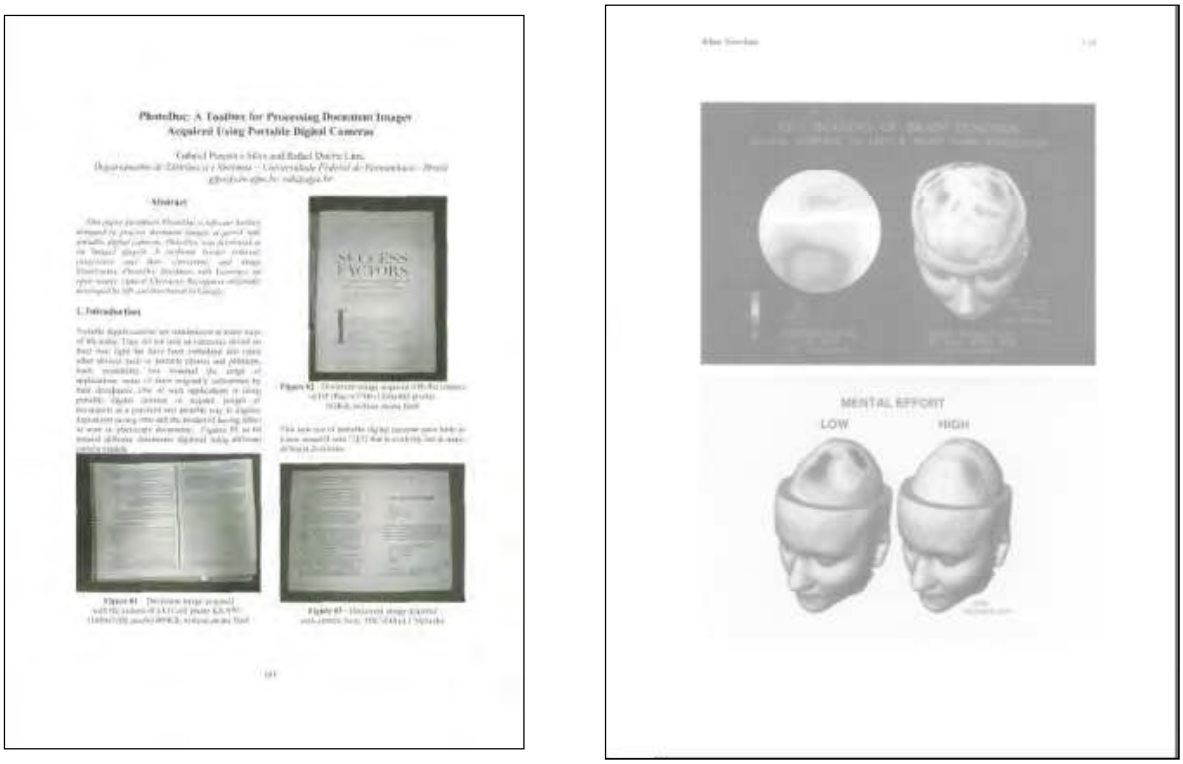
Pessoas Paisagens Documentos

Figura 4.3 - Exemplos de fotos (banco de dados).

Os 3.051 logotipos que existentes na base de dados foram coletados da Internet de várias fontes diferentes. Logotipos tendem a exibir uma paleta com um pequeno número de cores, embora muitas vezes eles apresentam-se no formato de arquivo *JPEG*, introduzindo artefatos inicialmente não existentes.

Figura 4.4 - Exemplos de logotipos (banco de dados).

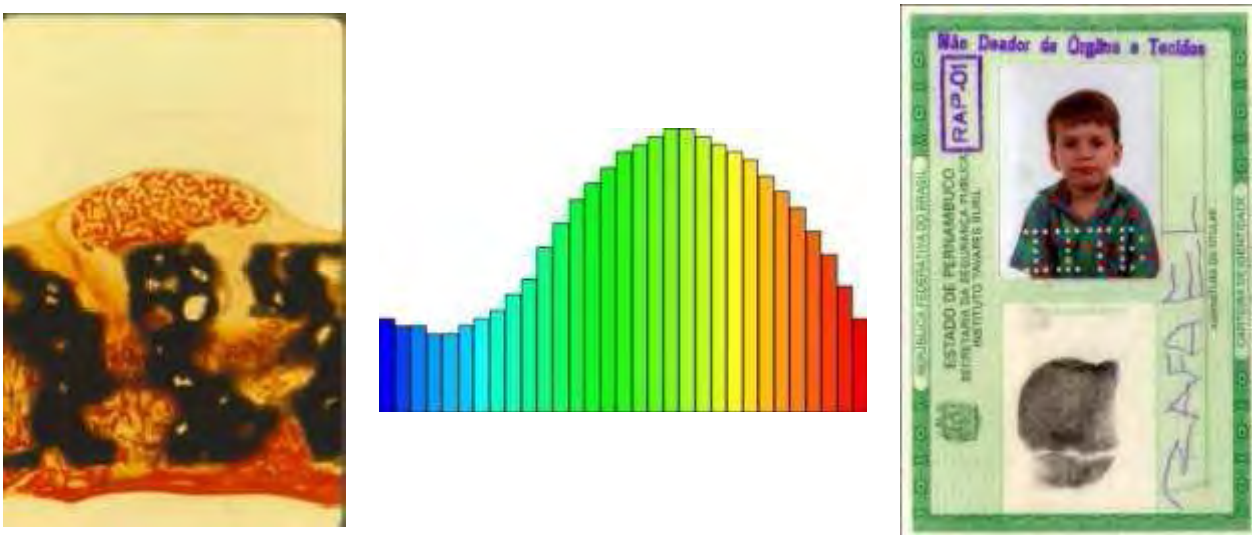
Já o conjunto de imagens de documentos é formado por 3.856 imagens de documentos adquiridos de diversas maneiras, quinhentos documentos foram fotografados com uma câmera digital a resoluções entre 3.2 e 7.2 *Mpixels*, com e sem apoio mecânico, com e sem flash e em seguida processados com o PhotoDoc [51], que corrigiu as distorções geométricas e eliminou a área externa ao documento. Outra porção desses documentos foi digitalizado por meio de *scanner* com diferentes resoluções (100-300 dpi) e salvo em *BMP*, *TIFF*, *JPEG* que embora não seja adequado para este tipo de imagem é muitas vezes utilizado por pessoas em geral [88]. Outra parte desses documentos foi obtida de imagens geradas a partir de arquivos *pdf*, com o auxílio do Adobe Acrobat 8.0 [81]. A Figura 4.5 ilustra alguns exemplos de documento de imagens utilizados neste trabalho.



PhotoDoc Escaneada

Figura 4.5 - Exemplos de documentos (banco de dados).

O último agrupamento de imagens é o conjunto de imagens que não pertence a nenhum dos três conjuntos de interesse ou possui uma característica híbrida que deixou dúvida quanto a qual grupo ela pertenceria. Estas imagens foram incluídas para aumentar a possibilidade de erros de classificação (para testar os classificadores mais extensivamente). Esta gama de imagens é formada a partir de imagens vetoriais, biomédica, gráficas (obtidos por software como o Excel ®, Powerpoint ®, etc.) e híbridas (imagens onde existem mais de uma classe de interesse). A Figura 4.6 ilustra exemplos dessas imagens.



Biomédicas Vetor gráfico Híbrida

Figura 4.6 - Exemplos de imagens classificadas como "Não Sei" (banco de dados).

A Tabela 4.1 mostra o número de imagens por arquivo no formato o ensaio definido.

	JPG	TIFF	BMP	Total
Foto	7.476	35	457	**7.968**
Logo	2.984	0	67	**3.051**
Documento	3.048	808	0	**3.856**
Não Sabe	202	0	327	**529**
Total:	**13.710**	**843**	**851**	**15.404**

Tabela 4.1- Imagens do conjunto de testes separadas por formato de arquivo.

4.2.1 Características Testadas

A escolha das características é uma importante etapa na construção de mecanismos de classificação/segmentação de informação. A escolha de características não representativas degrada consideravelmente o desempenho dos classificadores. O estudo da entropia da imagem é freqüentemente utilizado para a classificação dessas [89], porém possuem um alto custo computacional. O cálculo da entropia exige uma varredura na imagem para calcular a freqüência relativa de uma determinada cor, por exemplo, o que implica em varias divisões e operações logarítmicas.

O classificador desenvolvido nesta pesquisa é baseado na classificação binária apresentado em [89]. Nesse algoritmo assume-se uma distribuição gaussiana para cada uma das características, o que degrada o seu desempenho em relação a dados que possuem distribuições não-gaussianas. Já o novo classificador assume que a escala decrescente de uma imagem, analisadas juntamente com a sua escala de cinza e monocromática fornece elementos suficientes para uma rápida e eficiente classificação da imagem. Doze características foram testadas:

- Altura;
- Largura;
- Palheta (*true-color e grayscale*);
- *Gamute*;
- Conversão para escala de cinza (se RGB);
- *Gamute* em escala de cinza (se RGB);
- Binarização (Otsu);
- Número de pixels pretos na imagem binária;
- (Número de pixels pretos /área da imagem)*100%;
- (*Gamute*/Palheta)*100% (*true-color e grayscale*).

Essas características foram escolhidas devido ao seu baixo custo computacional e por apresentarem uma distribuição bem definida entre as classes de interesse, como por exemplo, em documentos, onde a relação pixels pretos e a área do documento varia entre 3% a 5%, a Tabela 4.2 apresenta uma comparação entre os tempos de extração das características propostas neste trabalho e a baseada em entropia apresentada em [89].

Ainda na Tabela 4.2 foi inserido o conceito de sub-amostragem das imagens, que consiste em reduzir o espaço de busca e extração, levando em conta as dimensões da imagem, todas essas foram submetidas à extração de características com (Sim) e sem (Não) sub-amostragem. O mecanismo ocorre da seguinte maneira:

size = height*width

- If size ≤ 300,000 break;
- If 300,000< size ≤ 500,000:
  remove even line or column (whatever the larger);
- If 500,000 < size ≤ 700,000:
  remove even lines and columns;
- If 700,000 < size ≤ 900,000:
  remove 2 lines in every 3 lines and even columns, (if height>width)
  remove even lines and 2 columns in every 3 columns, otherwise;
- If 900,000 < size remove 2 lines and 2 columns in every 3 lines/columns.

Sub-amostragem de imagens

Por fim, as características são extraídas para cada uma das imagens e formam um vetor de doze posições, este será o vetor de características de entrada para o classificador.

Extrator de características	Sub-Amostagem	Tempo de Extração (ms)	Linguagem
Proposta	Não	**0,52**	**C#**
Proposta	Sim	**0,192**	**C#**
Entropia [89]	Não	**1,4576**	**C#**
Entropia [89]	Sim	**0,497**	**C#**

Tabela 4.2- Tempo de extração de características.

4.2.2 Conjuntos de Treinamento

O conjunto de treinamento foi cuidadosamente selecionado para garantir uma boa representatividade no espaço de busca, ou seja, a interseção entre as três curvas de distribuição gaussiana deve ser a menor possível. Para tal optou-se por usar cerca de 10% das imagens de cada classe como instâncias de treinamento [68]. A Tabela 4.3 descreve a formação dos conjuntos de teste e treinamento.

Classes	**Teste**	**Treinamento**	**%**
Foto	**7968**	**668**	**8,34**
Logo	**3051**	**412**	**10,22**
Documento	**3856**	**276**	**4,70**
Não Sabe	**529**	**0**	**0**
Total:	**15404**	**1356**	**8,80**

Tabela 4.3 - Imagens do conjunto de testes separadas por formato de arquivo.

4.3 O classificar

O classificador foi desenvolvido inicialmente na versão Java sobre o código fonte do WEKA [90] em sua versão final ele encontra-se na linguagem C++. Esse algoritmo de classificação tomou como base o algoritmo “RandomForest” [87], cujas árvores sofreram uma pequena modificação. Ele utiliza treinamento supervisionado, onde as instâncias de treinamentos (imagens) comentadas na subseção 3.2.2 foram previamente rotuladas. O algoritmo “RandomForest” [87] permite “n” saídas enquanto o algoritmo proposto fornece uma saída binária. Optou-se pela escolha de um classificador binário por questões de simplicidade, é mais fácil modelar e visualizar o espaço de busca bidimensional em relação a um espaço n-ário.

Para a tarefa de classificar uma imagem em um dos três tipos de classes de interesse foi necessário usar três classificadores em cascata que podem ser visto como um classificador de três camadas, onde a primeira classifica em (Fotos/Não Foto) a segunda em (Logo/Não Logo) e a última em (Documento/Não Sei). O arranjo dos classificadores foi definido experimentalmente, onde

diversas combinações foram testadas para um conjunto de teste inicial e escolheu-se o melhor resultado de classificação (menor erro global). A Figura 4.7 ilustra a disposição de cada camada do classificador.

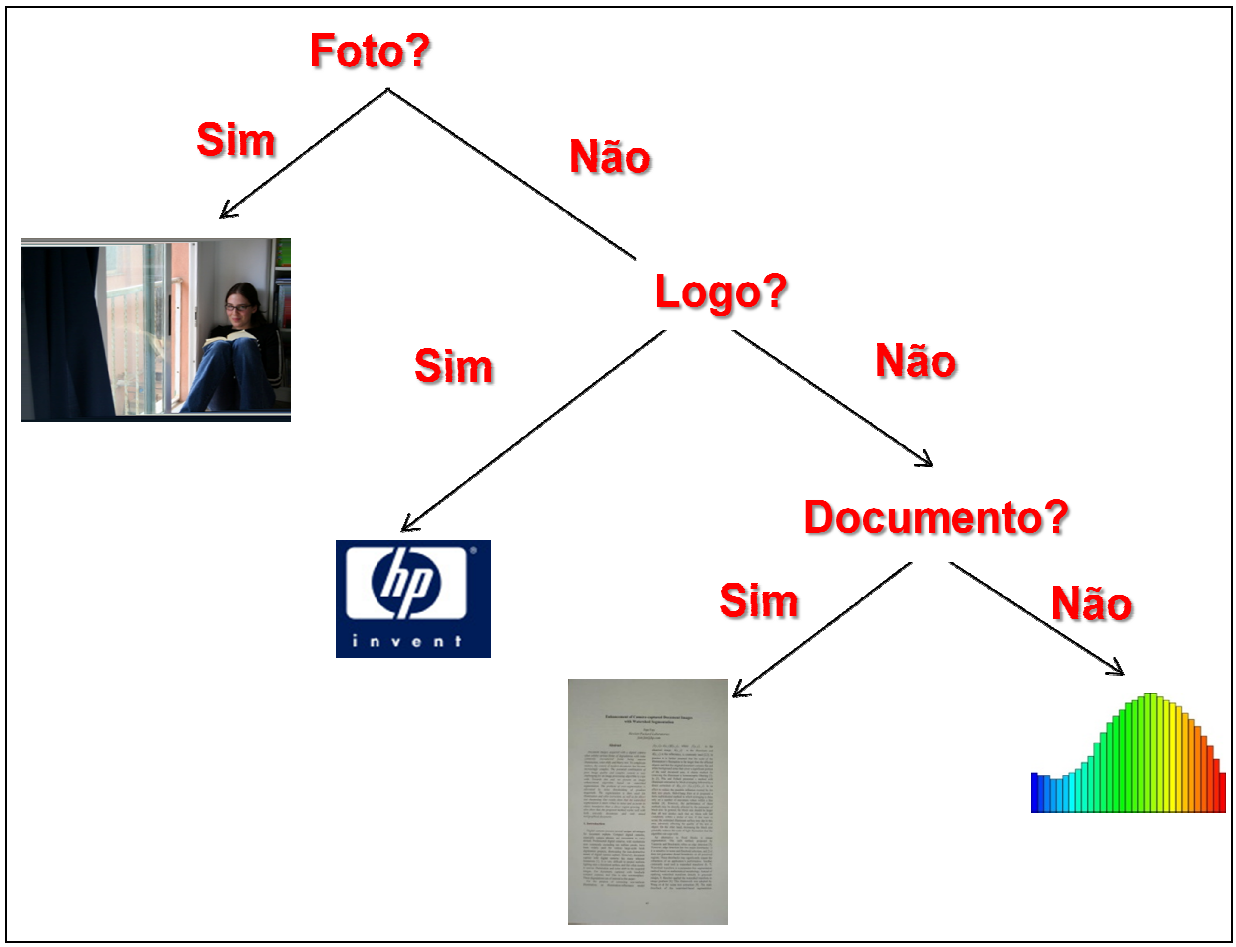


Figura 4.7 - Arranjo em cascata de classificadores.

Outro ponto importante desse novo classificador é o seu tempo de execução que é composto por: (tempo de extração) + (tempo de classificação). A média desse tempo fica abaixo de 400 ms. A Tabela 4.4 apresenta os tempos de classificação desse algoritmo.

Algoritmos	**Tempo de classificação**	**Linguagem**
Entropia [7]	**6,16**	**C#**
Proposto	**0,12**	**C#**

Tabela 4.4 - Tempo de classificação.

Ainda existe a possibilidade “expandir” este algoritmo, basta anexar um novo classificador em uma de suas folhar terminais, um exemplo de “expansão” é ilustrada na Figura 4.8.

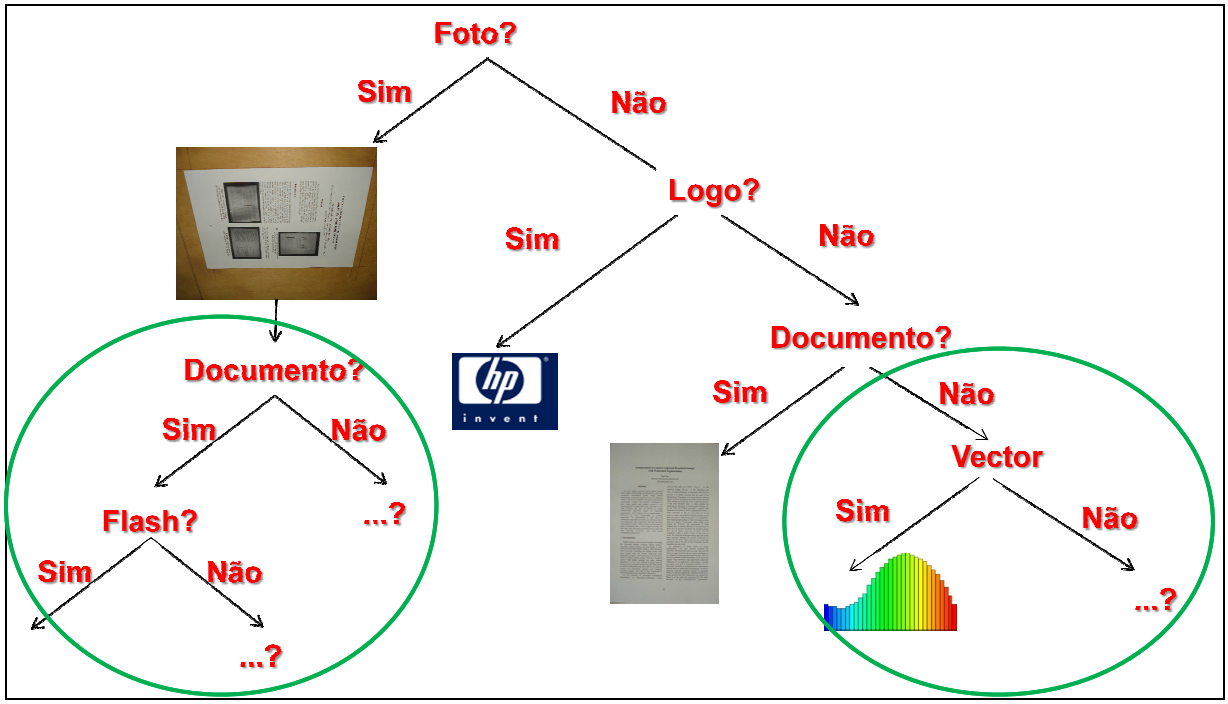


Figura 4.8 - Expansão do classificador.

4.3 Resultados

Esta subseção apresenta os resultados da classificação das imagens da base dados, esta foi descrita na seção 4.2 deste Capítulo, comparando o classificador [89] e o desenvolvido pela UFPE. Por fim, serão apresentados os dados referentes à classificação de base de dados anteriormente mencionada.

Antes de apresentar os resultados algumas informações devem ser apontadas:

- Ambos classificadores foram submetidos ao mesmo conjunto de teste e treinamento dessa forma buscando os modelos que obtiveram melhores resultados, após cento e cinqüenta rodadas para cada um dos classificadores;
- O código do classificador “Entropia [89]” foi fornecido pela Hewlett-Packard e reusado para criar o arranjo em cascata, código se encontra na linguagem C#;
- O classificador “Proposto [80]” foi desenvolvido pela equipe da UFPE e a versão utilizada neste experimento está na linguagem C#;
- Os dois conjuntos de características usadas separadamente por ambos classificadores estão descritas nas referências [89][80];
- Toda base de dados foi rotulada manualmente (verificada duas vezes) antes da realização dos experimentos que serão apresentados a seguir (Matrizes de confusão).

As tabelas a seguir apresentam uma comparação entre o resultado do algoritmo descrito em [89] e o desenvolvido pelo grupo de engenharia de documentos da UFPE. Ainda são demonstrados resultados de uma “expansão” do classificador proposto [79]. As Tabelas 4.5 a 4.12 apresentam os resultados obtidos com o arranjo de classificadores apresentado em [80], ou seja, a classificação das

imagens em três classes Foto, Logo e Documento. As Tabelas 4.13 a 4.24 apresentam os resultados obtidos em [79] onde se busca classificar as imagens de documentos quanto ao dispositivo de digitalização.

Paras as Tabelas 4.5 a 4.8 foram utilizadas as características desenvolvidas na UFPE [80], que formam os vetores de entrada de treinamento e teste dos classificadores "Entropia [89]" e "Proposto [80]", com o uso ou não da técnica de sub-amostragem.

Classificador Entropia[89]	**Foto**	**Logo**	**Documento**	**Não Sabe**	**Acertos**
Foto	7021	593	159	195	**0,881**
Logo	566	1684	203	598	**0,551**
Documento	374	408	2945	129	**0,763**
Não Sabe	62	174	95	198	**0,374**
				Taxa de Acerto Global	**0,769**

Tabela 4.5 - Matriz de confusão com o uso de sub-amostragem.

Classificador Entropia[89]	**Foto**	**Logo**	**Documento**	**Não Sabe**	**Acertos**
Foto	6957	593	187	231	**0,873**
Logo	564	1685	203	599	**0,552**
Documento	389	408	2912	147	**0,755**
Não Sabe	62	191	95	181	**0,342**
				Taxa de Acerto Global	**0,761**

Tabela 4.6 - Matriz de confusão sem o uso de sub-amostragem.

Classificador Proposto [80]	**Foto**	**Logo**	**Documento**	**Não Sabe**	**Acertos**
Foto	7740	164	34	30	**0,971**
Logo	258	2761	11	21	**0,904**
Documento	93	41	3722	0	**0,965**
Não Sabe	110	300	30	89	**0,343**
				Taxa de Acerto Global	**0,929**

Tabela 4.7 - Matriz de confusão com o uso de sub-amostragem.

Classificador Proposto [80]	**Foto**	**Logo**	**Documento**	**Não Sabe**	**Acertos**
Foto	7554	363	14	37	**0,948**
Logo	282	2730	23	16	**0,894**
Documento	277	266	3314	0	**0,859**
Não Sabe	151	309	17	52	**0,098**
				Taxa de Acerto Global	**0,886**

Tabela 4.8 - Matriz de confusão sem o uso de sub-amostragem.

As Tabelas 4.9 a 4.12 utilizam as características apresentadas em [89] para formar os vetores de treinamento e teste, para ambos classificadores. Ainda é possível verificar os resultados da classificação com o uso ou não da técnica de sub-amostragem.

Classificador Entropia[89]	**Foto**	**Logo**	**Documento**	**Não Sabe**	**Acertos**
Foto	7603	275	18	72	**0,954**
Logo	385	1929	135	602	**0,632**
Documento	311	373	3167	5	**0,821**
Não Sabe	77	174	128	150	**0,283**
				Taxa de Acerto Global	**0,834**
Tabela 4.9 - Matriz de confusão com o uso de sub-amostragem.					

Classificador Entropia[89]	**Foto**	**Logo**	**Documento**	**Não Sabe**	**Acertos**
Foto	7280	620	12	56	**0,914**
Logo	429	2104	96	422	**0,690**
Documento	206	351	3299	0	**0,856**
Não Sabe	70	225	0	234	**0,442**
				Taxa de Acerto Global	**0,838**
Tabela 4.10 - Matriz de confusão sem o uso de sub-amostragem.					

Classificador Proposto [80]	**Foto**	**Logo**	**Documento**	**Não Sabe**	**Acertos**
Foto	7062	576	143	187	**0,886**
Logo	551	1704	201	595	**0,558**
Documento	348	407	2989	112	**0,775**
Não Sabe	66	192	95	176	**0,332**
				Taxa de Acerto Global	**0,774**
Tabela 4.11 - Matriz de confusão com o uso de sub-amostragem.					

Classificador Proposto [80]	**Foto**	**Logo**	**Documento**	**Não Sabe**	**Acertos**
Foto	7282	620	12	54	**0,913**
Logo	427	2112	93	419	**0,692**
Documento	206	348	3302	0	**0,856**
Não Sabe	73	232	38	186	**0,351**
				Taxa de Acerto Global	**0,836**
Tabela 4.12 - Matriz de confusão sem o uso de sub-amostragem.					

A seguir será apresentado o desempenho dos classificadores submetidos ao contexto de qual dispositivo foi utilizado para a digitalização de dado documento, trata-se de uma modificação do classificador apresentado em [80]. As Tabelas 4.13 a 4.16 apresentam os resultados do uso da extração de características desenvolvidos pela equipe da UFPE [79].

Classificador Entropia[89]	**Foto**	**Escaneado**	**Acertos**
Foto	8332	1232	**0,871**
Escaneado	138	6306	**0,978**
		Taxa de Acerto Global	**0,914**
Tabela 4.13 - Matriz de confusão com o uso de sub-amostragem.			

Classificador Entropia[89]	**Foto**	**Escaneado**	**Acertos**
Foto	8334	1230	**0,871**
Escaneado	135	6309	**0,979**
		Taxa de Acerto Global	**0,914**

Tabela 4.14 - Matriz de confusão sem o uso de sub-amostragem.

Classificador Proposto [79]	**Foto**	**Escaneado**	**Acertos**
Foto	9569	4	**0,999**
Escaneado	0	6444	**1**
		Taxa de Acerto Global	**0,999**

Tabela 4.15 - Matriz de confusão com o uso de sub-amostragem.

Classificador Proposto [79]	**Foto**	**Escaneado**	**Acertos**
Foto	9570	3	**0,999**
Escaneado	0	6444	**1**
		Taxa de Acerto Global	**0,999**

Tabela 4.16 - Matriz de confusão sem o uso de sub-amostragem.

As Tabelas 4.17 a 4.20 apresentam os resultados do uso da extração de características [89].

Classificador Entropia[89]	**Foto**	**Escaneado**	**Acertos**
Foto	8394	1170	**0,876**
Escaneado	6	6438	**0,999**
		Taxa de Acerto Global	**0,926**

Tabela 4.17 - Matriz de confusão com o uso de sub-amostragem.

Classificador Entropia[89]	**Foto**	**Escaneado**	**Acertos**
Foto	8423	1150	**0,879**
Escaneado	4	6440	**0,999**
		Taxa de Acerto Global	**0,927**

Tabela 4.18 - Matriz de confusão sem o uso de sub-amostragem.

Classificador Proposto [79]	**Foto**	**Escaneado**	**Acertos**
Foto	8581	983	**0,897**
Escaneado	32	6412	**0,995**
		Taxa de Acerto Global	**0,936**

Tabela 4.19 - Matriz de confusão com o uso de sub-amostragem.

Classificador Proposto [79]	**Foto**	**Escaneado**	**Acertos**
Foto	8581	983	**0,897**
Escaneado	32	6412	**0,995**
		Taxa de Acerto Global	**0,936**
Tabela 4.20 - Matriz de confusão sem o uso de sub-amostragem.			

Determinar o uso ou não do dispositivo de *flash* é fundamental para uma melhor filtragem dos documentos adquiridos por meio de câmeras digitais, uma vez que o impacto da iluminação sobre os algoritmos é um fator muitas vezes determinante para qualidade da filtragem.

A seguir é apresentado um desdobramento "expansão" do classificador. Ele consiste em classificar os documentos classificados como fotos quanto ao uso ou não do dispositivo de *flash*, ver Tabelas 4.21 a 4.24. As duas primeiras Tabelas fazem uso das características descritas em [89] enquanto as duas últimas Tabelas usam as características apresentadas em [79].

Classificador Entropia[89]	**+Flash**	**-Flash**	**Escaneado**	**Acertos**
+Flash	3402	270	357	**0,844**
-Flash	69	4562	913	**0,822**
Escaneado	24	158	6262	**0,971**
			Taxa de Acerto Global	**0,888**
Tabela 4.21 - Matriz de confusão com o uso de sub-amostragem.				

Classificador Entropia[89]	**+Flash**	**-Flash**	**Escaneado**	**Acertos**
+Flash	3402	272	355	**0,844**
-Flash	71	4466	1007	**0,805**
Escaneado	32	152	6260	**0,971**
			Taxa de Acerto Global	**0,882**
Tabela 4.22 - Matriz de confusão sem o uso de sub-amostragem.				

Classificador Proposto [79]	**+Flash**	**-Flash**	**Escaneado**	**Acertos**
+Flash	4029	0	0	**1**
-Flash	0	5540	4	**0,999**
Escaneado	0	0	6444	**1**
			Taxa de Acerto Global	**0,999**
Tabela 4.23 - Matriz de confusão com o uso de sub-amostragem.				

Classificador Proposto [79]	**+Flash**	**-Flash**	**Escaneado**	**Acertos**
+Flash	4029	0	0	**1**
-Flash	4	5537	3	**0,998**
Escaneado	0	0	6,444	**1**
			Taxa de Acerto Global	**0,999**
Tabela 4.24 - Matriz de confusão sem o uso de sub-amostragem.				

Os experimentos foram realizados tomando como base o software *open souce* Weka [90] que demonstrou ser uma excelente plataforma de ensaio para análise estatística. A escolha de uma árvore de classificador foi tomada após a realização de vários experimentos com o grande número de alternativas oferecidas pelo Weka, embora os resultados não variassem muito. Entre eles, uma comparação entre o novo classificador proposto aqui e um classificador MLP (Multi Layer Perceptron) [68] apresentou resultados (91,37% Fotos, Logos 85,48%, 94,54% e Documentos) o classificador SVM (Support Vector Machine) [68] apresentou resultados muito próximos a esses, ou seja, inferiores ao do algoritmo proposto.

A escolha das imagens na formação do conjunto é de importância primordial para o desempenho do classificador. Outro ponto importante, observado com o uso da sub-amostragem, é que a qualidade das imagens tem-se revelado mais importante do que tamanho, ou seja, a sub-amostragem em imagens de "boa" qualidade não representa perdas expressivas aos classificadores. No entanto imagens muito pequenas parecem representar um maior grau de dificuldade para a classificação, uma vez que foram mais freqüentemente erros de classificação sobre essas imagens.

Observando-se os resultados apresentados pelas Tabelas 4.5 a 4.24 foi possível constatar o bom desempenho do sistema de classificação (classificador+características) proposto pela equipe da UFPE. Esse novo sistema diminuiu a taxa de erro global e trouxe ganhos muito superiores de classificação para as classes (Logo e Documento). Já a tempo de extração de característica foi reduzido por um fator de dez, 1458 para 147 mseg (média de processamento por imagem) usando-se [89] e [80], respectivamente.

O aumento da precisão e a diminuição do tempo de execução permitem que esse classificador possa ser embarcado em impressoras ou câmeras fotográficas digitais. Ainda foi demonstrada a facilidade de expansão do classificador para tratar novas classes ou até mesmo refinar classes já existentes.

4.4 Análise das imagens incorretamente classificadas

Esta subseção apresenta a análise das imagens incorretamente classificadas no experimento apresentado anteriormente. Uma vez que toda base utilizada neste experimento está rotulada foi possível estabelecer algumas características comuns a respeito dessas imagens.

A análise é referente aos resultados apresentados na Tabela 4.7, onde foi utilizado o classificador e características descritos em [80].

4.4.1 Fotos classificadas como Logo

Trata-se de 164 imagens onde foi observada pouca variação de cores, ou seja, possuem *Gamute* pobre. Essas imagens geralmente são de ambientes internos onde prevalecem paredes ou no caso de ambientes externos focam em placas, murais ou *outdoors*. Também foi observado em imagens de objetos que ocupam a maior parte da cena e que possuem pequenas variações de cores. A

Figura 4.8 apresenta alguns exemplos de imagens da classe Foto que foram incorretamente classificadas como Logo.

Ambiente Interno (Quadros) Ambiente Externo (Placa) Objetos

Figura 4.9 – Exemplos de fotos classificadas como logo.

4.4.2 Fotos classificadas como Documento

São 34 fotos de documentos que foram classificadas como documentos, sua análise indicou a presença de uma pequena parcela de borda. Essa borda é caracterizada por possuir *pixels* cujos valores se aproximam muito do valor dos *pixels* do papel. A Figura 4.9 ilustra um dessas imagens incorretamente classificada.

Figura 4.10 – Exemplos de fotos classificadas como documento.

4.4.3 Logos classificados como Foto

Já no caso das imagens de logos classificados como foto, o classificador apresentou um total de 258 erros. Todas as imagens incorretamente classificadas possuem uma característica em comum, trata-se de logos onde há uma textura de fundo complexa. A Figura 4.10 apresenta alguns dessas imagens de logos incorretamente classificados.

Figura 4.11 – Exemplos de logos classificados como foto.

4.4.4 Logos classificados como Documento

Resumem-se a apenas 11 imagens de logos que foram incorretamente classificadas como documento. Todas essas imagens são de logos cuja resolução é de 600 dpis e que possuem menos de 10% de sua área ocupada por texto. São caracterizados por *banners de* campanhas publicitárias.

4.4.5 Documentos classificados como Foto

Trata-se de 93 imagens de documentos digitalizados por meio *scanner* de mesa, para mais detalhes consultar o Capítulo 2 desta dissertação.

Pode-se dividir o montante de imagens de documentos classificadas como foto em dois grupos:

- Em 58 dessas imagens foram observados inserções de ruídos de digitalização sobre tudo o ruído de *warp*. São documentos oriundos de livros grossos que impedem que as páginas fiquem na posição ideal "paralela ao suporte do *scanner*" para serem digitalizadas.
- As 35 imagens restantes são imagens de documentos fotografados sem o uso de flash processadas pelo PhotoDoc [51]. Foram imagens adquiridas por câmeras de baixa resolução (1.3 *Mpixels*) e que não foram submetidas ao processo de normalização de iluminação.

A Figura 4.11 ilustra imagens de documentos que foram incorretamente classificadas como fotos.

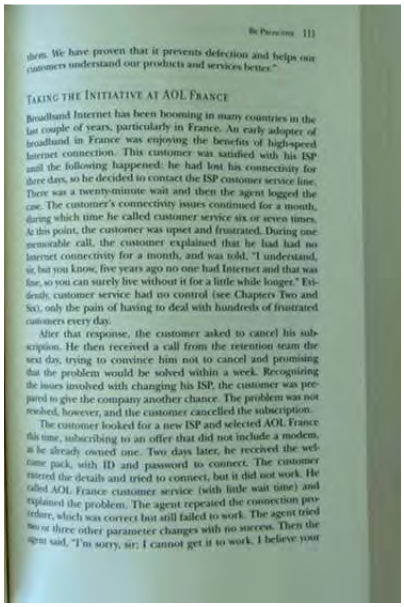


Imagem com ruído de *warp*

Imagem adquirida com câmera digital de 1.3 *Mpixels* sem o uso de *flash*

Figura 4.12 – Exemplos de documentos classificados como foto.

4.4.6 Documentos classificados como Logo

São 41 imagens de documentos geradas com o auxílio ferramenta Adobe Acrobat Pro 8.0 [81] a partir de arquivos *pdf* de *proceedings* de conferências científicas. Trata-se de 26 páginas onde existe apenas a presença do logo das entidades promotora e patrocinadoras do evento.

As 15 imagens de documentos restantes são provenientes de *pdf* onde há designer artístico inserido.

Logos

Designer artístico

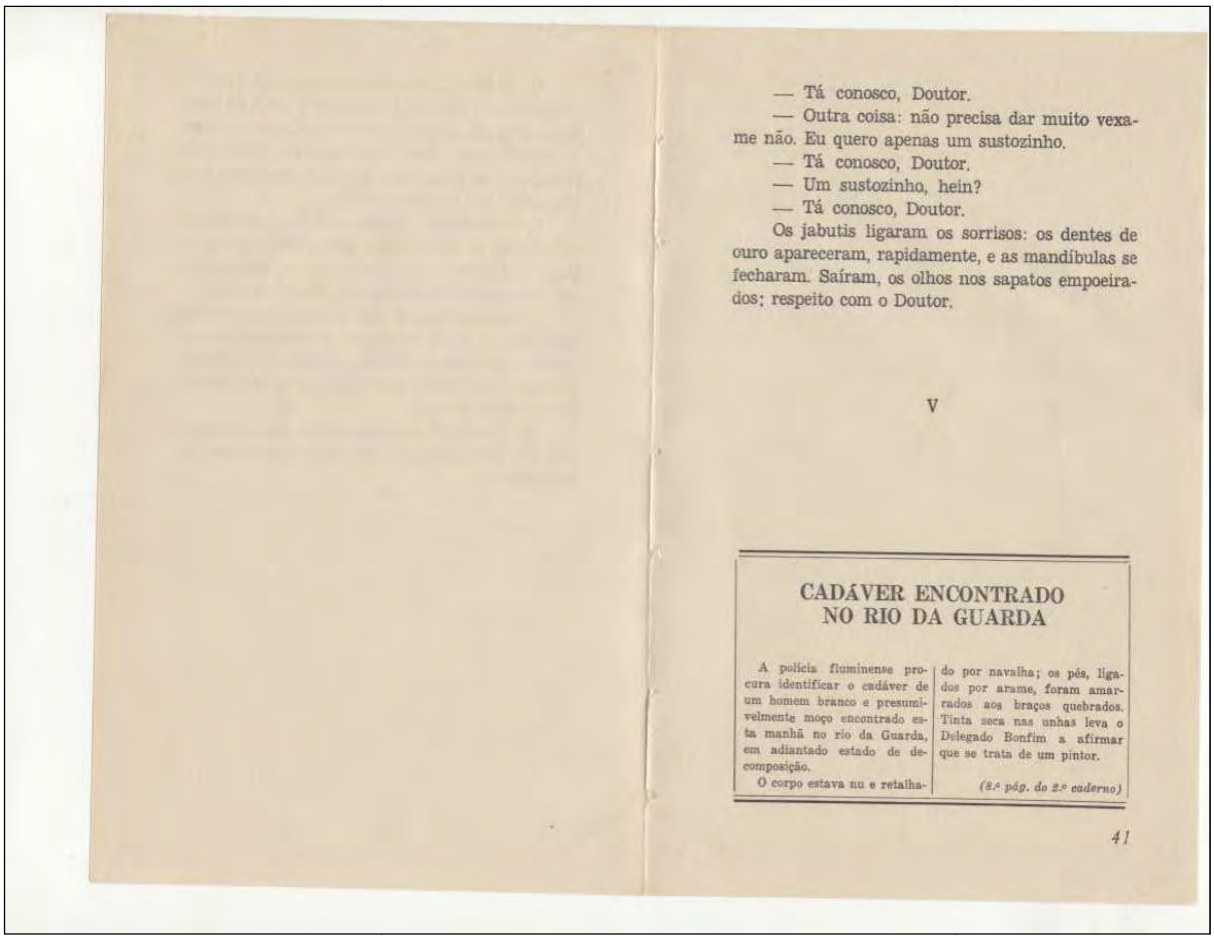
Figura 4.13 – Exemplos de documentos classificados como logo.

Capítulo 5
Detecção de Bordas

Neste capítulo será tratado o problema de detecção de bordas de documentos fotografados. Uma breve introdução sobre técnicas clássicas para detectar borda em imagens (geralmente em tons de cinza) e outros algoritmos que tratam imagens em *true color*, esses últimos forneceram os pontos de controle, para a correção de perspectiva e recorte da imagem, para o ambiente PhotoDoc.

5.1 Bordas

Bordas de documentos são definidas como os contornos que delimitam o conteúdo e a região externa dos documentos, ainda podem ser consideradas áreas que circundam a imagem do documento e que não representam informação relevante. Em todo caso, o aparecimento de bordas não é um fenômeno restrito apenas à digitalização de documentos por câmeras digitais. Há o surgimento desse fenômeno em imagens digitalizadas por *scanners*, tais bordas correspondem à área da bandeja do escaner não ocupada pelo documento. Pode-se observar esse efeito na Figura 5.1, a borda apresenta uma coloração clara e circunda o documento.


— Tá conosco, Doutor.
— Outra coisa: não precisa dar muito vexame não. Eu quero apenas um sustozinho.
— Tá conosco, Doutor.
— Um sustozinho, hein?
— Tá conosco, Doutor.
Os jabutis ligaram os sorrisos: os dentes de ouro apareceram, rapidamente, e as mandíbulas se fecharam. Saíram, os olhos nos sapatos empoeirados; respeito com o Doutor.

V

CADÁVER ENCONTRADO NO RIO DA GUARDA

A polícia fluminense procura identificar o cadáver de um homem branco e presumivelmente moço encontrado esta manhã no rio da Guarda, em adiantado estado de decomposição.
O corpo estava nu e retalhado por navalha; os pés, ligados por arame, foram amarrados aos braços quebrados. Tinta seca nas unhas leva o Delegado Bonfim a afirmar que se trata de um pintor.
(8.ª pág. do 2.º caderno)

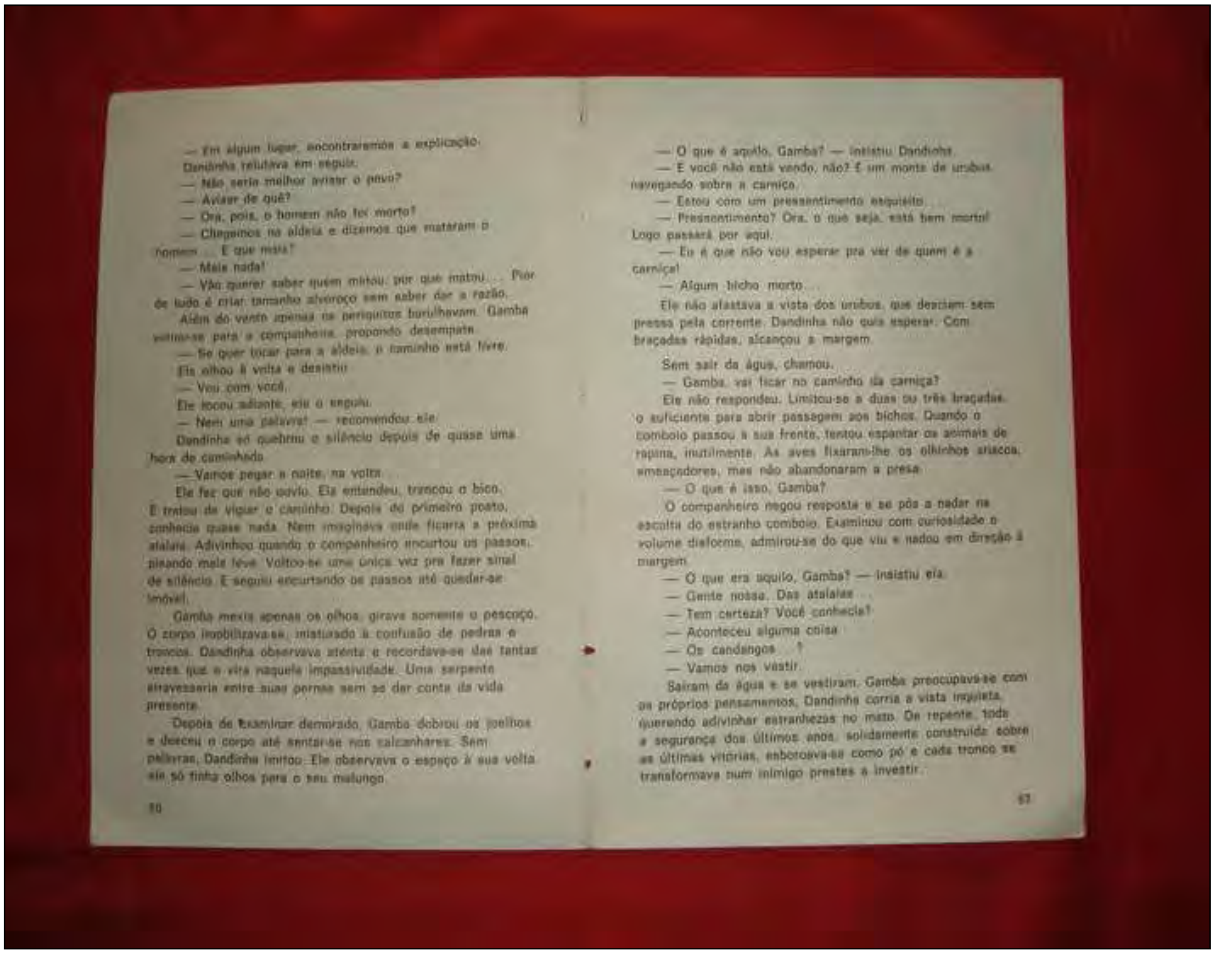
41

Figura 5.1 - Documento escaneado apresentando bordas.

A existência de bordas é um efeito indesejável, pois requer: espaço adicional para armazenamento em disco, maior ocupação da largura de banda na transmissão de imagens via canais de comunicação, reduz a resolução dos caracteres em dispositivos de saída (monitores, projetores, etc.), maior custo de impressão, dentre diversos outros problemas [12]. Um exemplo típico de como a presença de borda pode impor dificuldades aos algoritmos de processamento de imagens é ilustrado usando-se o algoritmo de binarização de Otsu [34]. O resultado final da aplicação desse algoritmo sobre a Figura 5.2 mostra o efeito negativo da presença da borda que fica evidente ao visualizar as Figura 5.3 e 5.4. Ainda é possível notar um aumento aparente na “resolução dos caracteres.

Já o efeito negativo exemplificado anteriormente é bastante comum em algoritmos que tenham como base a análise estatística da imagem, pois a borda irá adicionar informações estatísticas não relevantes à filtragem do documento. Como já foi discutido em capítulos anteriores, sabe-se que em documentos fotografados por câmeras digitais portáteis, o aparecimento de bordas é corriqueiro e por sua vez essas bordas tendem a serem variadas e complexas.

O uso de algoritmos onde a análise estatística é local, ou seja, realizada em pequenas partes das imagens também conhecidas como janelas de análise. Definir automaticamente o tamanho dessas janelas não é uma tarefa trivial e uma má escolha tende a degradar o texto próximo à extremidade dos documentos.



— Em algum lugar, encontraremos a explicação.
Dandinha relutava em seguir.
— Não seria melhor avisar o povo?
— Avisar de quê?
— Ora, pois, o homem não foi morto?
— Chegamos na aldeia e dizemos que mataram o homem ... E que mais?
— Mais nada!
— Vão querer saber quem matou, por que matou... Por de tudo é criar tamanho alvoroço sem saber dar a razão.
Além do vento apenas os periquitos burulhavam. Gamba voltou-se para o companheiro, propondo desempate.
— Se quer tocar para a aldeia, o caminho está livre.
Ela olhou à volta e desistiu.
— Vou com você.
Ele tocou adiante, ela o seguiu.
— Nem uma palavra! — recomendou ele.
Dandinha só quebrou o silêncio depois de quase uma hora de caminhada.
— Vamos pegar a noite, na volta.
Ele fez que não ouviu. Ela entendeu, trancou o bico. E tratou de vigiar o caminho. Depois do primeiro posto, conhecia quase nada. Nem imaginava onde ficaria a próxima atalaia. Adivinhou quando o companheiro encurtou os passos, pisando mais leve. Voltou-se uma única vez pra fazer sinal de silêncio. E seguiu encurtando os passos até quedar-se imóvel.
Gamba mexia apenas os olhos, girava somente o pescoço. O corpo imobilizava-se, misturado à confusão de pedras e troncos. Dandinha observava atenta e recordava-se das tantas vezes que o vira naquela impassividade. Uma serpente atravessaria entre suas pernas sem se dar conta da vida presente.
Depois de examinar demorado, Gamba dobrou os joelhos e desceu o corpo até sentar-se nos calcanhares. Sem palavras, Dandinha imitou. Ele observava o espaço à sua volta ele só tinha olhos para o seu mutungo.

50

— O que é aquilo, Gamba? — insistiu Dandinha.
— E você não está vendo, não? É um monte de urubus navegando sobre a carniça.
— Estou com um pressentimento esquisito...
— Pressentimento? Ora, o que seja, está bem morto! Logo passará por aqui.
— Eu é que não vou esperar pra ver de quem é a carniça!
— Algum bicho morto...
Ele não afastava a vista dos urubus, que desciam sem pressa pela corrente. Dandinha não quis esperar. Com braçadas rápidas, alcançou a margem.
Sem sair da água, chamou:
— Gamba, vai ficar no caminho da carniça?
Ele não respondeu. Limitou-se a duas ou três braçadas, o suficiente para abrir passagem aos bichos. Quando o comboio passou à sua frente, tentou espantar os animais de rapina, inutilmente. As aves fixaram-lhe os olhinhos sisudos, ameaçadores, mas não abandonaram a presa.
— O que é isso, Gamba?
O companheiro negou resposta e se pôs a nadar na escolta do estranho comboio. Examinou com curiosidade o volume disforme, admirou-se do que viu e nadou em direção à margem.
— O que era aquilo, Gamba? — insistiu ela.
— Gente nossa. Das atalaias...
— Tem certeza? Você conhecia?
— Aconteceu alguma coisa?
— Os candangos...?
— Vamos nos vestir.
Saíram da água e se vestiram. Gamba preocupava-se com os próprios pensamentos, Dandinha corria a vista inquieta, querendo adivinhar estranhezas no mato. De repente, toda a segurança dos últimos anos, solidamente construída sobre as últimas vitórias, esboroava-se como pó e cada tronco se transformava num inimigo prestes a investir.

51

Figura 5.2 - Documento fotografado com presença de bordas.

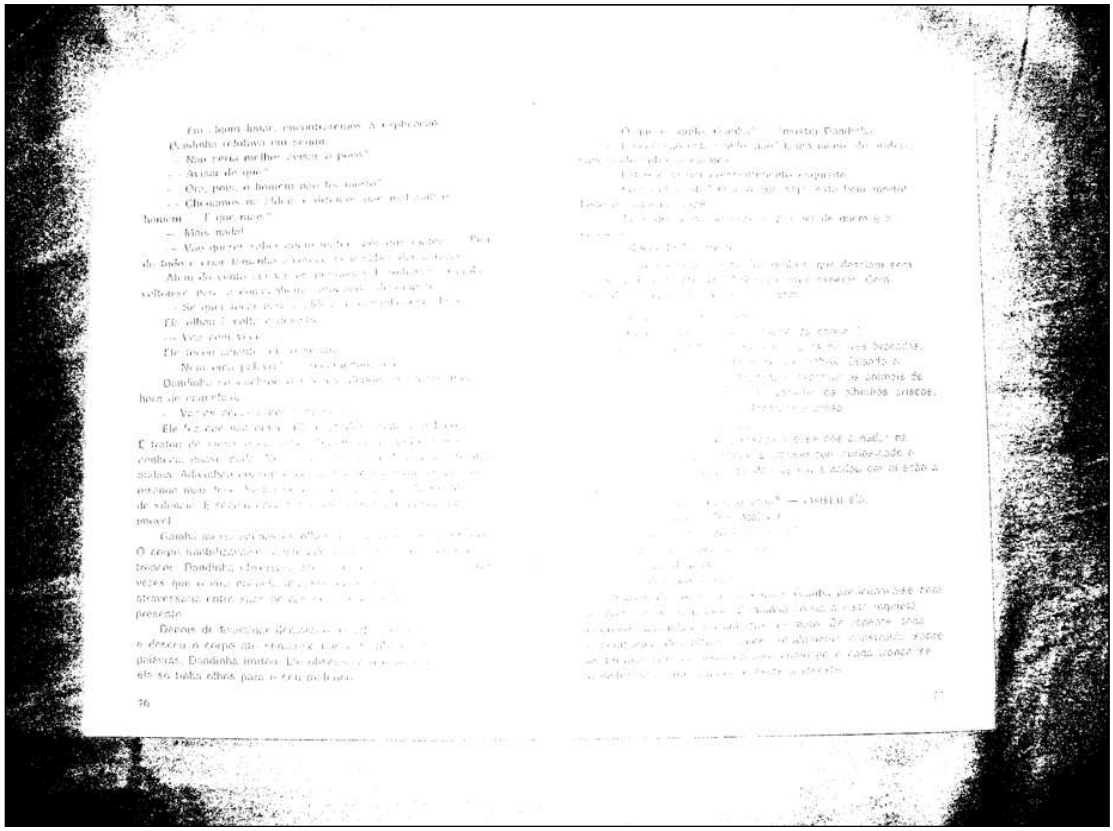
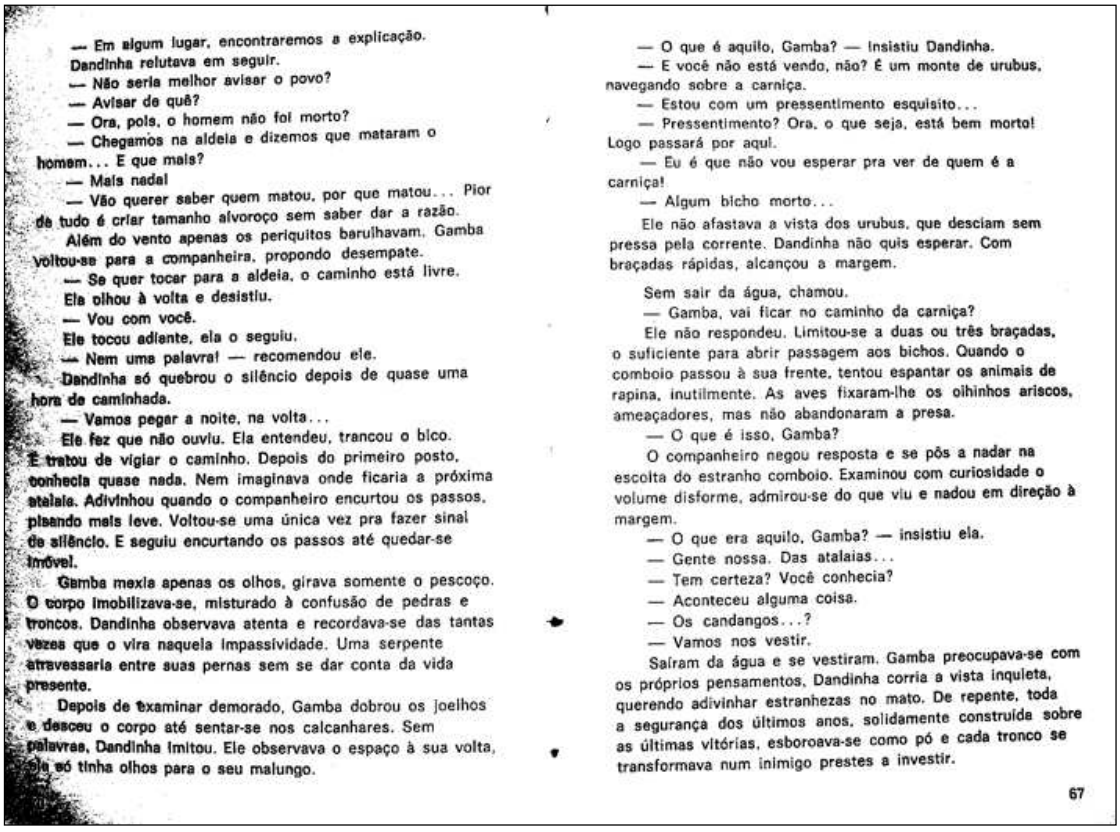


Figura 5.3 - Resultado da Binarização [34] da Figura 5.2.



— Em algum lugar, encontraremos a explicação.
Dandinha relutava em seguir.
— Não seria melhor avisar o povo?
— Avisar de quê?
— Ora, pois, o homem não foi morto?
— Chegamos na aldeia e dizemos que mataram o homem... E que mais?
— Mais nada!
— Vão querer saber quem matou, por que matou... Pior de tudo é criar tamanho alvoroço sem saber dar a razão.
Além do vento apenas os periquitos barulhavam. Gamba voltou-se para a companheira, propondo desempate.
— Se quer tocar para a aldeia, o caminho está livre.
Ela olhou à volta e desistiu.
— Vou com você.
Ele tocou adiante, ela o seguiu.
— Nem uma palavra! — recomendou ele.
Dandinha só quebrou o silêncio depois de quase uma hora de caminhada.
— Vamos pegar a noite, na volta...
Ele fez que não ouviu. Ela entendeu, trancou o bico. E tratou de vigiar o caminho. Depois do primeiro posto, conhecia quase nada. Nem imaginava onde ficaria a próxima atalaia. Adivinhou quando o companheiro encurtou os passos, pisando mais leve. Voltou-se uma única vez pra fazer sinal de silêncio. E seguiu encurtando os passos até quedar-se imóvel.
Gamba mexia apenas os olhos, girava somente o pescoço. O corpo imobilizava-se, misturado à confusão de pedras e troncos. Dandinha observava atenta e recordava-se das tantas vezes que o vira naquela impassividade. Uma serpente atravessaria entre suas pernas sem se dar conta da vida presente.
Depois de examinar demorado, Gamba dobrou os joelhos e desceu o corpo até sentar-se nos calcanhares. Sem palavras, Dandinha imitou. Ele observava o espaço à sua volta, ele só tinha olhos para o seu malungo.

— O que é aquilo, Gamba? — insistiu Dandinha.
— E você não está vendo, não? É um monte de urubus, navegando sobre a carniça.
— Estou com um pressentimento esquisito...
— Pressentimento? Ora, o que seja, está bem morto! Logo passará por aqui.
— Eu é que não vou esperar pra ver de quem é a carniça!
— Algum bicho morto...
Ele não afastava a vista dos urubus, que desciam sem pressa pela corrente. Dandinha não quis esperar. Com braçadas rápidas, alcançou a margem.
Sem sair da água, chamou.
— Gamba, vai ficar no caminho da carniça?
Ele não respondeu. Limitou-se a duas ou três braçadas, o suficiente para abrir passagem aos bichos. Quando o comboio passou à sua frente, tentou espantar os animais de rapina, inutilmente. As aves fixaram-lhe os olhinhos ariscos, ameaçadores, mas não abandonaram a presa.
— O que é isso, Gamba?
O companheiro negou resposta e se pôs a nadar na escolta do estranho comboio. Examinou com curiosidade o volume disforme, admirou-se do que viu e nadou em direção à margem.
— O que era aquilo, Gamba? — insistiu ela.
— Gente nossa. Das atalaias...
— Tem certeza? Você conhecia?
— Aconteceu alguma coisa.
— Os candangos...?
— Vamos nos vestir.
Saíram da água e se vestiram. Gamba preocupava-se com os próprios pensamentos, Dandinha corria a vista inquieta, querendo adivinhar estranhezas no mato. De repente, toda a segurança dos últimos anos, solidamente construída sobre as últimas vitórias, esboroava-se como pó e cada tronco se transformava num inimigo prestes a investir.

67

Figura 5.4 - Resultado da Binarização [34] após remoção das bordas da Figura 5.2.

5.2 Detecção de bordas

O problema detecção de bordas foi uns dos primeiros a ser tratado na área de processamento de imagens e ainda vem desafiando os pesquisadores da área. Sobre ele continuam sendo experimentadas novas técnicas, principalmente a detecção de bordas em cenas consideradas complexas, caso em que se enquadram as imagens adquiridas por meio de câmeras digitais portáteis.

Nos algoritmos clássicos, a idéia básica é realçar as áreas das imagens que apresentam mudanças abruptas de sinais, no caso de imagens coloridas esses sinais são as cores dos pixels. O tratamento deste problema em imagens fotografadas por câmeras digitais é relativamente novo, o uso de imagens de documentos digitalizados por *scanners* foi fundamental nesta pesquisa já que a mesmas encontra-se bem estudada [12][44], servindo como referência para o desenvolvimento de algoritmos em câmeras digitais.

A maioria das técnicas de detecção de bordas utiliza o mecanismo básico de definir um operador derivativo local de primeira ou segunda ordem, juntamente com alguma técnica de regularização para reduzir os efeitos de ruído, geralmente usa-se um filtro gaussiano passa-baixa. Os métodos de detecção de bordas, como por exemplo, detector de Söbel [60] e detector de Prewitt [13] são baseados no conceito de filtro derivativo espacial, onde os operadores gradiente local são usados para detectar bordas em certas orientações, que é ilustrada na Figura 5.5.

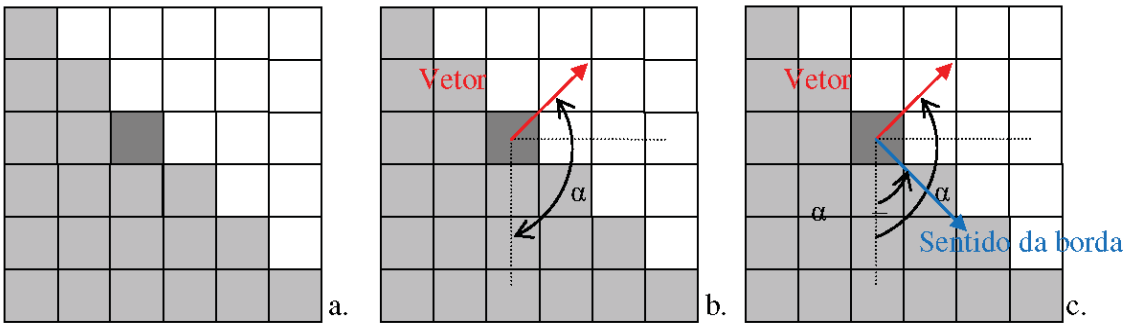


Figura 5.5 - Ponto analisado (a), vetor gradiente (b), vetor gradiente fazendo um ângulo reto com o sentido da borda (c)[74].

Em geral filtros derivativos não possuem bom desempenho quando as bordas estão difusas e com ruídos. A Figura 5.6 ilustra o resultado da convolução usando a máscara de Söbel [60] sobre a Figura 4.6, onde as cores do papel se aproximam das cores que compõem o plano de fundo onde se encontra a imagem. Já o filtro de Canny [70] foi proposto com o objetivo de contornar os problemas com imagens ruidosas, no qual a imagem é convoluída com as derivadas de primeira ordem do filtro Gaussiano para suavização na direção do gradiente local seguido pela detecção de bordas por limiarização [72].

Outro avanço foi o uso da transformada de Hough [73], essa transformada é um algoritmo que procura a reta que melhor se ajuste a um conjunto de pontos dado, este mecanismo ainda pode ser estendido para detectar curvas, círculos ou outras formas. Pode-se dizer que é uma melhoria dos métodos anteriormente já que se faz necessário o uso de algum desses métodos em associação com a

transformada. Para cada uma das infinitas retas que passam por cada ponto (x, y), existe um (r, θ) associado satisfazendo a Equação 5.1:

$$x\cos(\theta) + y\sin(\theta) = r \tag{5.1}$$

Os pontos (x, y) onde as retas candidatas devem passar estão definidos em uma imagem $J(x, y)$. Para cada um dos pontos definidos como cantos, somamos o valor de sua intensidade a imagem resultado $R(r, \theta)$, percorrendo θ e obtendo o r correspondente. Ao fim disso, a votação é normalizada e o ponto (r, θ) com maior número de votos correspondera ao (r, θ) da reta vencedora.

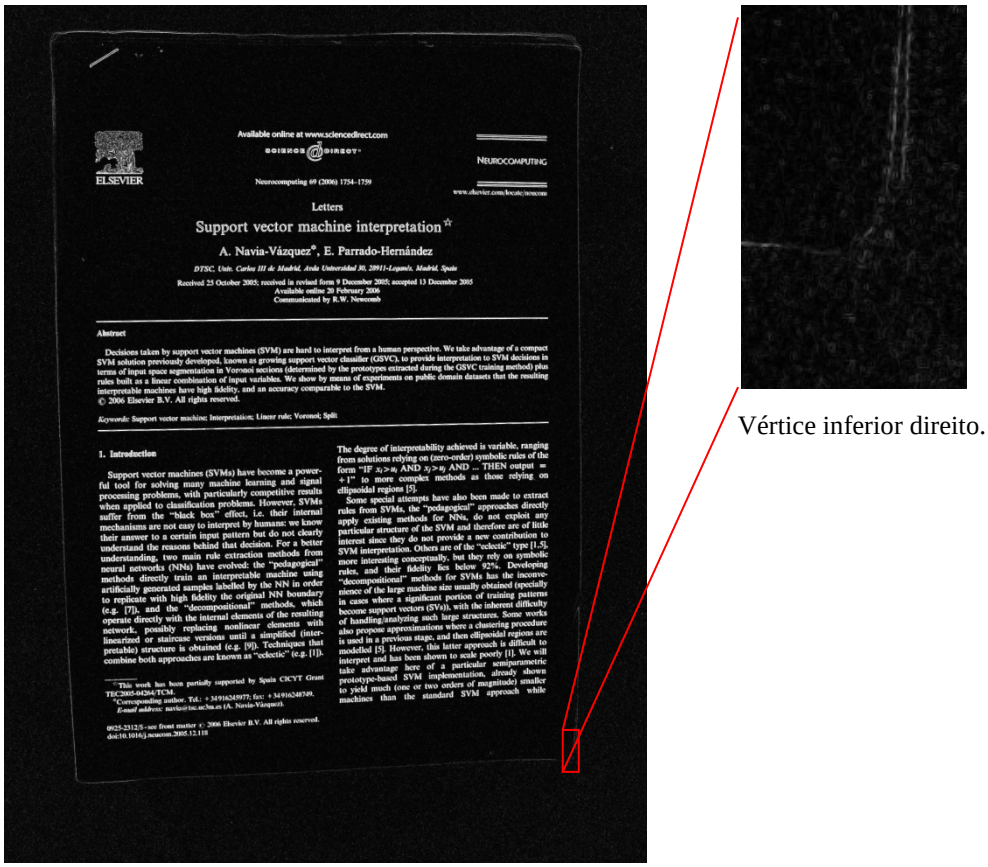


Figura 5.6 - Cores de papel e borda com valores próximos.

Técnicas recentes têm caracterizado a detecção de bordas como um problema de computação inteligente, empregando-se lógica fuzzy e redes neurais. Essas técnicas têm mostrado bons resultados e, portanto, promissoras nas áreas de processamento de imagens e visão computacional. Um breve resumo a respeito dessas técnicas é apresentado a seguir.

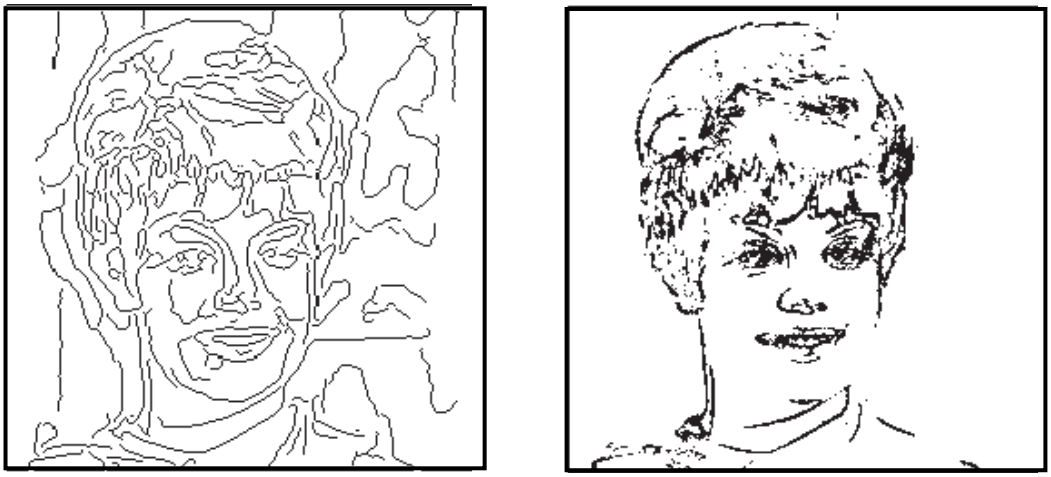
- As técnicas fuzzy permitem uma nova perspectiva para modelar as incertezas devido à imprecisão de valores de cinza presentes nas imagens. Desta forma, ao invés de atribuir valores de cinza a imagem, pode-se utilizar a pertinência fuzzy para definir os tons de cinza da imagem. As abordagens fuzzy para segmentação de imagem podem ser classificadas em abordagens baseadas em regras fuzzy e algoritmos de classificação fuzzy [71].

- Já o estudo das redes neurais artificiais (RNAs) aplicadas na detecção de bordas em imagens [67], consiste em treinar as redes neurais a reconhecer elementos de bordas (padrões) na imagem, assumindo que uma borda resulta da união de elementos básicos. Assim, os padrões utilizados nos treinamentos das redes neurais são padrões considerados como possíveis elementos de bordas. Geralmente são utilizados os modelos de redes neurais artificiais com aprendizagem supervisionada, tais como: *Perceptron* de Múltiplas Camadas (MLP) [68], Funções de Base Radial e Aprendizagem por Quantização Vetorial.

As Figuras 5.7 e 5.8 ilustram a aplicação de filtro de realce de contornos. Já a Figura 5.9 ilustra alguns padrões usados no treinamento dessas redes.

Por se tratar de um assunto extenso e rico apenas uma breve introdução foi apresentada neste Capítulo. Informações adicionais podem ser encontras em [13][66].

Figura 5.7 - Imagem Original.



(a) Filtro de Canny. (b) Filtro usando lógica fuzzy.

Figura 5.8 - Resultado da aplicação de filtros sobre a Figura 5.7.

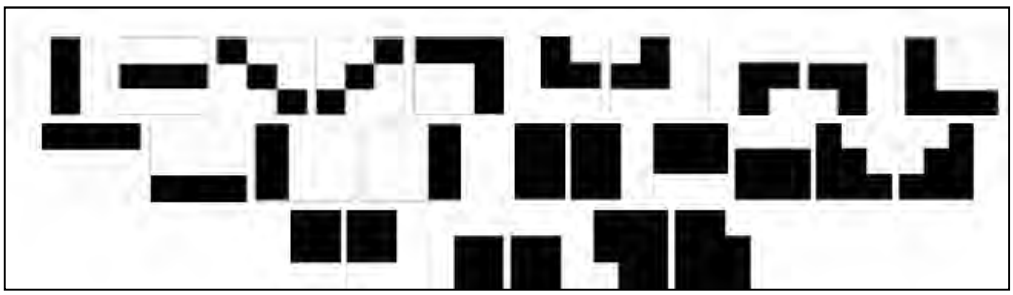
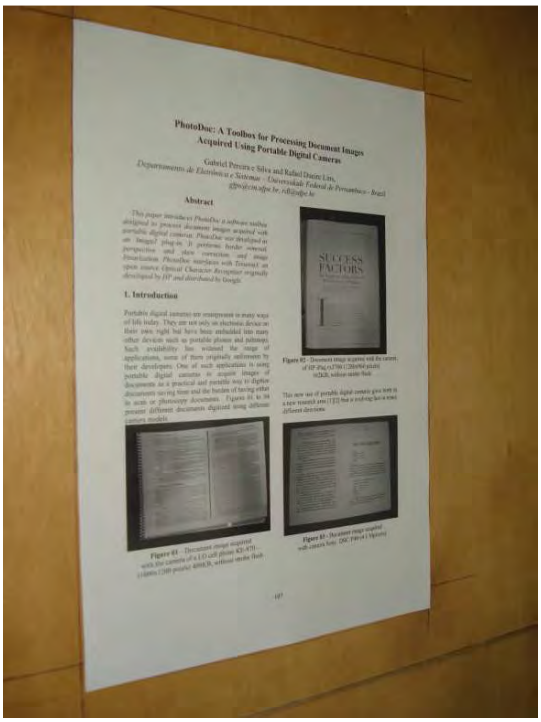


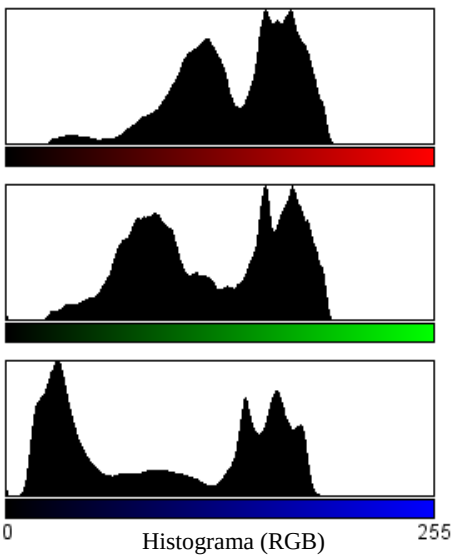
Figura 5.9 - Padrões de treinamentos para detecção de bordas usados por redes neurais.

5.2 Os algoritmos do PhotoDoc

Esta seção apresenta dois algoritmos de detecção de bordas para documentos fotografados. Esses algoritmos funcionam de maneira automática eliminando a necessidade de intervenção humana, já mencionado anteriormente como um processo lento e dispendioso [44].

Nos capítulos anteriores foi possível constatar que casos onde as imagens de documentos foram adquiridas por câmeras digitais, o surgimento de distorções geométricas é comum. Devido à existência desses ruídos, a remoção de bordas pode não ser totalmente eficiente, pode-se observar essa limitação nas Figuras 5.10 e 5.11. Ainda há problemas gerados por outros objetos presentes nas margens dos documentos, tais com: espirais de encadernação, página adjacente entre outros, as Figura 5.12 e 5.13 ilustram bem essa situação. De toda forma, os algoritmos que serão apresentados em seguida formam desenvolvidos de forma a buscar remover o máximo possível das bordas de forma a preservar a área referente ao documento.

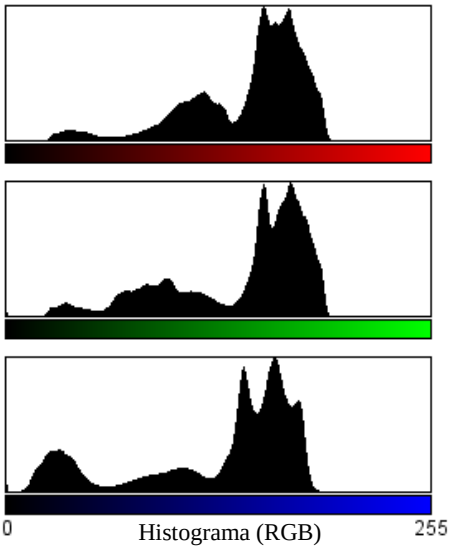




Histograma (RGB)

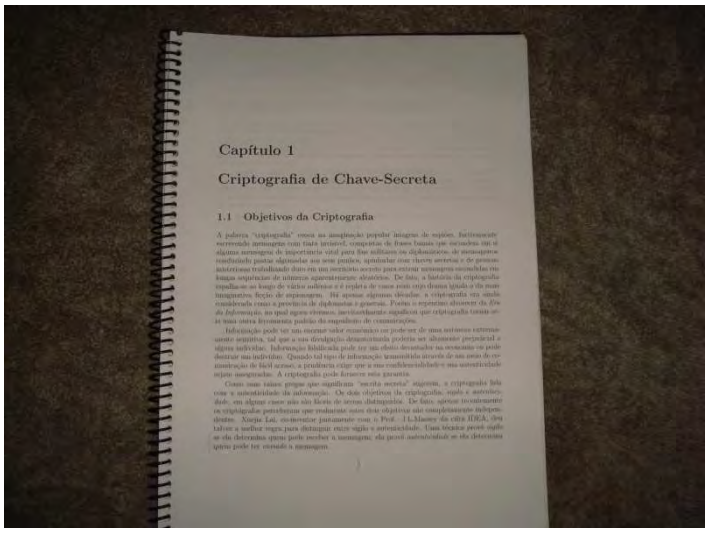
Documento fotografado com o uso do planetário.

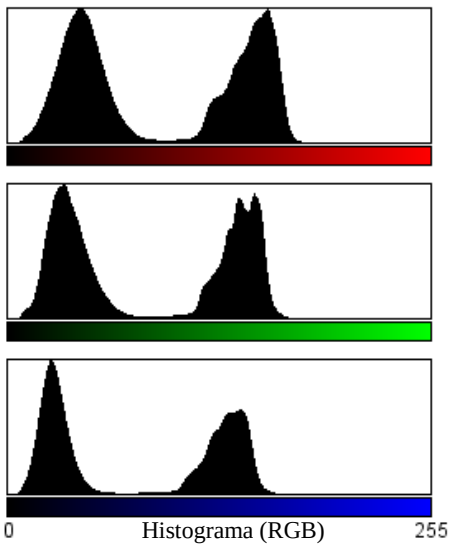
Figura 5.10 - Documento Fotografado em um ângulo de 30° sem o uso de flash.



Documento fotografado com o uso do planetário – Processado pelo PhotoDoc.

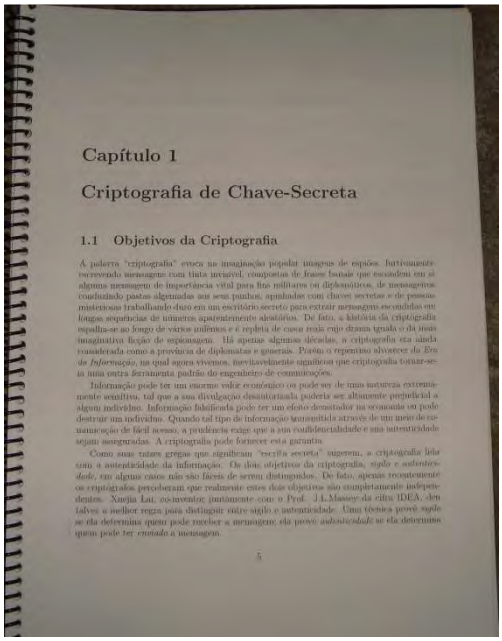
Figura 5.11 - Imagem da Figura 5.10 com remoção parcial de bordas.

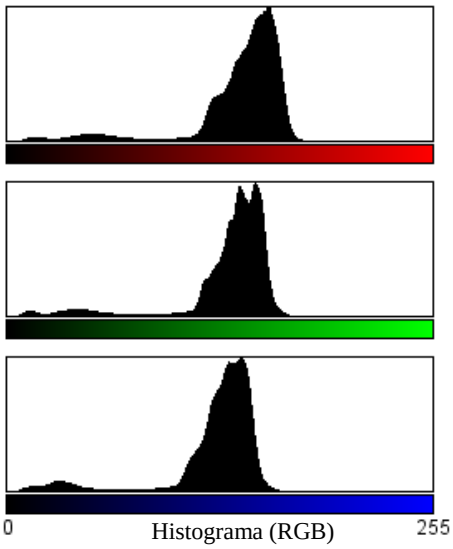




Documento fotografado a mão-livre.

Figura 5.12 - Documento Fotografado com o uso de flash.





Documento fotografado a mão-livre – Processado pelo PhotoDoc.

Figura 5.13 - Imagem ilustrada na Figura 5.12 com remoção parcial de bordas.

5.2.1 Detecção de Bordas (Algoritmo 1)

Visto que os usuários tendem a fotografar documentos nos mais variados ambientes o que leva a grandes variações na composição das bordas. Por este motivo optou-se pela identificação dos padrões do documento, já que esses padrões possuem menores variações. Observou-se ainda que pixels consecutivos pertencentes ao papel tendem a apresentar pequenas variações nas três componentes RGB, sendo estas variações para mais ou para menos. De acordo com estas observações, realizadas por meio de ferramentas estatísticas sobre o banco de dados, foi possível constatar que os pixels adjacentes pertencentes ao papel apresentavam uma variação media de até dezesseis níveis em duas de suas componentes (padrão RGB), já a terceira componente pode possui uma variação media de até o dobro da variação dos outros dois níveis.

As etapas a seguir descrevem o primeiro algoritmo para a detecção de bordas em documentos adquiridos por meio de câmeras digitais portáteis em *true color*, desenvolvido pelo grupo de engenharia de documentos da UFPE. A descrição desse algoritmo foi retirada de [11].

- **Etapa 1: Cálculo da moda no centro da imagem**

Levando em consideração o fato de a informação ocupar espaço muito menor do que o papel do documento calculou-se o pixel (RGB) que mais se repete na região central da imagem, correspondente a 1/9 do total da imagem. A decisão pela moda nesta região foi definida com base no brilho causado pelo flash ser mais intenso, e pela região mais provável de o documento estar localizado na imagem. Observando-se as variações identificadas entre pixels vizinhos pertencentes ao papel, esse pixel serviu como referência para a busca pelos limites do papel. Assumiu-se que o documento sempre preenche a maior parte da região central da imagem (Figura 5.14).

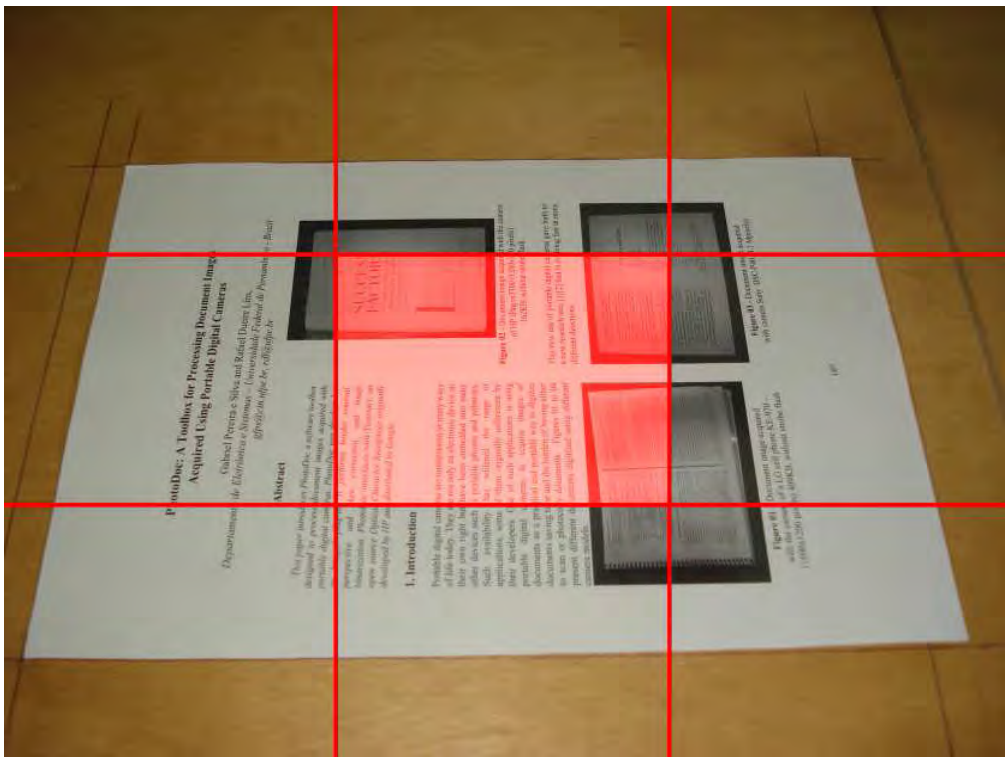


Figura 5.14 - Remoção de bordas (etapa 1).

- **Etapa 2: Estimativa dos limites do documento**

  Baseando-se na moda da região central da imagem, e utilizando-se o conhecimento das variações entre pixels vizinhos, buscou-se encontrar os limites do papel do documento, partindo do centro da imagem e em direção à área externa ao documento até que atinja os limites da imagem. Durante a varredura o pixel mais recentemente classificado como papel (α) e a moda do documento (β) são utilizados como referências para determinar se o próximo pixel deve ser classificado como papel ou não-papel. Observa-se que no padrão *true color* cada componente possui 8-bits cujos valores podem variar de 0 a 255. Duas tolerâncias são estabelecidas, com base na análise estatística da base de dados que se deseja filtrar.

  A primeira determina um limite no qual apenas uma das componentes do pixel atual pode ultrapassar quando comparada com a mesma componente em α e β. A segunda tolerância define um valor no qual as duas outras componentes podem variar. Caso mais de uma componente exceda esse limite, o pixel é classificado como informação ou borda. Estas tolerâncias determinam a flexibilidade do algoritmo, podendo ser ajustadas para diferentes conjuntos de imagens na obtenção de melhores resultados. Dessa forma, quatro pontos são identificados como limites dos documentos $(a_o = \{x_0, y_0\}, a_1 = \{x_1, y_1\}, a_2 = \{x_2, y_2\}, a_3 = \{x_3, y_3\})$, (Figura 5.15).

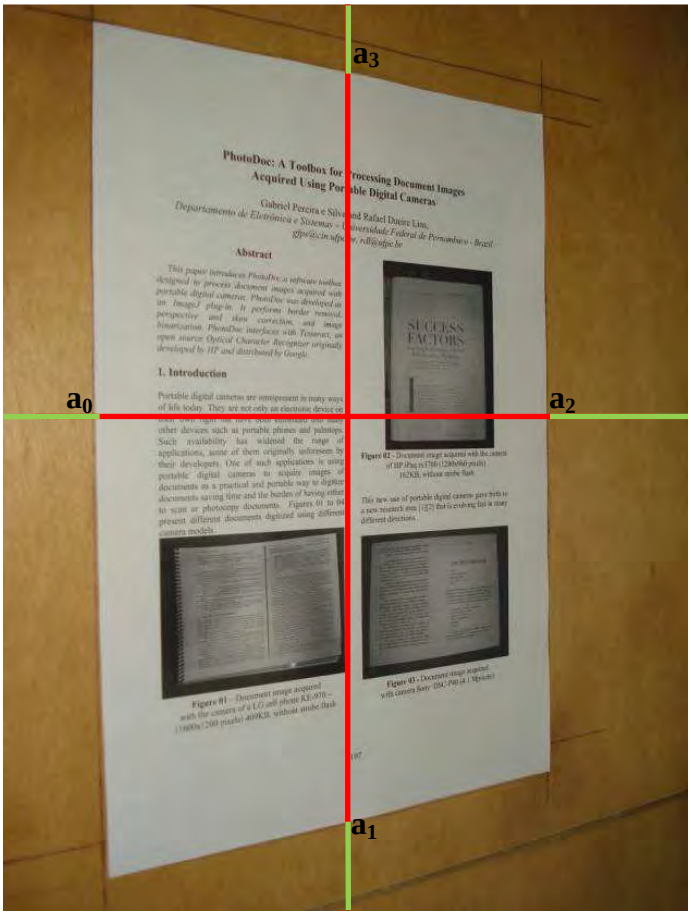


Figura 5.15 - Remoção de bordas (etapa 2).

- **Etapa 3: Cálculos para melhoria da precisão**

  Devido à variação de luz, que é freqüente ao utilizar-se o *flash* da câmera, a etapa 2 deste algoritmo pode não ser precisa. Ao aproximar-se do limite do documento, de dentro para fora, percebe-se que o brilho pode apresentar variações mais abruptas (superiores a 20%). Baseando-se na estimativa dos quatro pontos identificados como limites dos documentos, calcula-se a moda na região próxima a cada um destes pontos. Dois limites são determinados um interno e outro externo. Esses limites representam pixels consecutivos (linha reta) em direção à parte exterior e em direção à parte interior do documento, respectivamente (valores sugeridos em [11]). Partindo do ponto estimado, e cujos outros dois limites são dados pelos dois pontos encontrados em varredura ortogonal a este ponto. Por exemplo, a moda que contém o ponto a_0, tem como limites os pontos y_1, y_3, $x_0 - 15$, $x_0 + 60$, ou seja, $b_0 = \{x_0 - 15, y_3\}$, $b_1 = \{x_0 - 15, y_1\}$, $b_2 = \{x_0 + 60, y_1\}$, $b_3 = \{x_0 + 60, y_3\}$, (Figura 5.16). O valor de 15 pixels em direção à parte exterior do documento foi definido buscando-se assimilar valores do papel com menor iluminação, enquanto o valor de 60 pixels em direção ao centro do documento garante que a moda será pertencente ao papel, e não à borda, visto que esta pode apresentar composição de cores simples. Tais valores foram determinados baseando-se em valor percentual à quantidade de pixels das imagens adquiridas e apresentaram bons resultados para a base de dados apresentada em [59].

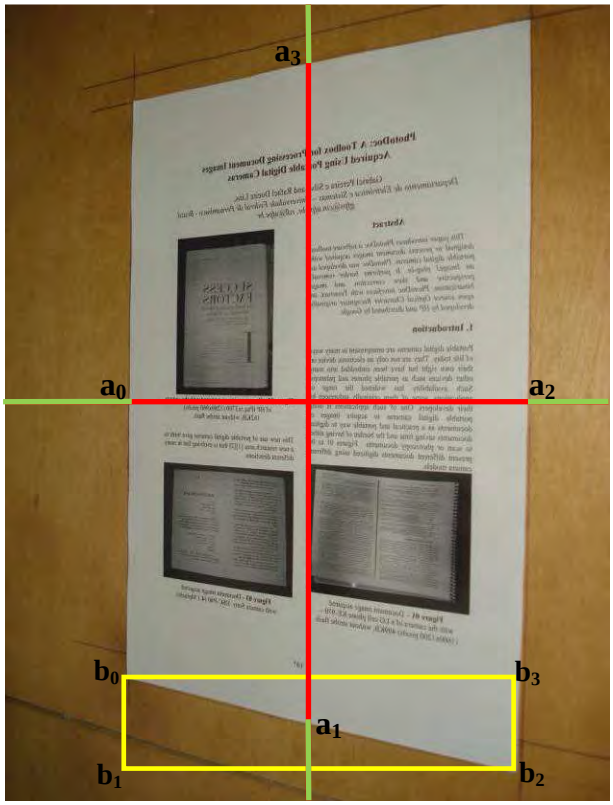


Figura 5.16 - Remoção de bordas (etapa 3).

- **Etapa 4: Encontrando os valores para corte da imagem**

  Com os valores das modas das regiões próximas aos pontos estimados como limites do documento, executa-se varredura partindo de um dos limites inferiores do documento - altura ou largura - dependendo da região do documento, em direção aos limites superiores, classificando os pixels como borda ou papel. A cada pixel classificado como papel, efetua-se o deslocamento de um pixel em direção à região externa e retrocedem-se cinco pixels em direção ao limite inferior atuante. Com este retrocesso busca-se evitar irregularidades nos limites dos documentos, como buracos ou papel irregularmente recortado. Esta verificação é realizada dois pixels por vez para aumentar a confiabilidade onde através das observações chegaram-se à definição da tolerância de 32 níveis para todos os canais. O último pixel classificado como papel indica o limite do documento, vertical ou horizontal, tal que $s_k = \{x_k, y_k\}$. Isto é, para verificar-se o limite esquerdo, parte-se da coordenada $(x_0, 0)$ em direção a $(x_k, y_{\max})$. O ponto de verificação deslocar-se-á verticalmente. Caso os pontos $\{x_0, y_n\}$ e $\{x_0 + 1, y_n\}$ sejam identificados como papel, então os próximos pixels a serem verificados seriam $\{x_0 - 1, y_n - 5\}$ e $\{x_0, y_n - 5\}$, (Figura 5.17). Este algoritmo é conservativo, ou seja, depois de estimados os quatro limites do documento, verificam-se as menores e maiores alturas e larguras, para então ser realizado o corte da imagem. O resultado pode ser observado na Figura 5.18.

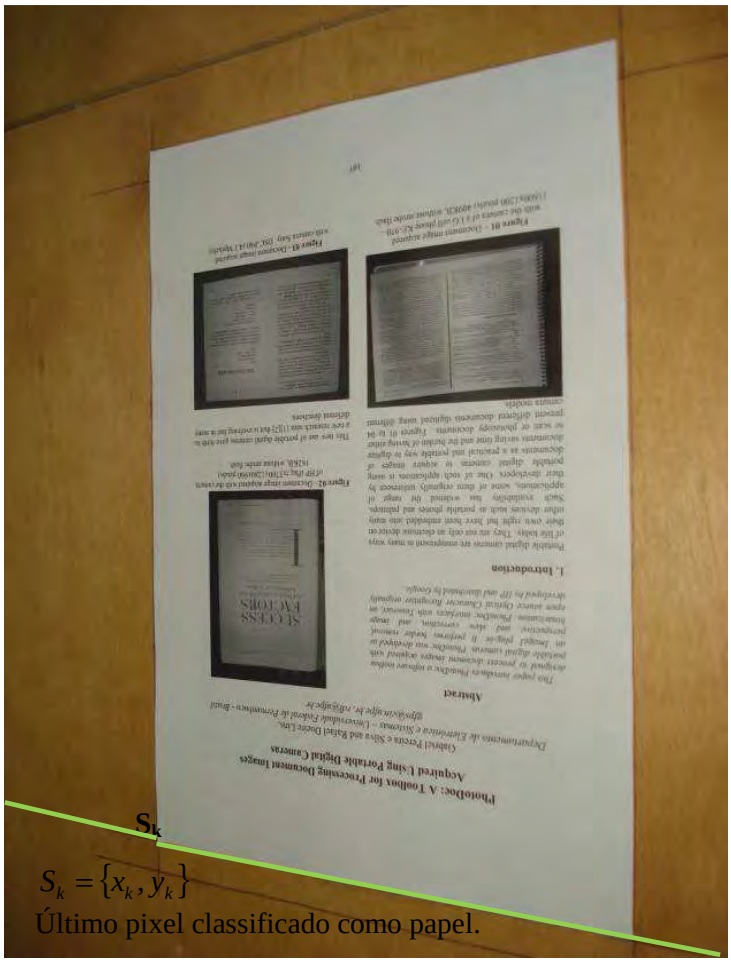


Figura 5.17 - Remoção de bordas (etapa 4).

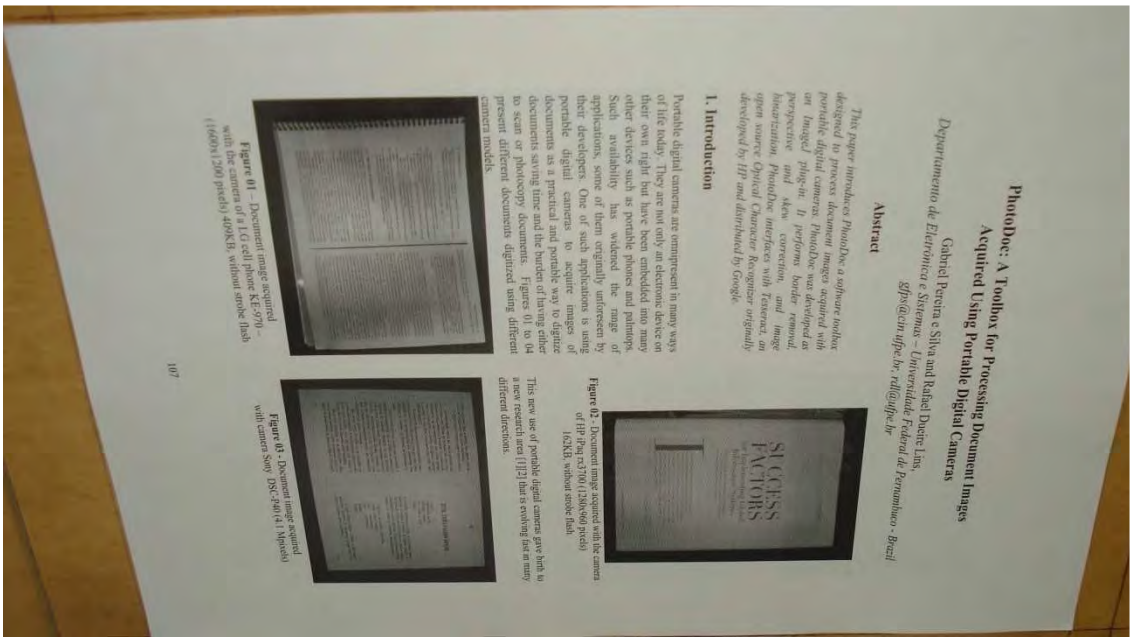


Figura 5.18 - Remoção de borda (final).

5.2.2 Detecção de Bordas (Algoritmo 2)

Devido às limitações presentes no algoritmo 1 (ver Tabela 5.1) um novo algoritmo ainda inédito de detecção de borda foi incluído em PhotoDoc.

Limitações	Algoritmo 1	Algoritmo 2
Parâmetros Fixos	Sim	Não
Documentos com Baixa Resolução	Não	Sim
***Layout* Complexo**	Baixo desempenho	Alto desempenho
Cor da Borda e do Papel Próximas	Baixo desempenho	Médio desempenho

Tabela 5.1 – Comparativo entre os algoritmos de remoção de bordas.

A partir das conclusões obtidas nos Capítulos anteriores pode-se afirmar que a estrutura básica de uma imagem de documentos fotografados é constituída por dois componentes: pelo conteúdo e pelos objetos ao redor do documento (plano de fundo). A Figura 5.19 destaca esses componentes. Ainda pode-se concluir que a única área que tem uma estrutura onde a variação da mesma é pequena entre todos os documentos da base que serviu de estudo é a área de conteúdo.

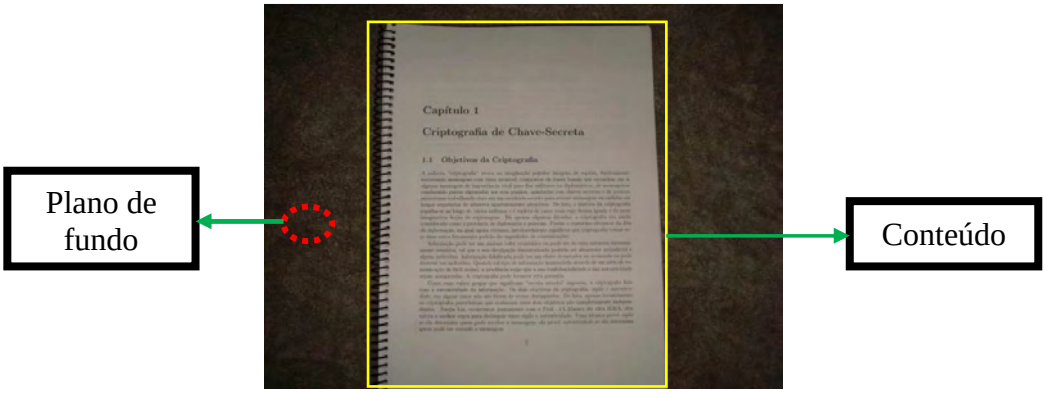


Figura 5.19 - Componentes básicos de um documento fotografado.

O novo algoritmo desenvolvido para compor o PhotoDoc é uma melhoria do algoritmo apresentado na subseção anterior. Antes de detalhar o funcionamento deste algoritmo, faz-se necessário definir algumas notações a serem utilizada no decorrer da explicação do mesmo.

- ***Bloco de Pixel* de origem:** representa o bloco de *pixels* a ser classificado (possui borda ou não);
- ***Bloco de Pixel* de apoio:** representa o bloco de *pixels* que será usado para comparação com o de origem;
- ***Bloco de borda:*** representa um bloco de origem que contém partes do exterior do documento.

A maioria dos métodos clássicos de detecção de bordas baseia-se em computações entre os pixels circunvizinhos. O novo mecanismo é baseado na entropia de Shannon [65][41] em cada componente do pixel (padrão RGB) entre blocos adjacentes. Basicamente o algoritmo é iniciado com cinco *blocos de pixels* de apoio pré-definidos construídos da seguinte forma: bloco central (1/9 da imagem) e os

blocos laterais (10% da altura e 10% da largura da imagem) esquerdo, direito, superior e inferior, a Figura 5.20 ilustra a posição desses blocos.

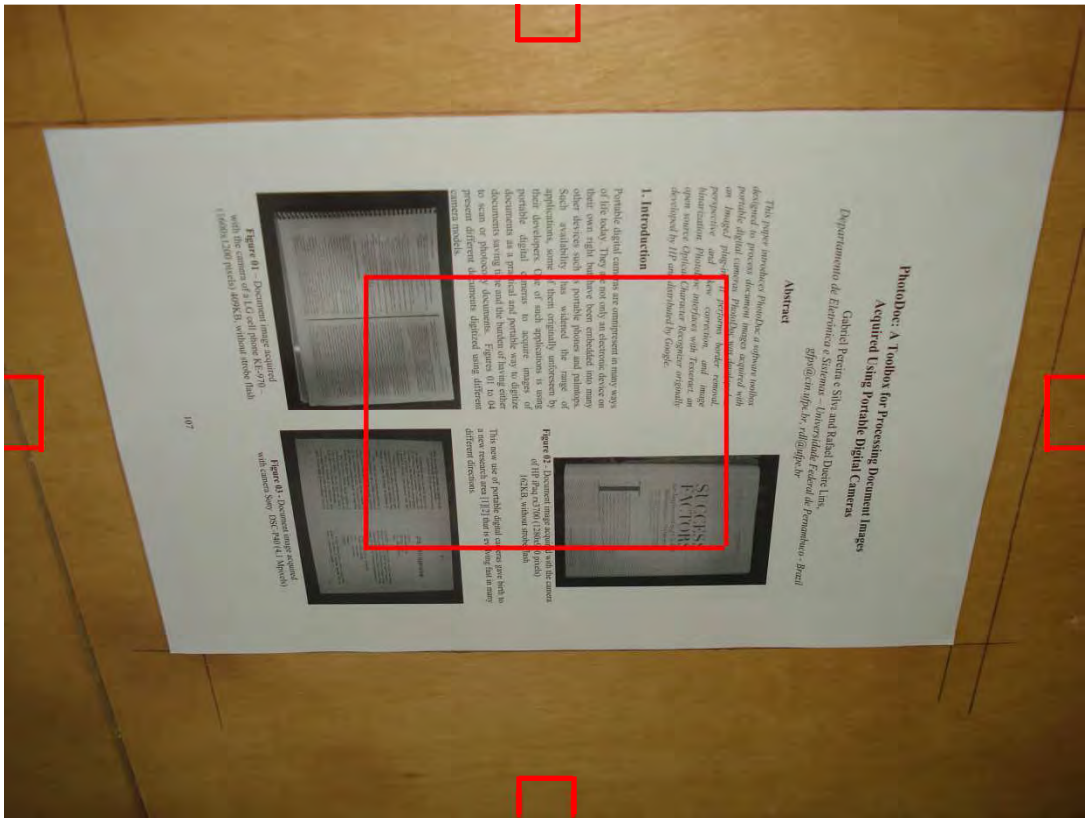


Figura 5.20 - Representação dos blocos de pixels iniciais.

As seguir as etapas desse algoritmo serão apresentadas.

- **Etapa 1: Cálculo das entropias dos *blocos de pixels***

  Calcula-se a entropia para cada componente R, G e B dos *blocos de pixels*, essa entropia irá definir um *threshold* para cada componente. O cálculo dar-se por:

$$H_{componente} = -\sum_{i=0}^{255} p_i \log_2(p_i),$$

  onde $\{p_0, p_1, ..., p_{255}\}$ é uma distribuição *a priori* dada por:

$$p_i = \frac{número\ \ de\ \ pixels\ \ de\ \ uma\ \ dada\ \ componente\ \ no\ \ bloco}{número\ \ tatal\ \ de\ \ pixels\ \ do\ \ bloco}.$$

  Para cada "*t*" calcula-se a distribuição a *posteriori.* O "*t*" é uma variável que determina a distribuição $\{1-P_t\}$ (distribuição após a partição). Ela representa um limiar, pois particiona o histograma em duas partes.

  $\left\{P_t = \sum_{i=0}^{t} p_i, 1-P_t\right\}$, enquanto $P_t \leq 0{,}5$ a entropia segue a seguinte distribuição:

$$H'(t) = -P_t \log(P_t) - (1-P_t)\log(1-P_t).$$

  Finalmente para cada componente é determinado o valor que minimiza a expressão dada por:

$$|e(t)| = \left| \frac{H'(t)}{H/\log(256)} - \alpha(H/\log(256)) \right|$$, onde α é dado por:

$$\alpha(H/\log(256)) = \begin{cases} -(3/7)(H/\log(256)) + 0{,}8 \ \ se \ \ H/\log(256) < 0{,}7 \\ H/\log(256) - 0{,}2 \ \ se \ \ H/\log(256) \geq 0{,}7 \end{cases}$$

Sobre as últimas expressões, a idéia é fazer o ajuste estatístico entre as duas entropias. $H'(t) = H/\log(256)$. Mas, como existe a interferência na imagem, houve a necessidade da inserção do α. Assim, o ideal seria: $H'(t) = \alpha(H/\log(256)) \Rightarrow H'(t)/[H/\log(256)] = \alpha$. Como é praticamente impossível de ocorrer a igualdade, buscou-se minimizar a diferença absoluta entre os dois membros da equação $|e(t)| = H'(t)/[H/\log(256)] - \alpha$. Já o valor de α foi determinado experimentalmente.

Por meio desses cálculos é possível estabelecer os *threshold* para cada uma das componentes dos blocos e servirão de base como critério de parada para o algoritmo.

- **Etapa 2: Busca dos limites do documento**

  Calculados os *threshold*, busca-se encontrar os limites do papel do documento, partindo do centro da imagem e em direção à área externa ao documento até que atinja os limites da imagem. Define-se um novo bloco de origem correspondente a 1/9 do central, iniciando-se a varredura, onde novos blocos serão definidos contendo sempre a metade horizontal do bloco anterior no caso da busca na vertical e a metade horizontal no caso vertical. Para cada novo bloco calcula-se o *threshold* das componentes R, G e B. Em seguida compara-se esse bloco com os cincos blocos de apoio. Caso mais de um *threshold* do bloco de origem ultrapasse em mais de 25% o seu correspondente *threshold* no bloco central e abaixo de 35% em relação à pelo menos dois blocos laterais, esse bloco é tido por bloco de borda, caso contrário compara-se o próximo bloco de origem.

- **Etapa 3: Cálculo das novas tolerâncias do algoritmo 1**

  Devido à variação de luz, que é freqüente ao utilizar-se o flash da câmera, a etapa 2 do algoritmo anterior pode não ser precisa. Para tal calculam-se novas tolerâncias para o algoritmo anterior a partir do bloco de borda. Neste bloco é aplicado um filtro passa banda e em seguida usa-se o algoritmo para detecção de bordas de Söbel [60] destacado o contorno do documento e assim permitindo a análise mais precisa das características dos pixels do conteúdo e do plano de fundo. As Figuras 5.21 e 5.22 ilustram o resultado da aplicação dos filtros a blocos de bordas. Por fim chama-se a etapa 2 do algoritmo anterior iniciando a análise a partir do pixel do bloco de borda mais próximo do centro da imagem.

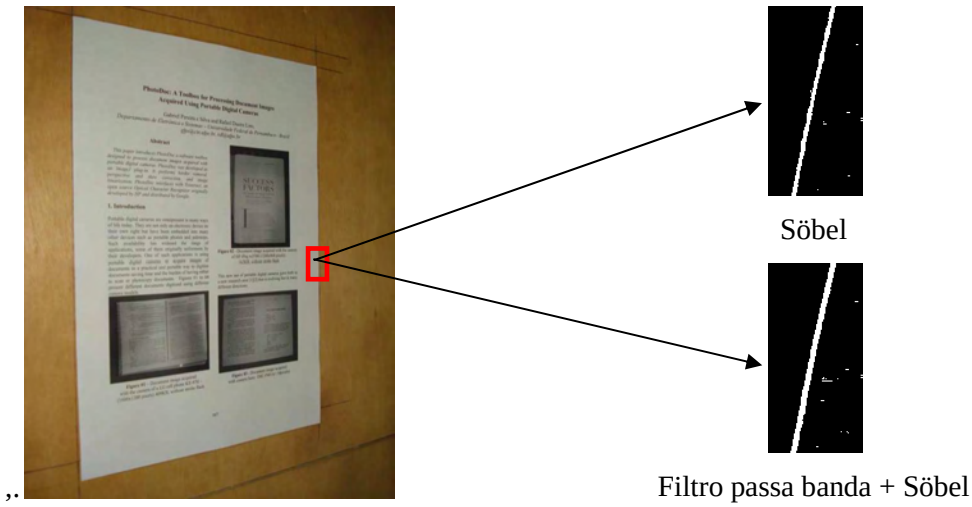


Figura 5.21 - Imagem de documento com cores do papel e borda distantes.

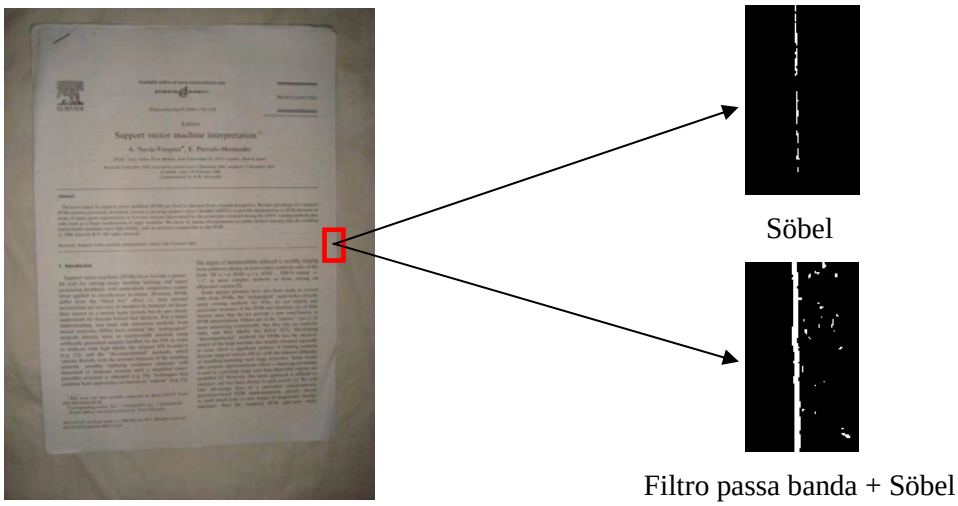


Figura 5.22 - Imagem de documento com cores do papel e borda próximas.

5.2.3 Algoritmo para busca dos vértices

Por se tratar de um algoritmo que foi derivado a partir do algoritmo apresentado na subseção 5.2.1, decidiu-se apresentá-lo neste Capítulo. Basicamente a sua função é identificar os vértices do documento e fornecê-los ao algoritmo de correção de perspectiva, sobre esse assunto trataremos no próximo Capítulo desta dissertação. É valido acrescentar que o uso do algoritmo 2 apresentado na subseção 5.2.2 também é empregado aqui, mas para um melhor entendimento, optou-se por descrever esse algoritmo de busca dos vértices usando os princípios do algoritmo 1.

De fato os algoritmos de correções de distorções geométricas em imagens requerem a entrada de informações adicionais, fornecidas pelo usuário ou obtidas automaticamente, no caso estudado buscam-se capturar parâmetros necessários à correção de perspectiva. Existe uma rica literatura científica acerca desse assunto onde essas informações podem ser obtidas por meio de: características do layout [6][29][32][39], limites dos documentos (contornos ou vértices) [5] e características de conteúdo específico (tipo de texto, símbolos conhecidos etc.) [28][16].

Para identificar os parâmetros de entrada do algoritmo de correção de perspectiva, optou-se pela identificação dos vértices do documento, tendo em vista sua simplicidade computacional e na aplicação deste algoritmo a documentos, os quais tiveram suas características observadas ao longo deste trabalho. Neste algoritmo se busca identificar pontos nos contornos da imagem, para então se calcular quatro equações de retas. A partir da intersecção dessas retas estimam-se os vértices dos documentos. Para se encontrar esses pontos nos contornos, há a necessidade da utilização de características dos documentos, para assim diferenciá-lo da borda. Optou-se pela identificação das cores mais freqüentes nas laterais da imagem como informação para decisão da localização do contorno. Criou-se então um banco de contendo 3072 documentos, dentre eles foram selecionados mil para validar os algoritmos, nos ângulos de 0° com alturas baixa (40 cm) e alta (60 cm), 15° e 30° (altura alta e direções Sul e Oeste ver Figura 3.1) e a mão-livre. A partir da análise dos resultados se chegou à conclusão que o novo algoritmo é quase duas vezes mais eficiente que o primeiro, a Tabela 5.1 apresenta mais detalhes.

Resolução (*Mpixels)*	Flash	Método de captura	Ângulo	Número de imagens	Acertos algoritmo 1	Acertos algoritmo 2
3.1	Sim	Mão-livre	Não se aplica	50	64%	88%
3.1	Não	Mão-livre	Não se aplica	50	62%	88%
4.2	Sim	Mão-livre	Não se aplica	50	70%	91%
4.2	Não	Mão-livre	Não se aplica	50	70%	91%
5.1	Sim	Mão-livre	Não se aplica	100	58%	98%
5.1	Não	Mão-livre	Não se aplica	100	51%	95%
5.1	Sim	Planetário	0° Baixa	30	63,33%	100%
5.1	Sim	Planetário	0° Alta	30	60%	100%
5.1	Sim	Planetário	15°S - 0°O	35	34,28%	97,14%
5.1	Sim	Planetário	30°S - 0°O	35	28,57%	97,14%
5.1	Não	Planetário	15°S -15°O	35	48,57%	97,14%
5.1	Não	Planetário	15°S - 30°O	35	35,71%	100%
7.2	Sim	Mão-livre	Não se aplica	100	51%	98%
7.2	Não	Mão-livre	Não se aplica	100	53%	98%
7.2	Sim	Planetário	0° Baixa	30	60%	100%
7.2	Sim	Planetário	0° Alta	30	60%	100%
7.2	Sim	Planetário	15°S - 0°O	35	34,28%	100%
7.2	Sim	Planetário	30°S - 0°O	35	28,57%	100%
7.2	Não	Planetário	15°S -15°O	35	34,28%	100%
7.2	Não	Planetário	15°S - 30°O	35	28,57%	100%
				Acerto médio:	61%	96%

Tabela 5.2 - Teste de validação dos algoritmos de busca de vértices.

As etapas do algoritmo que busca os pontos a serem fornecidos ao algoritmo de correção de perspectiva para documentos fotografados são descritas a seguir:

- **Etapa 1: Estimativa dos limites laterais**

  A estimativa dos limites laterais torna-se necessária para o cálculo das cores nas regiões próximas aos contornos. Tal estimativa ocorre da mesma forma descrita pelas etapas um e dois do algoritmo apresentado na subseção 5.2.1. De forma que os pontos encontrados são: $\left(a_0 = \{x_0, y_0\}, a_1 = \{x_1, y_1\}, a_2 = \{x_2, y_2\} \text{ e } a_3 = \{x_3, y_3\}\right)$.

- **Etapa 2: Região de estimativa dos contornos**

  Para melhor precisão da estimativa dos vértices é necessário o cálculo da moda nas regiões mais distantes do centro do documento. Estas novas modas são calculadas apenas nas regiões próximas aos pontos a_1 e a_2, visto que as regiões laterais contêm os vértices e possuem maior variação de brilho. Para o cálculo dessa moda utilizou-se dois limiares que representam pixels consecutivos e em linha reta apontando em direção à região externa ao e em direção à área interna.

- **Etapa 3: Cálculo das equações das retas**

  Utilizando-se os valores adquiridos pelos cálculos das modas nas regiões laterais do documento, executa-se varredura partindo dos limites inferiores e superiores do documento (altura ou largura) em direção aos limites superiores e inferiores, respectivamente, classificando os pixels como documento ou borda. Esta varredura tem início em uma das coordenadas centrais do documento, vertical ou horizontal, e a outra coordenada no ponto máximo ou mínimo, em direção ao documento. Por exemplo, para se encontrar os dois pontos utilizados para o cálculo da reta no contorno inferior do documento, inicia-se a varredura a partir do ponto P_0 = (L/2, 0), em direção ao ponto (L/2, H), onde L e H representam a largura e a altura do documento, respectivamente. Ao encontrar um pixel classificado como documento, desloca-se a varredura para a esquerda e para a direita, sendo os dois últimos pixels, um da esquerda e um da direita, classificados como documento para posterior utilização no cálculo da equação da reta (Figura 5.23). A cada pixel classificado como papel, retrocedem-se cinco pixels em direção à área externa ao documento, para se evitar as possíveis irregularidades nos contornos. Como critério de classificação, as cores constituintes de um pixel precisam estar dentro de uma tolerância de 32 níveis em relação ao valor das modas nas laterais dos documentos para a classificação como parte do documento. Este valor foi encontrado através das análises realizadas durante o desenvolvimento do algoritmo descrito na subseção 5.2.1. Da mesma forma da etapa anterior esse valor pode ser alterado pelo algoritmo descrito na subseção 5.2.2. Depois de encontrados dois pontos pertencentes a cada um dos lados do documento, procede-se então com a definição das equações das retas, e em seguida, o cálculo das intersecções entre elas, resultando nos vértices estimados do documento (Figura 5.24).

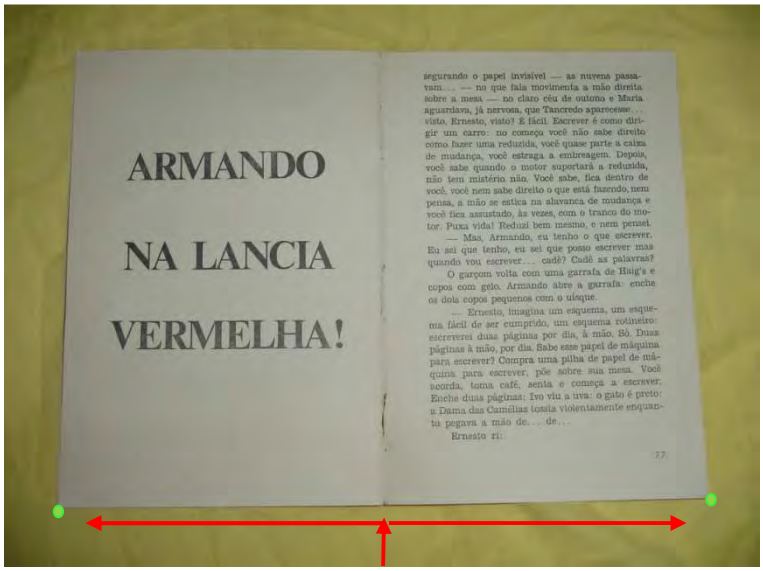


Figura 5.23 - Localização dos pontos no contorno.

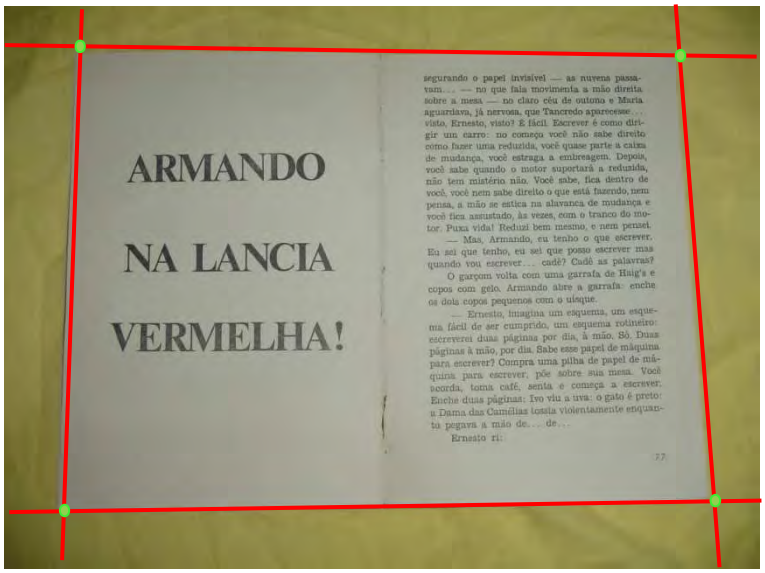
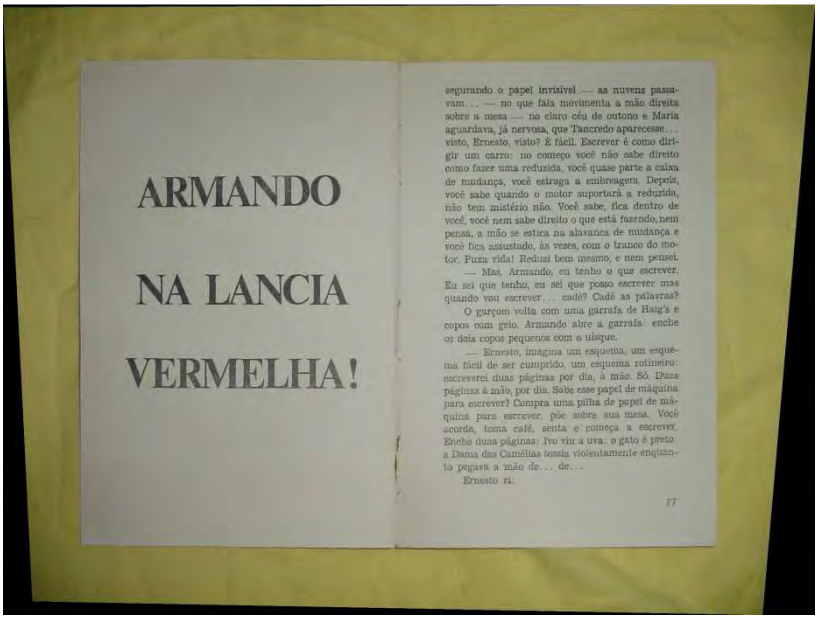


Figura 5.24 - Pontos coincidentes das retas.

- **Etapa 4: Transformação da imagem**

  De posse dos vértices estimados do documento procede-se então com o cálculo da razão de aspecto para definir o tamanho da nova imagem, e com isso calcular também os vértices do documento na imagem destino. Uma vez estimados os vértices do documento nas imagens origem e destino, é possível o cálculo do homógrafo, e posteriormente, a multiplicação deste homógrafo pela imagem original, resultando na correção da perspectiva (Figura 5.25). Essa transformação será tratada no próximo Capítulo desta dissertação.



segurando o papel invisível — as nuvens passavam... — no que fala movimenta a mão direita sobre a mesa — no claro céu de outono e Maria aguardava, já nervosa, que Tancredo aparecesse... visto, Ernesto, visto? É fácil. Escrever é como dirigir um carro: no começo você não sabe direito como fazer uma reduzida, você quase parte a caixa de mudança, você estraga a embreagem. Depois, você sabe quando o motor suportará a reduzida, não tem mistério não. Você sabe, fica dentro de você, você nem sabe direito o que está fazendo, nem pensa, a mão se estica na alavanca de mudança e você fica assustado, às vezes, com o tranco do motor. Puxa vida! Reduzi bem mesmo, e nem pensei.

— Mas, Armando, eu tenho o que escrever. Eu sei que tenho, eu sei que posso escrever mas quando vou escrever... cadê? Cadê as palavras?

O garçom volta com uma garrafa de Haig's e copos com gelo. Armando abre a garrafa: enche os dois copos pequenos com o uísque.

— Ernesto, imagina um esquema, um esquema fácil de ser cumprido, um esquema rotineiro: escreverei duas páginas por dia, à mão. Só. Duas páginas à mão, por dia. Sabe case papel de máquina para escrever? Compra uma pilha de papel de máquina para escrever, põe sobre sua mesa. Você acorda, toma café, senta e começa a escrever. Enche duas páginas: Ivo viu a uva: o gato é preto: a Dama das Camélias tosse violentamente enquanto pegava a mão de... de...

Ernesto ri:

Figura 5.25 - Localização dos pontos no contorno.

Capítulo 6
Correção de Perspectiva para documentos fotografados

O presente capítulo estuda e corrige a distorção de perspectiva em documentos fotografados com câmeras digitais portáteis. Ainda neste capítulo será introduzido o algoritmo para correção de perspectiva [50] que integra o PhotoDoc .

6.1 Correção de distorções de perspectiva

Correção de perspectiva pode ser assumida como a recuperação da visão frontal de uma imagem através da obtenção do homógrafo a partir de uma visão arbitrária. Essa correção é um passo corriqueiro no processamento de imagem de documentos fotografados, mesmo com o auxílio de suporte mecânico se o mesmo não estiver bem posicionado e calibrado fatalmente acorrerá o fenômeno de distorções geométricas na imagem, ou seja, sempre que o documento a ser fotografado não estiver paralelo ao plano da objetiva da câmera haverá distorção de perspectiva na imagem gerada.

A distorção causada por uma má perspectiva da câmera gera considerável grau de dificuldade na análise dos documentos, incluindo a remoção de bordas, como foi mencionado no Capítulo 3, e as ferramentas de OCR, visto que o texto do documento sofre em algumas áreas diminuição de resolução.

A maior parte dos algoritmos empregados na correção de distorções tira proveito das estruturas de texto conhecidas como: distribuição de palavras, linhas igualmente espaçadas etc [62][63][64]. Antes da aplicação de algoritmos de correção de perspectiva, faz-se necessário o conhecimento do processo de captura de imagens, descrito na próxima seção.

6.1.1 A formação de imagens

A formação de imagens em câmera fotográfica foi rapidamente discutida no Capítulo 2 dessa dissertação, grosso modo, pode-se definir uma câmera fotográfica como uma câmera escura com um furo minúsculo onde a imagem é projetada do lado oposto ao furo sobre um anteparo, esse modelo é conhecido como *pinhole* [13]. Neste capítulo será apresentada uma modelagem matemática dessas câmeras.

A imagem captada pelas câmeras pode ser modelada através de uma transformação espacial tridimensional (mundo real) para um bidimensional (plano da imagem), para esse estudo a perda de uma componente espacial, no caso a profundidade, não será prejudicial ao desenvolvimento do estudo aqui proposto. Uma vez conhecidas às coordenadas da câmera, pode-se então modelar o

sistema óptico de funcionamento da mesma. Embora existam diferentes modelos de câmera, decidiu-se adotar o modelo *pinhole* por motivos já mencionados no Capítulo 2. Observando-se a Figura 6.1 é possível estabelecer as relações entre os objetos no espaço tridimensional e uma imagem em um plano, conforme a Equação 6.1. Observa-se ainda que os raios de luz provenientes de um objeto em $\left(X_1, X_2, X_3 \right)$ e que passam através do orifício atingindo o plano da imagem em (x_1, x_2, d_i).

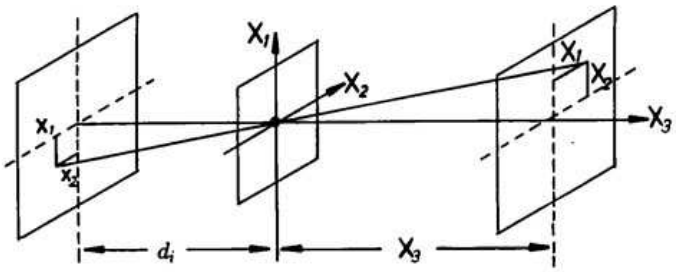


Figura 6.1 - Formação de imagens com uma câmera *pinhole*.

$$(X_1, X_2, X_3) \rightarrow (x_1, x_2) = (\frac{d_i X_1}{X_3}, \frac{d_i X_2}{X_3}) \qquad (6.1)$$

Ao utilizar coordenadas generalizadas, divide-se a transformação dada na Equação 6.1 por (d_i), que representa a distância do orifício à imagem projetada, obtendo a transformação:

$$(X_1, X_2, X_3) \rightarrow (x_1, x_2) = (\frac{X_1}{X_3}, \frac{X_2}{X_3}) \qquad (6.2)$$

6.1.2 Coordenadas homogêneas

As matrizes de coordenadas generalizadas não suportam as operações de translação, rotação, redimensionamento e projeção de perspectiva, porém sabe-se que é possível realizar essas transformações fazendo-se a troca das coordenadas generalizadas [16]. Para contornar essas limitações as coordenadas homogêneas serão adotadas nesse estudo. Coordenadas homogêneas são definidas como um vetor de quatro componentes, $X = (tX_1, tX_2, tX_3, t)$, através do qual as coordenadas tridimensionais são obtidas dividindo-se os três primeiros componentes pelo quarto. As operações elementares estão ilustradas na Tabela 6.1.

As transformações completas de pontos no espaço tridimensional para coordenadas em imagens podem ser entendidas através da aplicação de cada uma das transformações acima, seqüencialmente, visto que multiplicação de matrizes é propriedade associativa, pode-se resumir todo o processo com uma única matriz M. A matriz M pode seguir a seguinte decomposição:

$$\mathrm{M} = \mathrm{TR_X}, \mathrm{R_Y R_Z} PEC \qquad (6.3)$$

onde as matrizes T , R_X , R_Y , R_Z , P, E e C correspondem a translação, rotação nos eixos x, y e z, projeção de perspectiva, redimensionamento e recorte, respectivamente.

Portanto, conclui-se que a formação de imagem utilizando o modelo de câmera *pinhole* e aplicando-se coordenadas generalizadas é dada por [14][16]:

$$p = MP \tag{6.4}$$

onde p é o ponto na imagem com dimensões 3x1 (coordenadas homogêneas), P é o ponto no espaço tridimensional com dimensões 4x1 (coordenadas homogêneas) e M é a matriz da câmera com dimensões 4x1 (coordenadas homogêneas) composta por parâmetros internos e externos dela, como a posição da câmera. O referido modelo preserva a incidência de linhas e induz transformações lineares no espaço projetivo.

$T = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ -T_1 & -T_2 & -T_3 & 1 \end{bmatrix}$	$R_x = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & \cos\theta & \sin\theta & 0 \\ 0 & -\sin\theta & \cos\theta & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$
Translação por ($-T_1$, $-T_2$, $-T_3$).	Rotação em torno do eixo x por θ.
$R_y = \begin{bmatrix} \cos\psi & 0 & \sin\psi & 0 \\ 0 & 1 & 0 & 0 \\ -\sin\psi & 0 & \cos\psi & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$	$R_z = \begin{bmatrix} \cos\phi & 0 & \sin\phi & 0 \\ -\sin\phi & \cos\phi & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$
Rotação em torno do eixo y por ψ.	Rotação em torno do eixo z por ϕ.
$E = \begin{bmatrix} s_1 & 0 & 0 & 0 \\ 0 & s_2 & 0 & 0 \\ 0 & 0 & s_3 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$	$P = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & -1/d_i \\ 0 & 0 & 0 & 1 \end{bmatrix}$
Mudança de escala.	Projeção de perspectiva.

Tabela 6.1 - Matrizes de transformações na formação da imagem [59].

6.1.3 Transformação de Perspectiva

Quando imagens de objetos planos, caso em que se encontram os de documentos, são capturadas, as imagens observadas de diferentes posições são relacionadas por uma transformação projetiva linear é possível simplificar a relação da Equação 6.4 em:

$$x_i^{'} = Hx_i \tag{6.5}$$

$x_i^{'}$ e x_i são vetores com dimensões 3x1 (coordenadas homogêneas) e poderiam corresponder a imagens de um mesmo ponto. Já a matriz H possui dimensões 3x3 (coordenadas homogêneas), definida por um fator de escala e oito graus de liberdade, o que torna possível formar um sistema com

oito equações e oito variáveis. Dados quatro pontos correspondentes entre duas imagens, *H* pode ser unicamente calculado [14].

Se $x' = \left[x', y' \right]^T$ e $x = \left[x, y \right]^T$, são os pontos correspondentes em duas imagens relacionadas por um homógrafo, então:

$$x' = \frac{h_{11x} + h_{12y} + h_{13}}{h_{31x} + h_{32y} + h_{33}} \qquad y' = \frac{h_{21x} + h_{22y} + h_{23}}{h_{31x} + h_{32y} + h_{33}} \tag{6.6}$$

Visto que o homógrafo possui oito variáveis, é necessário o número mínimo de oito equações para solucionar o sistema. Tais variáveis podem ser calculadas a partir de quatro pontos correspondentes entre duas imagens, esses pontos são fornecidos pelo algoritmo descrito no Capítulo 4. O principal fator a ser analisado nos pontos de saída é a manutenção da proporção real do documento. A outra questão é relativa ao tamanho da imagem, como está se trabalhando com imagens digitais, é desejável que o mapeamento de um ponto da imagem de entrada corresponda a, pelo menos, um ponto da de saída, caso contrário há perda de informação. Caso a proporção seja conhecida, é trivial definir-se os pontos de saída com o tamanho desejado. Nos outros casos é necessário levar em consideração a informação *a priori* da forma do documento ser um retângulo como será visto nos três métodos que serão apresentados a seguir.

6.2 Métodos de Interpolação

Interpolação é um método que permite construir um novo conjunto de dados a partir de um conjunto de dados conhecidos. Em imagens os dados a serem interpolados são os *pixels*, através da interpolação destes pode-se construir uma nova imagem com características semelhantes à imagem original.

6.2.1 Interpolação pelos vizinhos mais próximos

A interpolação pelos vizinhos mais próximos *(Nearest Neighbor Interpolation)* é um dos métodos mais rápido e simples de interpolação, também conhecida por interpolação de ordem zero, ela simplesmente determina o valor de um ponto P na imagem destino calculando a média dos pontos vizinhos (*pixels* adjacentes) a esta região na imagem original. Por ser um método que leva em conta apenas a informação dos vizinhos esse método de interpolação possui alguns efeitos indesejáveis como as distorções conhecidas como *jaggies*, que ocorre nos contornos presentes na imagem deixando-os serrilhados ou imprecisos.

6.2.2 Interpolação linear

O método de interpolação linear é aplicado em uma dimensão, determinando-se um ponto baseando-se em outros dois. Conhecendo-se as coordenadas $P_0 = (x_0, y_0)$ e $P_1 = (x_1, y_1)$, ilustradas na Figura 6.2, deseja-se determinar os pontos na linha formada por P_0 e P_1 com um dado x no intervalo $[x_0, x_1]$. Fazendo:

$$\alpha = \frac{y - y_0}{y_1 - y_0} = \frac{x - x_0}{x_1 - x_0} \qquad (6.7)$$

em que α é denominado coeficiente de interpolação. Como o valor de x já é conhecido, pode-se determinar α. Manipulando-se matematicamente a equação, têm-se:

$$y = (1 - \alpha) y_0 + \alpha y_1 \qquad (6.8)$$

através do qual é possível o cálculo de y diretamente. Como a aplicação é em uma dimensão, determina-se que a intensidade da cor c do pixel p é representada pelo y acima descrito, enquanto x denota a posição de p na linha (ou coluna) da imagem.

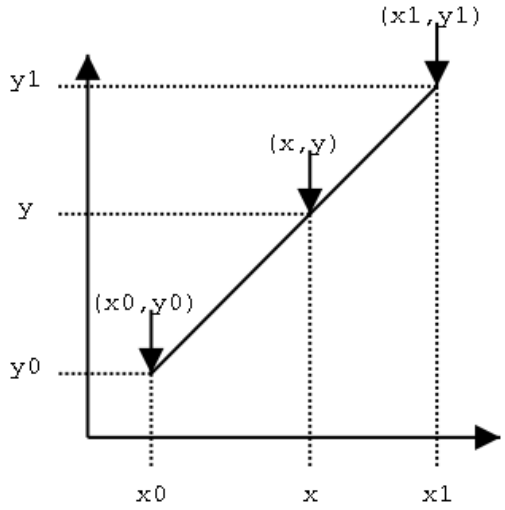


Figura 6.2 - Interpolação Linear.

6.2.3 Interpolação bilinear

Interpolação bilinear é uma extensão da interpolação linear para aplicação em funções de duas variáveis. A idéia principal é executar a interpolação linear em uma direção e depois em outra ortogonal à primeira.

Supondo que se deseja encontrar o valor de uma função desconhecida *f* no ponto P = (x, y). Assume-se que são conhecidos quatro valores de f, $Q_{11} = (x_1, y_1)$, $Q_{12} = (x_1, y_2)$, $Q_{21} = (x_2, y_1)$, $Q_{22} = (x_2, y_2)$, ilustrados na Figura 6.3.

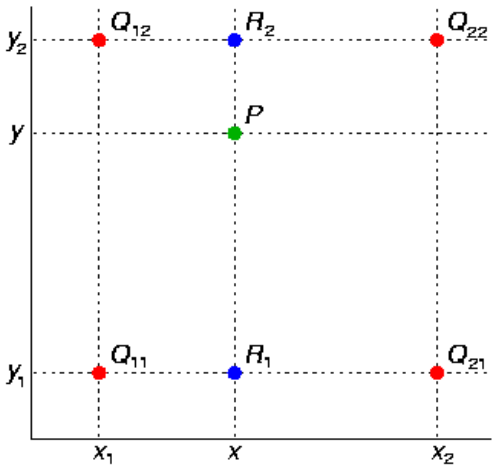


Figura 6.3 - Interpolação bilinear [59].

Primeiro executa-se interpolação linear na direção x:

$$f(R_1) = \frac{x_2 - x}{x_2 - x_1} f(Q_{11}) + \frac{x_2 - x}{x_2 - x_1} f(Q_{21}) \quad (6.9)$$

$$f(R_2) = \frac{x_2 - x}{x_2 - x_1} f(Q_{12}) + \frac{x_2 - x}{x_2 - x_1} f(Q_{22}) \quad (6.10)$$

Em que $R_1 = (x, y_1)$ e $R_2 = (x, y_2)$. Em seguida aplica-se a interpolação na direção y:

$$f(P) = \frac{y_2 - y}{y_2 - y_1} f(R_1) + \frac{y_2 - y}{y_2 - y_1} f(R_2) \quad (6.11)$$

Ao contrário do que o nome sugere, observa-se que a interpolação bilinear não é linear.

6.2.4 Interpolação bicúbica

A interpolação bicúbica preserva detalhes presentes na imagem ao custo de tempo adicional para a execução da interpolação. Neste método, o valor *f* (x, y) de uma função *f* no ponto (x, y) é calculado como uma média ponderada dos dezesseis pontos mais próximos a ele, montando uma matriz 4x4. Dois polinômios cúbicos de interpolação são utilizados, um para cada direção [22].
A interpolação bicúbica é calculada da seguinte forma:

$$f(x, y) = \sum_{i=0}^{3} \sum_{j=0}^{3} a_{ij} x^i y^j \quad (6.12)$$

O procedimento utilizado para encontrar os coeficientes a_{ij} depende das propriedades dos dados de origem. Supondo que se deseja encontrar o ponto P = (j + x, k + y) ilustrado na Figura 6.4, as equações são dadas por:

• Interpolação segundo o eixo horizontal

$$a_{j+x,k} = \frac{1}{6}(a_{j-1,k}R_1 + a_{j,k}R_2 + a_{j+1,k}R_3 + a_{j+x+2,k}R_4) \quad (6.13)$$

• Interpolação segundo o eixo vertical

$$a_{j+x,k} = \frac{1}{6}(a_{j+x,k-1}R_1 + a_{j+x,k}R_2 + a_{j+x,k+1}R_3 + a_{j+x,k+2}R_4) \quad (6.14)$$

Onde os coeficientes R_1 a R_4 são:

$$R_1 = (3+x)^3 - 4(2+x)^3 + 6(1+x)^3 - 4x_3$$

$$R_2 = (2+x)^3 - 4(2+x)^3 + 6x^3$$

$$R_3 = (1+x)^3 - 4(2+x)^3$$

$$R_4 = x^3$$

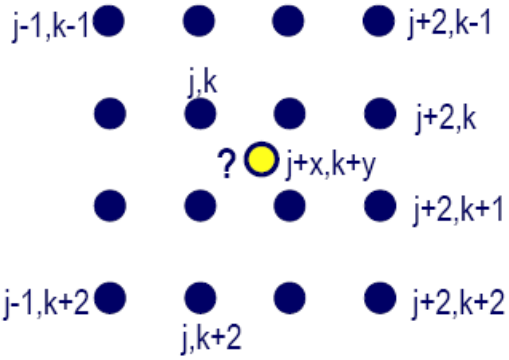


Figura 6.4 - Interpolação bicúbica [59].

A Figura 6.5 mostra os resultados obtidos pela rotação de uma linha vertical de cor preta em 17 graus, utilizando os algoritmos de interpolação pela vizinhança mais próxima, bilinear e bicúbica. É observado que o algoritmo de interpolação pela vizinhança mais próxima não produz níveis de cinza, gerando descontinuidades na linha. A interpolação bilinear produz linhas contínuas, enquanto a interpolação bicúbica preserva o melhor contraste da linha.

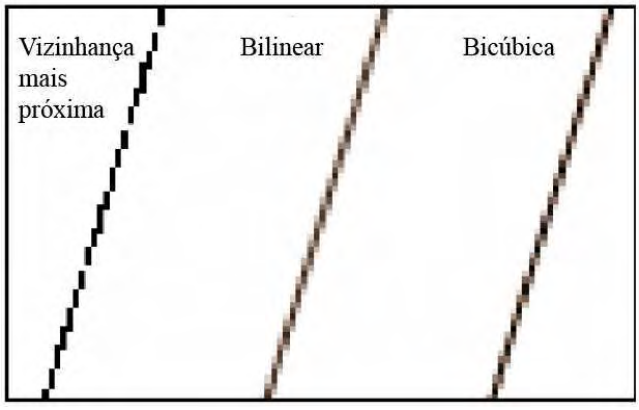


Figura 6.5 - Comparação dos métodos de interpolação.

6.3 PhotoDoc - correção de perspectiva

O método descrito em [49][59] apresenta um algoritmo simples e eficiente, fundamentado no cálculo da estimativa da razão largura/altura .

A Figura 6.6 ilustra um exemplo de quadrilátero que simula o contorno de um documento com distorção geométrica. Seus vértices são representados pelos pontos V_0 (vértice esquerdo superior), V_1 (vértice direito superior), V_2 (vértice esquerdo inferior) e V_3 (vértice direito inferior).

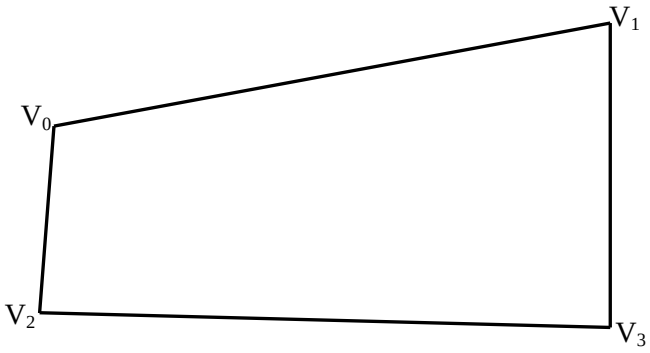


Figura 6.6 - Quadrilátero que representa o contorno de um documento.

A razão largura/altura $\left(R_{\frac{L}{H}}\right)$ é definida pelas somas das distâncias entre os vértices (V_0 e V_1) e os vértices (V_2 e V_3) dividido pelas distâncias entre os vértices (V_0 e V_2) e os vértices (V_1 e V_3), essa razão poder melhor visualizada na Equação 5.15.

$$R_{\frac{L}{H}} = \frac{\|V_0 - V_1\| + \|V_2 - V_3\|}{\|V_0 - V_2\| + \|V_1 - V_3\|} \tag{6.15}$$

Uma vez calculada a razão pode-se definir as novas dimensões do documento na imagem final. De forma a garantir que não existirá perda de informação é necessário impor a condição que para cada *pixel* da imagem de entrada seja mapeado, pelo menos, em um ponto da imagem de saída. Para tal calcula-se o teto dos máximos das larguras (L_{MAX}) e da altura (H_{MAX}) do quadrilátero.

$$\begin{cases} L_{MAX} = \max\left(\|V_0 - V_1\|, \|V_2 - V_3\|\right) \\ H_{MAX} = \max\left(\|V_0 - V_2\| + \|V_1 - V_3\|\right) \end{cases} \tag{6.16}$$

Em seguida verifica-se qual dessas dimensões abrange uma maior área, levando-se em consideração a $\left(R_{\frac{L}{H}}\right)$ neste cálculo é possível definir a largura e altura final da imagem. A Equação 6.17 ilustra o cálculo da largura e altura final.

$$\begin{cases} L_{FINAL} = H_{MAX} \times R_{\frac{L}{H}} \;\; e \;\; H_{FINAL} = H_{MAX} \quad se \quad L_{MAX} > H_{MAX} \times R_{\frac{L}{H}} \\ L_{FINAL} = L_{MAX} \;\; e \;\; H_{FINAL} = L_{MAX} \Big/ R_{\frac{L}{H}} \end{cases} \tag{6.17}$$

Tendo em mãos as dimensões finais do documento, pode-se definir as coordenadas dos quatro pontos do vértice na imagem transformada. Resolvendo o sistema de equações, envolvendo as coordenadas dos vértices da imagem de entrada com a de saída, obtendo a matriz homográfica de transformação *H*. Para calcular a transformação, observa-se a relação entre as coordenadas da imagem original e transformada, extraídas durante a aplicação do algoritmo descrito na subseção 5.2.3, que fornece os quatro pontos de origem, e seus correspondentes no destino, portanto têm-se oito equações no total, em um sistema com nove variáveis. Em [14] são fornecidos os meios matemáticos para encontrar a nona equação do sistema e assim solucioná-lo.

Capítulo 7
Realce de imagens de documentos adquiridos por câmeras digitais portáteis

Este capítulo irá apresentar um algoritmo de filtragem que podem ser utilizados para melhorar a qualidade da imagem de documentos fotografados. O objetivo central do realce é transformar os pontos das imagens pertencentes ao papel do documento em uma cor uniforme, removendo os efeitos indesejáveis da iluminação irregular.

O filtro que será apresentado é capaz de compensar o efeito da iluminação irregular, possibilitando uma melhor transcrição por ferramentas de OCR.

7.1 Realce de imagens no domínio do espaço

Realçar imagens significa processar uma imagem de modo que o resultado seja o mais apropriado possível para uma dada aplicação, no caso desta dissertação busca-se compensar a iluminação irregular das imagens de documentos fotografados para obter um melhor resultado dos algoritmos de binarização e assim obter uma melhor transcrição automática.

Realçar uma imagem f é transformá-la em outra imagem g utilizando um operador T definido em sua vizinhança (x, y). Os valores dos pixels das imagens f e g podem ser definidos por s e r. Onde T é a função de transformação de níveis de cinza da forma: $s = T(r)$, onde r e s denotam os níveis de cinza de $f(x, y)$ e $g(x, y)$ no ponto (x, y). A Figura 7.1 ilustra duas funções de transformação.

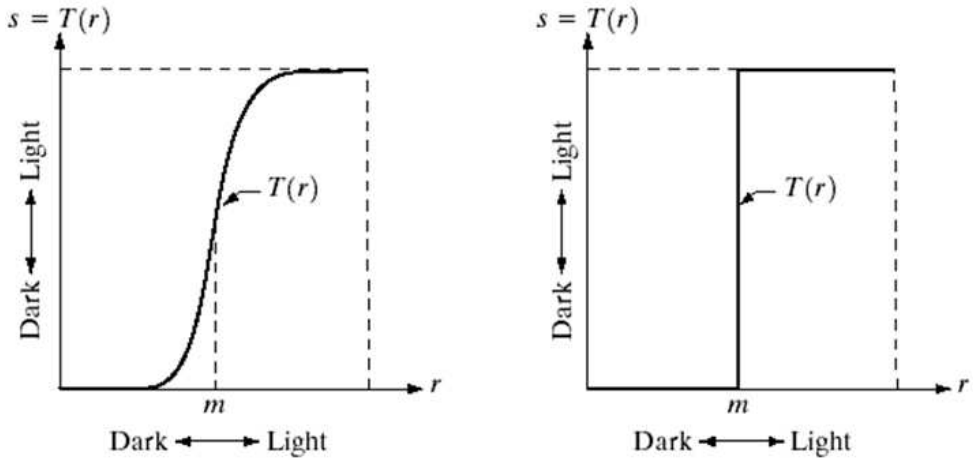


Figura 7.1 - Transformação de níveis de cinza por contraste de realce [13]

Filtragem espacial é definida como qualquer técnica ou processo de tratamento de imagens, que modificam o conteúdo da imagem, e tendem a enfatizar feições de interesse do usuário, enquanto suprime outras indesejáveis, como por exemplo, iluminação não uniforme. Os três principais tipos de filtros são o passa-alta, passa-baixa e passa-faixa. Denomina-se passa-alta quando ocorre diminuição dos componentes de baixa freqüência e aumento dos de alta freqüência, ocorrendo um realce das bordas e detalhes da imagem. Os filtros do tipo passa-baixa tendem a aumentar os componentes de baixa freqüência e diminuir os de alta freqüência, ocorrendo perda de detalhes e redução do contraste da imagem, porém atenua a influência de processos ruidosos provocados, por exemplo, por defeitos do sensor e erros na transformação matemática. Os filtros passa-faixa atenuam influências de ruídos periódicos. A seguir, são descritas algumas das principais técnicas de realce de imagens.

7.1.1 Ampliação de contraste

É freqüente obter-se imagens com baixo contraste em situações que envolvem iluminação não uniforme ou de baixa intensidade ou ainda devido a deficiências do sensor de visão. A operação de espalhamento de contraste busca uniformizar a distribuição de um histograma de forma a preencher toda faixa do espectro de cinza. Por exemplo, uma imagem de documento histórico codificada com *8-bits*, terá os seus valores de *pixels* originais transformados para a faixa de valores entre 0 e 255. As transformações nos níveis de cinza aumentam o contraste de faixas de intensidades e a depender do caso pode até mesmo binarizar essas imagens. Embora a transformação mais comum seja a linear, pode-se implementar qualquer outro tipo de transformação, dependendo do histograma original e do alvo ou feição de interesse, ou seja, essa transformação pode ainda ser: logarítmica, exponencial, raiz quadrada etc. A Figura 7.2 ilustra a curva de alguma destas transformações.

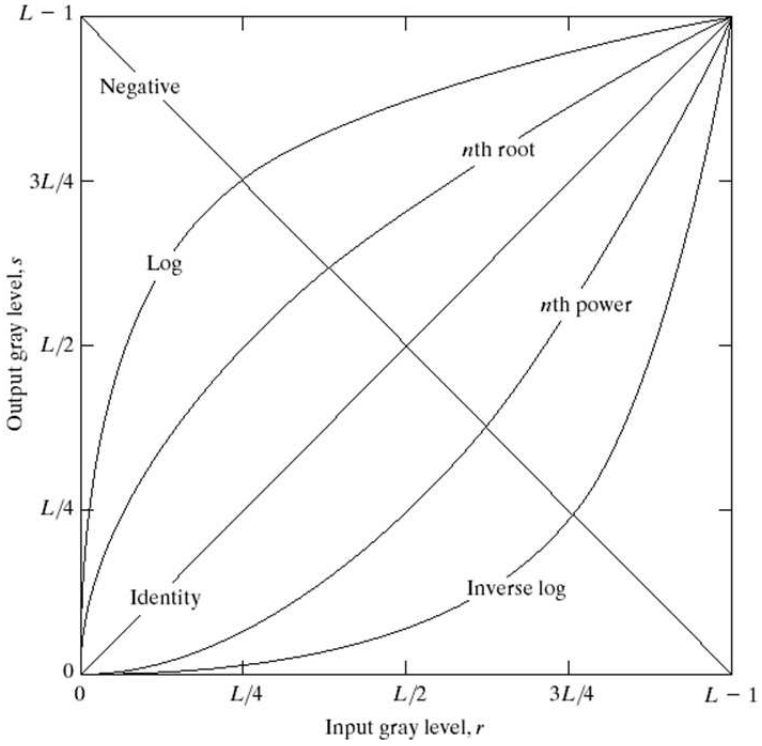


Figura 7.2 - Transformações não-lineares [13].

- **Transformação linear:** O aumento de contraste por uma transformação linear é a forma mais simples de transformação de níveis de cinza. A função de transferência é uma reta e apenas dois parâmetros são controlados: a inclinação da reta e o ponto de intersecção com o eixo X (Figura 7.3). A inclinação controla o aumento de contraste e o ponto de intersecção com o eixo X controla a intensidade média da imagem final [13]. A função de mapeamento linear pode ser representada por:

$$y = aX + b\,, \tag{7.1}$$

onde *y* é o novo valor, *X* é o valor original, *a* é a inclinação da reta (tangente do ângulo), b é o fator de incremento, definido pelos limites mínimo e máximo fornecidos pelo usuário. No aumento linear de contraste as barras que formam o histograma da imagem de saída são espaçadas igualmente, uma vez que a função de transferência é uma reta. Como se pode observar na Figura 7.3, o histograma de saída será idêntico, em formato, ao histograma de entrada, exceto que ele terá um valor médio e um espalhamento diferentes.

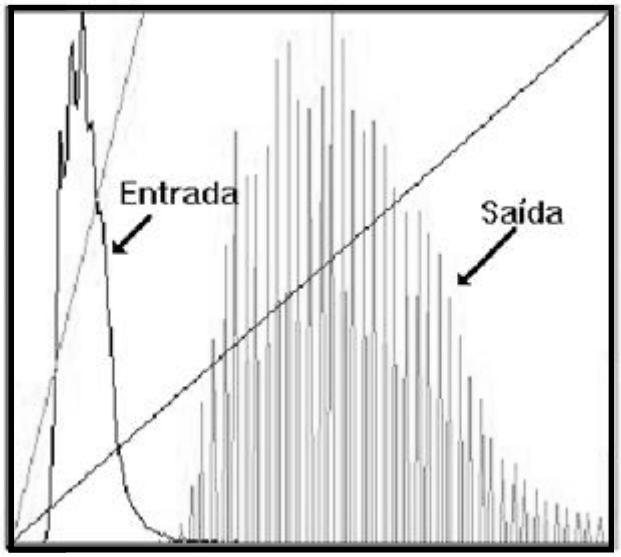


Figura 7.3 - Exemplo de transformação linear [8].

- **Transformação raiz:** Utiliza-se a opção de transformação por raiz quadrada para aumentar o contraste das regiões escuras da imagem original. A função de transformação é representada pela curva, como mostra a Figura 7.4. Observa-se que a inclinação da curva é inversamente proporcional aos valores de níveis de cinza. Uma transformação raiz pode ser expressa pela equação:

$$y = a \times \sqrt{X}\ , \tag{7.2}$$

onde *y* é o nível de cinza resultante, *X* é o nível de cinza original e *a* é o fator de ajuste para os níveis de saída (0 e 255). Este mapeamento difere do logarítmico porque realça um intervalo maior de níveis de cinza baixos (escuros), enquanto o logarítmico realça um pequeno intervalo.

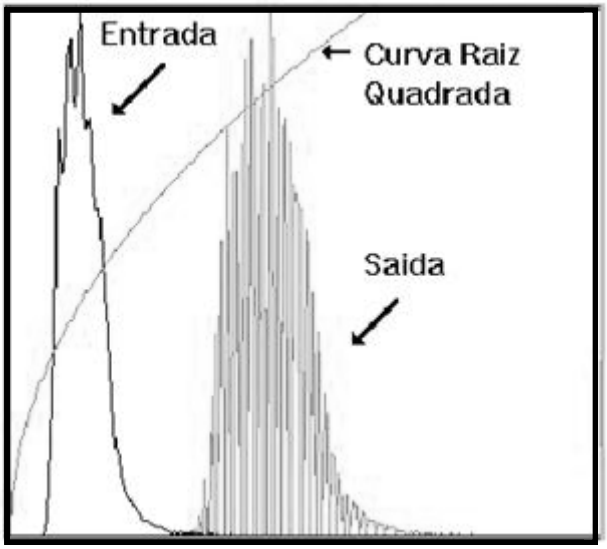


Figura 7.4 - Exemplo de transformação raiz [8].

- **Transformação logarítmica:** O mapeamento logarítmico de valores de níveis de cinza é útil para aumento de contraste em feições escuras (valores de cinza baixos). Equivale a uma curva logarítmica como mostrado na Figura 7.5. A função de transformação é expressa pela equação:

$$y = a \times \sqrt{X} \ , \qquad (7.3)$$

onde *y* é o novo valor de nível de cinza, *X* é o valor original de nível de cinza e *a* é o fator definido a partir dos limites mínimo e máximo da tabela, para que os valores de saída (0 e 255).

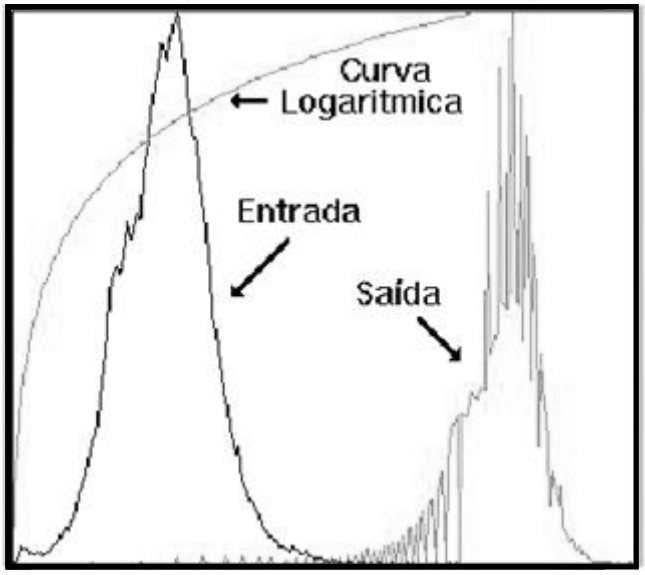


Figura 7.5 - Exemplo de transformação logarítmica [8]

- **Transformação negativa:** É uma função de mapeamento linear inversa, ou seja, o contraste ocorre de modo que as áreas escuras (baixos valores de nível de cinza) tornam-se claras (altos valores de nível de cinza) e vice-versa. A função de mapeamento negativa pode ser representada por:

$$y = -(a \times X + b)\,, \qquad (7.4)$$

onde *y* é o novo valor de nível de cinza, *X* é o valor original de nível de cinza, *a* é a inclinação da reta (tangente do ângulo) e *b* é o fator de incremento, definido pelos limites mínimo e máximo fornecidos pelo usuário. A Figura 7.6 ilustra essa transformação.

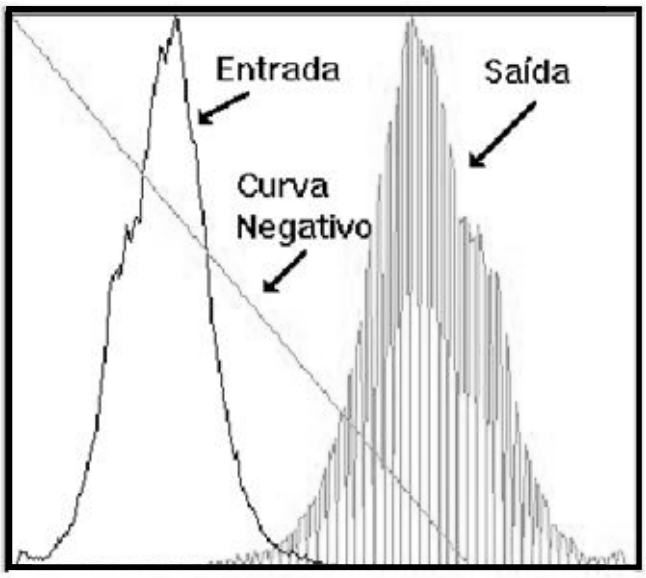


Figura 7.6 - Exemplo de transformação negativa [8].

7.1.2 Composição colorida

A utilização de composições coloridas é fundamentada pelo fato que o olho humano é capaz de discriminar mais facilmente matiz de cores do que tons de cinza. Para cada banda, associa-se uma cor primária (azul, verde ou vermelha) ou, ainda, as suas complementares (amarela, magenta ou ciano), de modo que para cada alvo diferente da cena associa-se uma cor ou uma combinação de cores diferentes. Essa técnica comumente utiliza imagens que já estejam realçadas por ampliação de contraste. Uma das poucas restrições a este método é a utilização simultânea de bandas, limitada ao máximo de três. A reconstituição das cores na imagem advém do processo aditivo de formação das cores primárias (azul, verde e vermelho). A imagem resultante é costumeiramente denominada imagem colorida RGB (*red*, *green* e *blue*). Outro tipo de realce por composição colorida é a transformação IHS (*Intensity, Hue, Saturation*), que envolve uma decomposição de uma imagem RGB em componentes de intensidade, matiz e saturação.

7.1.3 Divisão de bandas

Esta técnica consiste na divisão do valor digital dos pixels de uma banda pelos correspondentes valores de outra banda. Ao se efetuar uma razão entre bandas, os quocientes variam em um intervalo que compreende valores reais contínuos. Para a discretização desses valores multiplicam-se os quocientes por um "ganho" e adiciona-se um "*off-set*", cujos valores ideais, do ganho e do *off-set*, variam de acordo com a imagem e com o tipo de "*ratio*" (divisão de bandas ou

razão de canais). Esses valores devem atribuir à imagem resultante uma maior variância possível dos níveis de cinza (números digitais) sem saturá-la, e a média deve estar próxima da média do intervalo máximo dos valores digitais da imagem. Essa técnica possui a vantagem de atenuar os efeitos multiplicativos de ruídos, além de enfatizar a separação dos alvos com comportamento de gradiente diferente nas curvas de refletância. Possui, também, a capacidade de reduzir a dimensão dos dados, ou seja, as informações de quatro bandas podem ser obtidas através de uma única composição colorida, usando-se três imagens *ratio*. Porém, possui a desvantagem de perder as características espaciais da cena, devido à atenuação das influências de iluminação, além de atenuar a discriminação de alvos com comportamento de gradiente semelhante nas curvas de refletância, e perder as informações espectrais originais.

7.2 PhotoDoc - normalização da iluminação

Esta seção apresenta um algoritmo de realce para imagens de documentos fotografados. Este algoritmo é baseado no algoritmo apresentado em [57], onde as diferenças são dadas na definição do tamanho da janela e classificação dos blocos. O algoritmo calcula o fator de iluminação e o divide pelo valor da imagem original, obtendo assim uma figura sem a interferência da luz. Os passos deste algoritmo são descritos a seguir:

- Dividir a imagem em nove blocos de igual tamanho, binariza-se o bloco central com o algoritmo de Sauvola [56] e estima-se o tamanho médio dos caracteres do texto (*projection profiler*), para definição de um bloco de tamanho fixo;
- Para cada um desses blocos, as cores dos *pixels* são armazenas sendo essas ordenadas de maneira crescente de acordo com o valor da luminância. A Equação 7.5 para o cálculo da luminância *L*, muito usado para converter *pixels* coloridos em tons de cinza, é dada da seguinte forma:

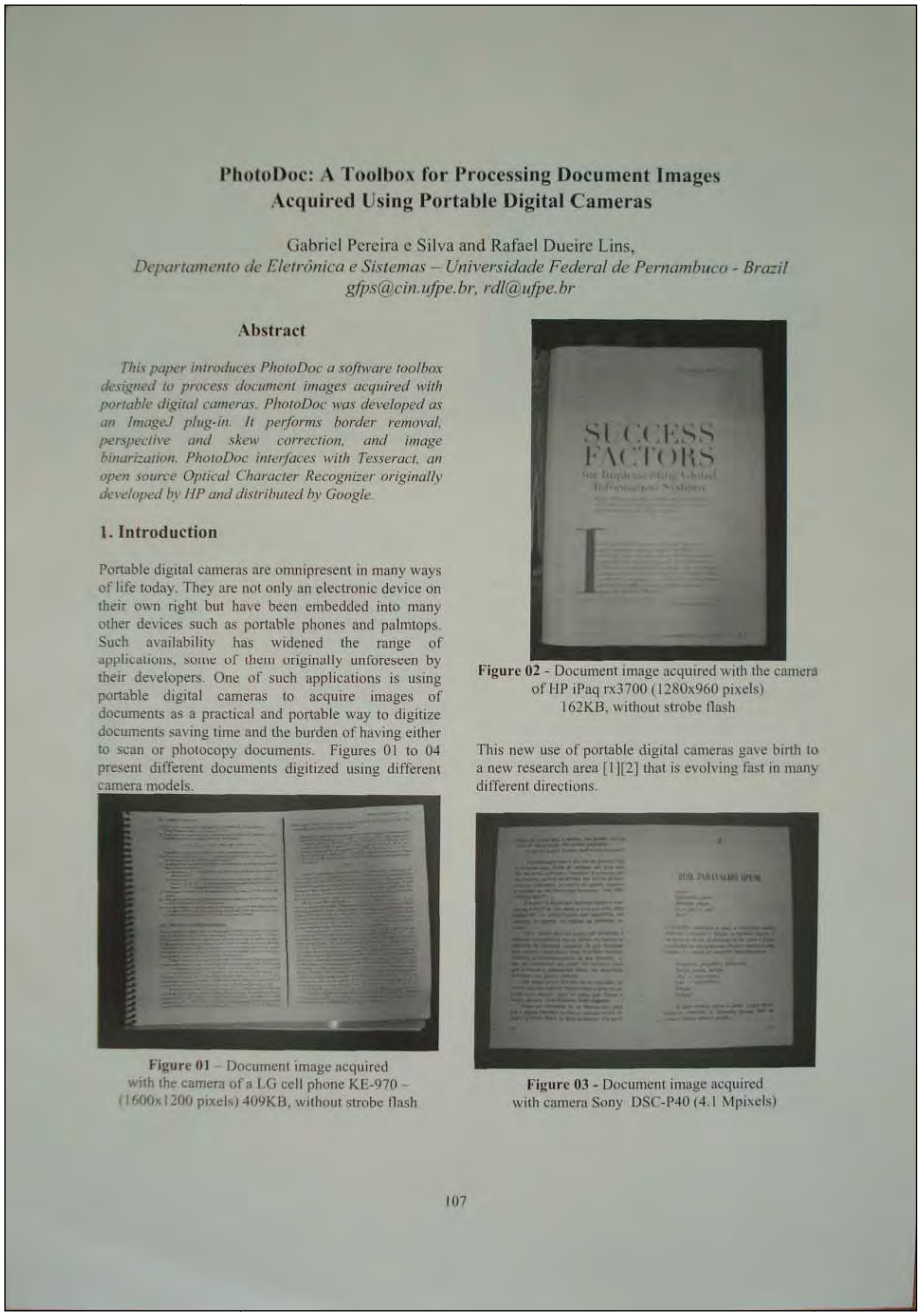
$$L = 0{,}30 \times R + 0{,}59 \times G + 0{,}11 \times B \ ; \tag{7.5}$$

- Calcula-se o valor médio de cada componente *R*, *G* e *B* de 25% dos pixels com maior luminância. Essa média irá definir o valor do fundo local C_{luz}, em cada uma dessas componentes;
- Removem-se blocos que estão fora da média em relação a média de todos os outros blocos (tolerância de 25%). Isso ocorre quando o bloco engloba um caractere ou parte dele, grande o suficiente para interferir no cálculo do fundo local. Essas cores são rejeitadas, pois são desconformes em relação ao fundo;
- Aplica-se sobre cada pixel da imagem a função *S* descrita na Equação 7.6,

$$S(x,y) = \begin{cases} 1 \ se \ x \geq 1 \\ 0{,}5 - 0{,}5 \times \cos\left(x^p\right) \ se \ x < 1 \end{cases} \tag{7.6}$$

Onde *x* é dado pela razão $C_{entrada} / C_{luz}$, onde $C_{entrada}$ é o valor do pixel original e *p* é um parâmetro variável, a partir de dados experimentais sugere-se *p* = 0,7.

O resultado da aplicação desse algoritmo sobre a Figura 7.7 é ilustrado na Figura 7.8.



PhotoDoc: A Toolbox for Processing Document Images Acquired Using Portable Digital Cameras

Gabriel Pereira e Silva and Rafael Dueire Lins,
Departamento de Eletrônica e Sistemas – Universidade Federal de Pernambuco - Brazil
gfps@cin.ufpe.br, rdl@ufpe.br

Abstract

This paper introduces PhotoDoc a software toolbox designed to process document images acquired with portable digital cameras. PhotoDoc was developed as an ImageJ plug-in. It performs border removal, perspective and skew correction, and image binarization. PhotoDoc interfaces with Tesseract, an open source Optical Character Recognizer originally developed by HP and distributed by Google.

1. Introduction

Portable digital cameras are omnipresent in many ways of life today. They are not only an electronic device on their own right but have been embedded into many other devices such as portable phones and palmtops. Such availability has widened the range of applications, some of them originally unforeseen by their developers. One of such applications is using portable digital cameras to acquire images of documents as a practical and portable way to digitize documents saving time and the burden of having either to scan or photocopy documents. Figures 01 to 04 present different documents digitized using different camera models.

Figure 01 – Document image acquired with the camera of a LG cell phone KE-970 – (1600x1200 pixels) 409KB, without strobe flash

Figure 02 - Document image acquired with the camera of HP iPaq rx3700 (1280x960 pixels) 162KB, without strobe flash

This new use of portable digital cameras gave birth to a new research area [1][2] that is evolving fast in many different directions.

Figure 03 - Document image acquired with camera Sony DSC-P40 (4.1 Mpixels)

107

Sony Cyber-shot 7.2 Mega Pixels modelo DSC-W55, com lentes Carl-Zeiss Vario-Tessar 2.8-5.2/6.3-18,9

Figura 7.7 - Imagem fotografa corrigida a perspectiva e removida as bordas.

PhotoDoc: A Toolbox for Processing Document Images Acquired Using Portable Digital Cameras

Gabriel Pereira e Silva and Rafael Dueire Lins,
Departamento de Eletrônica e Sistemas – Universidade Federal de Pernambuco - Brazil
gfps@cin.ufpe.br, rdl@ufpe.br

Abstract

This paper introduces PhotoDoc a software toolbox designed to process document images acquired with portable digital cameras. PhotoDoc was developed as an ImageJ plug-in. It performs border removal, perspective and skew correction, and image binarization. PhotoDoc interfaces with Tesseract, an open source Optical Character Recognizer originally developed by HP and distributed by Google.

1. Introduction

Portable digital cameras are omnipresent in many ways of life today. They are not only an electronic device on their own right but have been embedded into many other devices such as portable phones and palmtops. Such availability has widened the range of applications, some of them originally unforeseen by their developers. One of such applications is using portable digital cameras to acquire images of documents as a practical and portable way to digitize documents saving time and the burden of having either to scan or photocopy documents. Figures 01 to 04 present different documents digitized using different camera models.

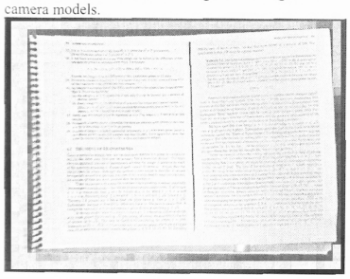


Figure 01 – Document image acquired with the camera of a LG cell phone KE-970 – (1600x1200 pixels) 409KB, without strobe flash

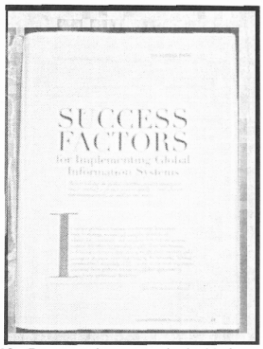


Figure 02 - Document image acquired with the camera of HP iPaq rx3700 (1280x960 pixels) 162KB, without strobe flash

This new use of portable digital cameras gave birth to a new research area [1][2] that is evolving fast in many different directions.

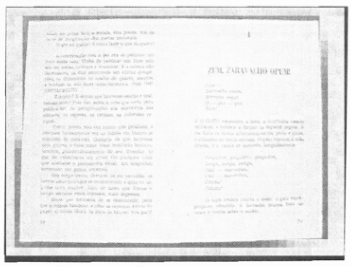


Figure 03 - Document image acquired with camera Sony DSC-P40 (4.1 Mpixels)

Figura 7.8 - Resultado do Realce da Figura 7.7.

Capítulo 8
A binarização de documentos

Este capítulo apresenta um estudo sobre a binarização de documentos fotografados por meio de câmeras digitais portáteis. Um total de dezoito algoritmos de binarização foi testado sobre as imagens fotografadas e as respectivas imagens escaneadas.

8.1 Binarização de imagens digitais

A idéia da binarização é reduzir a partir de um dado limiar (*thresholding*) o espaço de cores de uma imagem para um espaço binário, ou seja, apenas dois valores possíveis para cada *pixel* seram adotados. No âmbito de processamento de imagens é desejável que essas imagens apresentem apenas a cor preta e branca para o caso da imagem que contenham elementos textuais (texto, tabela, fluxo gramas, etc), já que as informações contidas por elementos não textuais (imagens, textura de fonte, etc.) seriam reduzidas drasticamente. Por isso essa transformação é mais indicada em imagens de documentos onde a informação predominante é textual. O nível de cinza de uma imagem pode ser definido como a distribuição dos tons de cinza em uma imagem monocromática (escala de cinza). A conversão de uma imagem colorida em imagem monocromática pode ser realizada através de diferentes cálculos, inclusive substituindo-se os valores das componentes R (vermelha) e B (azul), pelo valor da componente G (verde), entretanto, a fórmula mais conhecida para o cálculo do tom de cinza [8][10] para conversão de um pixel no padrão RGB é dada pela Equação 7.5, vista no Capítulo anterior.

Os algoritmos de binarização podem ser classificados em: globais e adaptativos locais. Os algoritmos globais estabelecem um valor de limiar, aplicando esse valor à imagem como um todo, pode-se fazer uma equivalência a um filtro passa-faixa onde se elimina os tons de cinza abaixo do limiar (cor branca) e os tons restantes para o símbolo que representa a cor preta. Já os algoritmos locais adaptativos possuem um valor inicial de limiar, podendo este ser alterados para cada localidade da imagem.

A aplicação de métodos de binarização desenvolvidos para atuação em *scanners* geralmente não produz bons resultados quando aplicados diretamente a documentos fotografados por câmeras digitais portáteis, pois eles são muito suscetíveis a variações abruptas nas imagens. A aplicação de algoritmos locais adaptativos produz melhores resultados do que os algoritmos globais. Essa diferença de desempenho é sobretudo decorrente dos problemas de digitalização, os quais foram descritos no Capítulo 2. Já que eles produzem uma maior variação na imagem, a definição de um

limiar adaptativo local tende a ser mais eficiente. Porém existe certa dificuldade em estimar o tamanho ideal da localidade usada para calcular esse tipo de *threshold*.

A necessidade de binarização de documentos digitalizados por *scanners* levou ao desenvolvimento de muitos algoritmos [43]. Tais algoritmos são utilizados como referência para o desenvolvimento de algoritmos mais eficientes no caso da digitalização por câmeras digitais [40][42]. A questão principal relacionada ao problema de binarização é como escolher o critério e os parâmetros de separação (*threshold*) mais adequados para a solução do problema.

Basicamente podemos classificar os algoritmos de binarização em seis categorias:

- Métodos baseados no formato do histograma: os picos, vales e curvaturas dos histogramas suavizados são analisados;
- Métodos baseados em agrupamento: as amostras de tons de cinza são agrupadas em plano de fundo e primeiro plano (objetos);
- Métodos baseados em entropia: usam a entropia do plano de fundo e do primeiro plano ou a entropia entre a imagem original e a binarizada;
- Métodos baseados nos atributos dos objetos: procuram uma métrica de similaridade entre as imagens em tons de cinza e as binarizadas;
- Métodos espaciais: utilizam a distribuição da probabilidade de primeira ordem e correlação entre os pixels;
- Métodos locais: adaptam o valor do *threshold* para cada pixel em cada região da imagem;
- Métodos neurais: esses podem ser vistos como métodos locais adaptativos, onde o valor do pixel final é determinador por uma rede neural.

Em [56] publicado no Journal of Electronic Imaging apresenta 40 dos principais algoritmos de *thresholding* descritos na literatura, no entanto não trata de documentos fotografados. Já em [92] pode-se encontrar os mais eficientes algoritmos para binarização de documentos fotografados, apresentados conforme a sua eficiência.

8.2 Algoritmos de binarização aplicados a documentos fotografados

Imagens binarizadas de documentos são visualmente mais “confortáveis” para o leitor, requerem menos *toner* para impressão, ocupam menos banda dos canais de comunicação e algumas ferramentas de OCR processam apenas imagens binárias, como é o caso do Tesseract [53]. Por esses motivos, realizou-se esta etapa do pré-processamento com o objetivo de analisar possíveis melhorias na qualidade do documento digitalizado.

O experimento testou dezoito algoritmos de limiarização (*thresholding*), sendo onze globais e sete locais. Para avaliação dessas técnicas optou-se por usar 100 imagens fotografadas adquiridas

com o uso do planetário, com angulação de 0° em relação à normal do plano que continha o documento, no formato *JPG*, juntamente com as suas equivalentes escaneadas a 300 *dpi* no formato *JPG*. Essas imagens foram adquiridas dos anais do CBDAR 2007 e as imagens de comparação foram geradas a partir do Adobe Acrobat 8.0 Pro no formato *PNG* binário, a uma resolução de 200 *dpis*. O algoritmo de correção de perspectiva sofreu uma pequena modificação no cálculo da largura e altura final apresentada no Capítulo anterior (Equação 6.17), para realizar o alinhamento entre o documento fotografado e o gerado a partir do arquivo *pdf*. A avaliação se deu por meio da análise do PSNR (*Peak-to-Signal Noise Ratio*) em relação às imagens geradas a partir do arquivo "pdf". O PSNR é uma medida de quão próxima é uma imagem de outra comparando *pixel* a *pixel*. Conseqüentemente, quanto mais elevado o valor do PSNR, mais elevada é a similaridade entre duas imagens.

$$PSNR = 10\log\left(\frac{C^2}{MSE}\right) \quad e \quad MSE = \frac{\sum_{x=1}^{M}\sum_{y=1}^{N}\left(I(x,y) - I'(x,y)\right)^2}{MN}, \text{ (ver referência [92])} \qquad (8.1)$$

onde *C* é dado como a diferença entre o *foreground* e o *background.*

A presença de figuras dentro dos documentos é um fato que degrada a análise dos caracteres, para contornar esse problema um pré-processamento foi realizado sobre essas imagens buscando eliminar os blocos pertencentes a figuras. As Figuras 8.1 a 8.6 apresentam um conjunto de imagens usadas por esse procedimento. A Tabela 8.1 apresenta o resultado do PSNR para cada algoritmo, submetido a um conjunto de 100 imagens fotografadas a 5.1 e 7.2 *Mpixels* com o auxílio do planetário a 0°, na altura baixa e com e sem uso de flash (+Fs e -Fs).

	Algoritmos	**5.1 *Mpixels***		**7.2 *Mpixels***		***Scanner***
		-Fs	**+Fs**	**-Fs**	**+Fs**	**####**
Globais	MelloLins_Threshold [30]	4,51	9,73	10,14	10,31	18,04
	KapurSahooWong_Threshold [21]	8,78	7,73	7,80	7,75	14,81
	WuSongdeHanqing_Threshold[45]	7,96	5,99	7,48	8,32	14,77
	Otsu_Threshold [34]	4,62	6,03	8,03	8,03	15,85
	Pun_Threshold[36]	4,47	8,5	8,11	8,73	13,77
	Yen_ChangChang_Threshold[47]	4,91	5,41	9,54	9,91	18,87
	SLR_Improved_Threshold [8]	7,62	**9,76**	**11,83**	**12,03**	**18,94**
	Kavallieratou_Antonopoulou_Threshold [100]	7,81	8,37	9,5	9,5	16,91
	Khashman_Sekeroglu_Threshold [102]	**9,13**	9,29	11,17	11,41	17,73
	RidlerCalvard_Threshold [104]	7,67	8,94	10,49	10,76	15,78
	Kittler_Illingworth_Threshold [105]	5,18	8,28	8,53	9,37	15,87
	Média:	**6,60**	**8,0**	**9,32**	**9,64**	**16,48**
Locais	Niblack [106]	9,93	10,38	11,94	12,29	15,75
	Sauvola_Pietaksinen [107]	14,98	15,48	15,90	16,10	17,91
	WhiteRohrer [108]	13,17	15,41	16,02	16,28	17,88
	Palumbo_Swaminathan_Srihari [103]	14,07	**15,73**	**16,11**	16,38	17,89
	Bernsen [101]	13,04	14,18	14,72	15,29	16,89
	Oliveira_Lins [93]	**15,45**	15,59	15,91	**16,48**	**18,91**
	MR_Otsu [109]	8,21	9,93	15,27	15,71	18,31
	Média:	**12,69**	**13,81**	**15,12**	**15,5**	**17,64**

Tabela 8.1 - Análise da binarização por PSNR.

É possível observar que dentre os algoritmos globais o que apresentou melhor resultado médio foi o da_Silva-Lins-Rocha [7], enquanto o Oliveira_Lins [93] mostrou-se melhor entre os locais. Em geral os algoritmos locais tiveram um resultado melhor, porém os globais também produziram bons resultados, entretanto as falhas que surgem na imagem devido ao problema de iluminação não uniforme podem causar um aumento na quantidade de erros nas transcrições por ferramentas de OCR [49]. Outro fator a ser observado é a degradação do desempenho dos algoritmos quando aplicado a imagens fotografadas, essa degradação se deve ao fato que em algumas imagens existe a presença de sombras, fraca iluminação ambiente ou até mesmo a influência do uso de flash.

Buscando-se avaliar o fator de degradação dos algoritmos de binarização em imagens de documentos adquiridas por meio de câmeras digitais portáteis, aplicou-se o algoritmo de realce descrito no Capítulo 7 sobre o mesmo conjunto de imagens utilizadas durante o experimento anterior. A Tabela 8.2 apresenta o resultado do PSNR para cada algoritmo de binarização aplicado após o realce dessas imagens.

	Algoritmos	**5.1 *Mpixels***		**7.2 *Mpixels***		***Scanner***
		-Fs	**+Fs**	**-Fs**	**+Fs**	**####**
Globais	MelloLins_Threshold [30]	11,41	**12,52**	12,81	14,81	18,04
	KapurSahooWong_Threshold [21]	9,71	10,11	8,86	10,86	14,81
	WuSongdeHanqing_Threshold[45]	9,87	9,88	10,33	12,33	14,77
	Otsu_Threshold [34]	10,23	12,3	13,37	14,22	15,85
	Pun_Threshold[36]	8,02	8,81	8,11	10,11	13,77
	Yen_ChangChang_Threshold[47]	7,43	10,6	12,59	14,99	18,87
	SLR_Improved_Threshold [8]	**11,19**	11,9	**13,57**	**15,53**	**18,94**
	Kavallieratou_Antonopoulou_Threshold [100]	9,21	10,76	11,41	13,67	16,91
	Khashman_Sekeroglu_Threshold [102]	10,26	10,97	12,33	15,42	17,73
	RidlerCalvard_Threshold [104]	10,17	10,61	13,42	14,57	15,78
	Kittler_Illingworth_Threshold [105]	10,24	10,69	10,74	12,74	15,87
	Média:	**9,7**	**10,83**	**11,59**	**13,56**	**16,48**
Locais	Niblack [106]	10,8	10,81	12,12	13,62	15,75
	Sauvola_Pietaksinen [107]	15,9	15,7	**16,51**	**17,46**	17,91
	WhiteRohrer [108]	13,8	14,41	16,27	16,64	17,88
	Palumbo_Swaminathan_Srihari [103]	14,91	**15,73**	16,45	16,05	17,89
	Bernsen [101]	13,33	14,57	15,07	15,96	16,89
	Oliveira_Lins [93]	**15,44**	15,7	15,04	16,34	**18,91**
	MR_Otsu [109]	10,73	11,0	15,34	17,37	18,31
	Média:	**13,55**	**13,98**	**15,25**	**16,20**	**17,64**

Tabela 8.2 - Análise da binarização por PSNR após aplicação de Realce.

Conclui-se que a aplicação do realce (normalização de iluminação) sobre as imagens fotografadas trouxe ganhos expressivos aos algoritmos globais enquanto que nos locais o ganho foi um pouco menor. Observa-se ainda uma aproximação dos resultados do algoritmo de Sauvola_Pietaksinen [56] quando sobre as imagens fotografadas em relação as imagens escaneadas. Por outro lado, o algoritmo Oliveira_Lins [56] apresentou perda de desempenho, devido ao fato que esse algoritmo faz uso da variação de iluminação para definir os valores dos pixels da imagem binária.

PhotoDoc: A Toolbox for Processing Document Images
Acquired Using Portable Digital Cameras

Gabriel Pereira e Silva and Rafael Dueire Lins,
Departamento de Eletrônica e Sistemas Universidade Federal de Pernambuco - Brazil
gfps@cin.ufpe.br, rdl@ufpe.br

Abstract

This paper introduces PhotoDoc a software toolbox designed to process document images acquired with portable digital cameras. PhotoDoc was developed as an ImageJ plug-in. It performs border removal, perspective and skew correction, and image binarization. PhotoDoc interfaces with Tesseract, an open source Optical Character Recognizer originally developed by HP and distributed by Google.

1. Introduction

Portable digital cameras are omnipresent in many ways of life today. They are not only an electronic device on their own right but have been embedded into many other devices such as portable phones and palmtops. Such availability has widened the range of applications, some of them originally unforeseen by their developers. One of such applications is using portable digital cameras to acquire images of documents as a practical and portable way to digitize documents saving time and the burden of having either to scan or photocopy documents. Figures 01 to 04 present different documents digitized using different camera models.

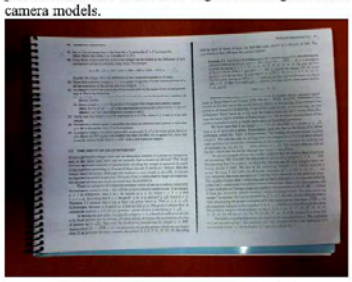


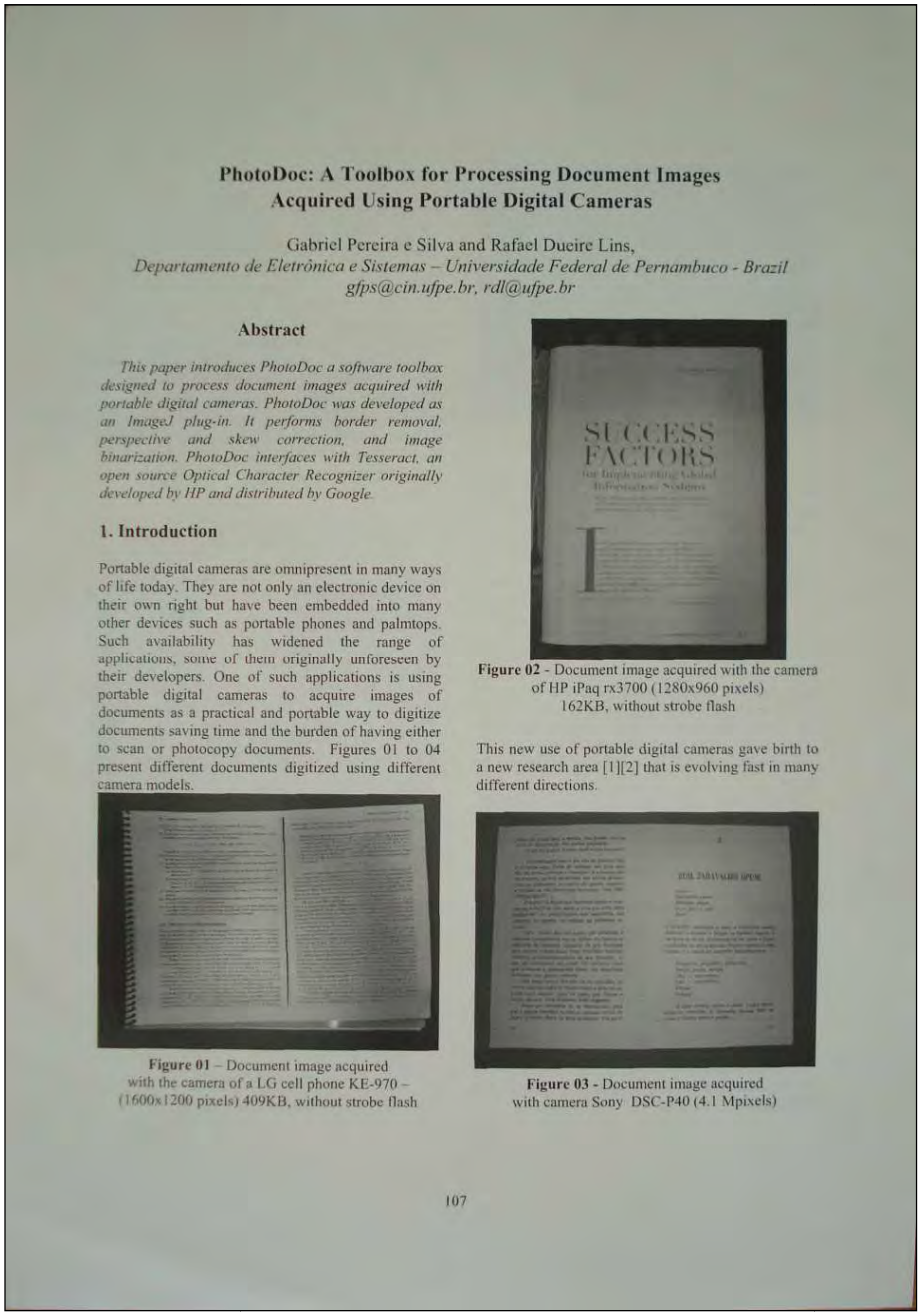
Figure 01 Document image acquired with the camera of a LG cell phone KE-970 (1600x1200 pixels) 409KB, without strobe flash

Figure 02 - Document image acquired with the camera of HP iPaq rx3700 (1280x960 pixels) 162KB, without strobe flash

This new use of portable digital cameras gave birth to a new research area [1][2] that is evolving fast in many different directions.

Figure 03 - Document image acquired with camera Sony DSC-P40 (4.1 Mpixels)

Figura 8.1 - Imagem gerada pelo Adobe Acrobat no formato (png).



PhotoDoc: A Toolbox for Processing Document Images Acquired Using Portable Digital Cameras

Gabriel Pereira e Silva and Rafael Dueire Lins,
Departamento de Eletrônica e Sistemas – Universidade Federal de Pernambuco - Brazil
gfps@cin.ufpe.br, rdl@ufpe.br

Abstract

This paper introduces PhotoDoc a software toolbox designed to process document images acquired with portable digital cameras. PhotoDoc was developed as an ImageJ plug-in. It performs border removal, perspective and skew correction, and image binarization. PhotoDoc interfaces with Tesseract, an open source Optical Character Recognizer originally developed by HP and distributed by Google.

1. Introduction

Portable digital cameras are omnipresent in many ways of life today. They are not only an electronic device on their own right but have been embedded into many other devices such as portable phones and palmtops. Such availability has widened the range of applications, some of them originally unforeseen by their developers. One of such applications is using portable digital cameras to acquire images of documents as a practical and portable way to digitize documents saving time and the burden of having either to scan or photocopy documents. Figures 01 to 04 present different documents digitized using different camera models.

Figure 01 – Document image acquired with the camera of a LG cell phone KE-970 – (1600x1200 pixels) 409KB, without strobe flash

Figure 02 - Document image acquired with the camera of HP iPaq rx3700 (1280x960 pixels) 162KB, without strobe flash

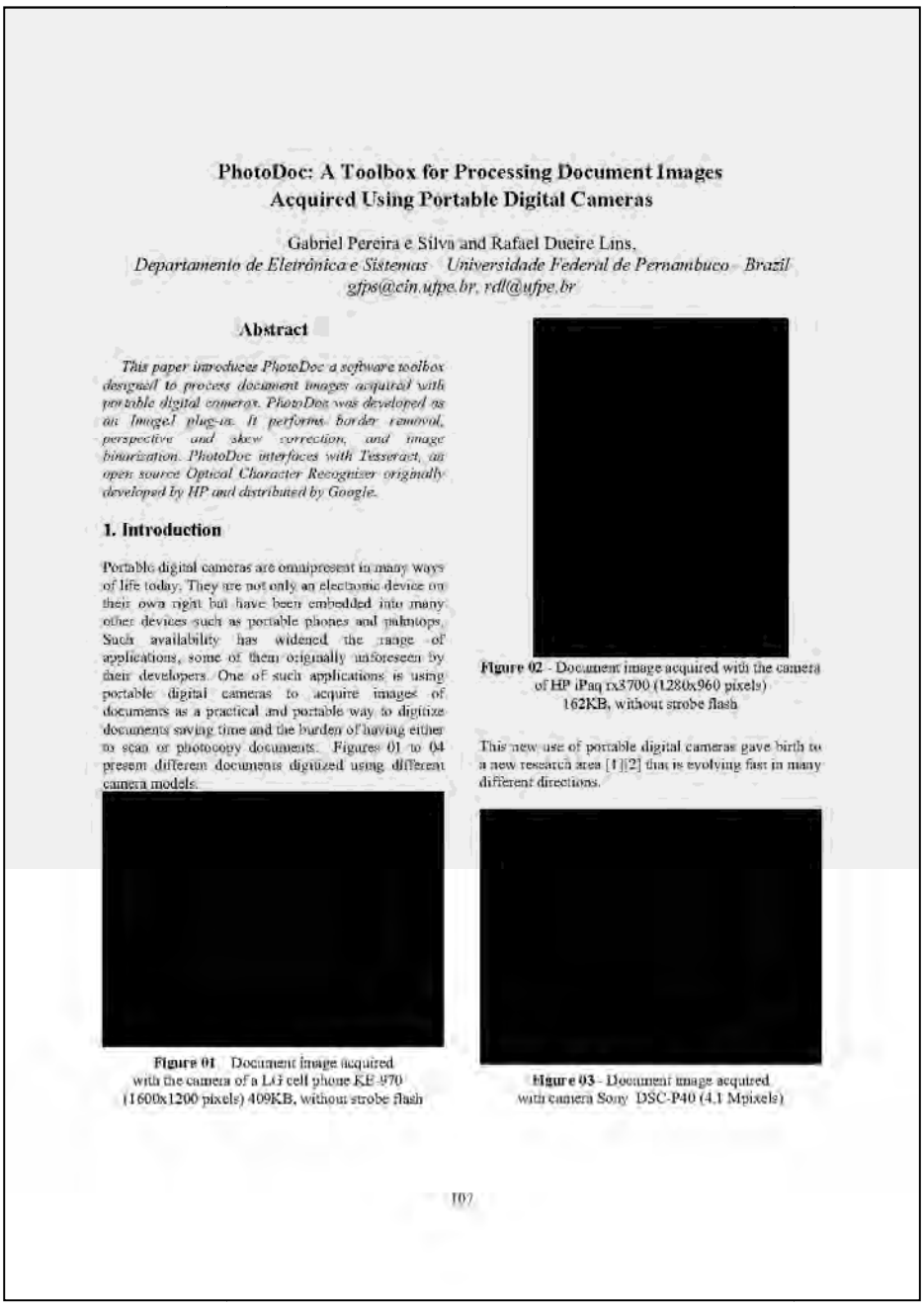
This new use of portable digital cameras gave birth to a new research area [1][2] that is evolving fast in many different directions.

Figure 03 - Document image acquired with camera Sony DSC-P40 (4.1 Mpixels)

107

Sony Cyber-shot 7.2 Mega Pixels modelo DSC-W55, com lentes Carl-Zeiss Vario-Tessar 2.8-5.2/6.3-18,9

Figura 8.2 - Imagem fotografada a 7.2 *Mpixels* adquirida por meio do planetário e processada pelo PhotoDoc.



PhotoDoc: A Toolbox for Processing Document Images Acquired Using Portable Digital Cameras

Gabriel Pereira e Silva and Rafael Dueire Lins,
Departamento de Eletrônica e Sistemas Universidade Federal de Pernambuco Brazil
gfps@cin.ufpe.br, rdl@ufpe.br

Abstract

This paper introduces PhotoDoc a software toolbox designed to process document images acquired with portable digital cameras. PhotoDoc was developed as an ImageJ plug-in. It performs border removal, perspective and skew correction, and image binarization. PhotoDoc interfaces with Tesseract, an open source Optical Character Recognizer originally developed by HP and distributed by Google.

1. Introduction

Portable digital cameras are omnipresent in many ways of life today. They are not only an electronic device on their own right but have been embedded into many other devices such as portable phones and palmtops. Such availability has widened the range of applications, some of them originally unforeseen by their developers. One of such applications is using portable digital cameras to acquire images of documents as a practical and portable way to digitize documents saving time and the burden of having either to scan or photocopy documents. Figures 01 to 04 present different documents digitized using different camera models.

Figure 01 Document image acquired with the camera of a LG cell phone KE-970 (1600x1200 pixels) 409KB, without strobe flash

Figure 02 - Document image acquired with the camera of HP iPaq rx3700 (1280x960 pixels) 162KB, without strobe flash

This new use of portable digital cameras gave birth to a new research area [1][2] that is evolving fast in many different directions.

Figure 03 - Document image acquired with camera Sony DSC-P40 (4.1 Mpixels)

107

Figura 8.3 - Imagem gerada pelo Adobe Acrobat no formato (png) binário.

PhotoDoc: A Toolbox for Processing Document Images Acquired Using Portable Digital Cameras

Gabriel Pereira e Silva and Rafael Dueire Lins,
Departamento de Eletrônica e Sistemas – Universidade Federal de Pernambuco - Brazil
gfps@cin.ufpe.br, rdl@ufpe.br

Abstract

This paper introduces PhotoDoc a software toolbox designed to process document images acquired with portable digital cameras. PhotoDoc was developed as an ImageJ plug-in. It performs border removal, perspective and skew correction, and image binarization. PhotoDoc interfaces with Tesseract, an open source Optical Character Recognizer originally developed by HP and distributed by Google.

1. Introduction

Portable digital cameras are omnipresent in many ways of life today. They are not only an electronic device on their own right but have been embedded into many other devices such as portable phones and palmtops. Such availability has widened the range of applications, some of them originally unforeseen by their developers. One of such applications is using portable digital cameras to acquire images of documents as a practical and portable way to digitize documents saving time and the burden of having either to scan or photocopy documents. Figures 01 to 04 present different documents digitized using different camera models.

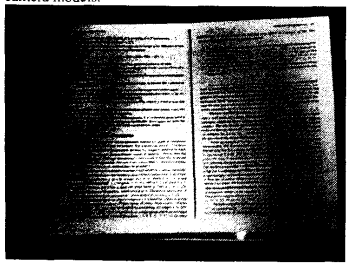


Figure 01 - Document image acquired with the camera of a LG cell phone KE-970 – (1600x1200 pixels) 409KB, without strobe flash

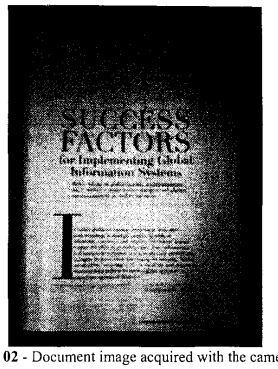


Figure 02 - Document image acquired with the camera of HP iPaq rx3700 (1280x960 pixels) 162KB, without strobe flash

This new use of portable digital cameras gave birth to a new research area [1][2] that is evolving fast in many different directions.

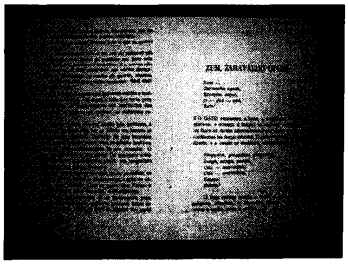


Figure 03 - Document image acquired with camera Sony DSC-P40 (4.1 Mpixels)

Figura 8.4 - Imagem da Figura 8.2 binarizada pelo algoritmo Silva_Lins_Rocha [8].

PhotoDoc: A Toolbox for Processing Document Images Acquired Using Portable Digital Cameras

Gabriel Pereira e Silva and Rafael Dueire Lins,
Departamento de Eletrônica e Sistemas – Universidade Federal de Pernambuco - Brazil
gfps@cin.ufpe.br, rdl@ufpe.br

Abstract

This paper introduces PhotoDoc a software toolbox designed to process document images acquired with portable digital cameras. PhotoDoc was developed as an ImageJ plug-in. It performs border removal, perspective and skew correction, and image binarization. PhotoDoc interfaces with Tesseract, an open source Optical Character Recognizer originally developed by HP and distributed by Google.

1. Introduction

Portable digital cameras are omnipresent in many ways of life today. They are not only an electronic device on their own right but have been embedded into many other devices such as portable phones and palmtops. Such availability has widened the range of applications, some of them originally unforeseen by their developers. One of such applications is using portable digital cameras to acquire images of documents as a practical and portable way to digitize documents saving time and the burden of having either to scan or photocopy documents. Figures 01 to 04 present different documents digitized using different camera models.

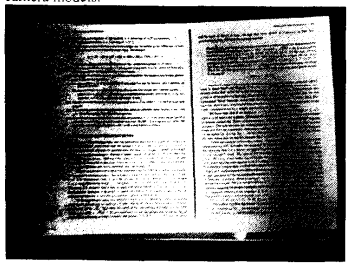


Figure 01 Document image acquired with the camera of a LG cell phone KE-970 (1600x1200 pixels) 409KB, without strobe flash

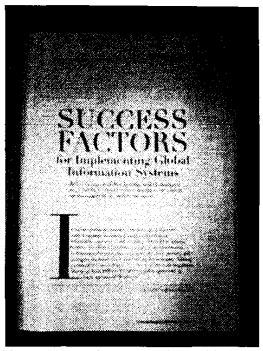


Figure 02 - Document image acquired with the camera of HP iPaq rx3700 (1280x960 pixels) 162KB, without strobe flash

This new use of portable digital cameras gave birth to a new research area [1][2] that is evolving fast in many different directions.

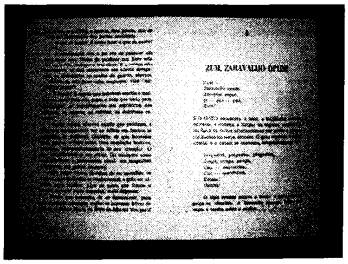


Figure 03 - Document image acquired with camera Sony DSC-P40 (4.1 Mpixels)

Figura 8.5 - Imagem da Figura 8.2 binarizada pelo algoritmo Oliveira_lins [93].

PhotoDoc: A Toolbox for Processing Document Images Acquired Using Portable Digital Cameras

Gabriel Pereira e Silva and Rafael Dueire Lins,
Departamento de Eletrônica e Sistemas Universidade Federal de Pernambuco - Brazil
gfps@cin.ufpe.br, rdl@ufpe.br

Abstract

This paper introduces PhotoDoc a software toolbox designed to process document images acquired with portable digital cameras. PhotoDoc was developed as an ImageJ plug-in. It performs border removal, perspective and skew correction, and image binarization. PhotoDoc interfaces with Tesseract, an open source Optical Character Recognizer originally developed by HP and distributed by Google.

1. Introduction

Portable digital cameras are omnipresent in many ways of life today. They are not only an electronic device on their own right but have been embedded into many other devices such as portable phones and palmtops. Such availability has widened the range of applications, some of them originally unforeseen by their developers. One of such applications is using portable digital cameras to acquire images of documents as a practical and portable way to digitize documents saving time and the burden of having either to scan or photocopy documents. Figures 01 to 04 present different documents digitized using different camera models.

Figure 01 Document image acquired with the camera of a LG cell phone KE-970 (1600x1200 pixels) 409KB, without strobe flash

Figure 02 - Document image acquired with the camera of HP iPaq rx3700 (1280x960 pixels) 162KB, without strobe flash

This new use of portable digital cameras gave birth to a new research area [1][2] that is evolving fast in many different directions.

Figure 03 - Document image acquired with camera Sony DSC-P40 (4.1 Mpixels)

Figura 8.6 - Imagem da Figura 8.3 após remoção dos blocos de imagem.

Capítulo 9
Resultados

O termo qualidade em imagens é muito abrangente. Podem-se encontrar tanto métodos baseados na percepção humana para identificar os fatores que influenciam na qualidade de imagens, quanto pesquisas baseadas em características das imagens em relação a padrões pré-estabelecidos. Mesmo restringindo o grupo de imagens a documentos digitalizados, ainda não se pode determinar um método que realize essa medição de maneira geral e satisfatória.

Nesta dissertação optou-se por analisar a qualidade dos documentos digitalizados por meio de câmeras digitais portáteis, usando ferramentas comerciais de transcrição automática. Essa escolha já era prevista, uma vez que nos primeiros Capítulos deste trabalho é exposto o interesse em melhorar a qualidade de documentos buscando um melhor desempenho das ferramentas de OCR. Como a transcrição automática de documentos é uma aplicação habitual, ela pode servir para medição de qualidade, não como índice, mas como fator de comparação [1].

No Capítulo 2 foram apresentados alguns desafios da digitalização de documentos por câmeras digitais portáteis, juntamente com uma breve comparação entre câmeras digitais e *scanners*. A partir dessas considerações podemos afirma que imagens de documentos escaneados tende a apresentarem uma melhor transcrição automática. Em resumo o objetivo desse Capítulo é justamente medir o efeito do processamento de imagens de documentos pelo PhotoDoc sobre a transcrição por ferramentas de OCR, comparando-se as imagens processadas pelo PhotoDoc e o mesmo documento digitalizado por escaner. Convém observar que este método de análise limita-se a documentos impressos.

9.1 Metodologia de medição de qualidade

Ao longo desta dissertação quatro problemas têm sido citados com freqüência como fatores indesejáveis na digitalização de documentos: iluminação irregular, presença de bordas, inclinações e distorções de perspectiva. Para medir-se o grau de influência desses quatros problemas faz-se necessário estabelecer critérios para estimar-se a qualidade destas imagens. Sabe-se que imagens adquiridas por *scanners* também podem possuir bordas ou inclinações, portanto podem ser utilizadas como conjunto de referência para medição da qualidade, assumindo-se que as imagens adquiridas pelo escaner apresentam boa qualidade.

Devido à extrema complexidade na atribuição de índices de qualidade, buscou-se apenas analisar a qualidade das saídas providas pela transcrição de documentos através de ferramentas comerciais de OCR. Para a transcrição automática dos documentos utilizou-se a ferramenta ABBYY FineReader 9.0 [95], sem o uso de dicionário, já que a intenção aqui é analisar a melhora do reconhecimento dos caracteres. A análise cumulativa da transcrição automática das imagens foi comparada com o texto extraído dos arquivos *pdfs* dos anais do CBDAR 2007, por meio da API Java

PDFBOX [94]. Foram utilizadas 100 imagens de documentos com texto na língua inglesa nas resoluções de 5.1 e 7.2 *Mpixels* no formato *JPEG* adquiridos com uma câmera digital Sony Cyber-shot 7.2 Mega Pixels modelo DSC-W55, com lentes Carl-Zeiss Vario-Tessar 2.8-5.2/6.3-18,9, juntamente com as respectivas imagens escaneadas a 100, 200 e 300 *dpis* nos formatos *TIFF*, *PNG* e *JPEG*.

Por fim todo o conjunto usado neste experimento seguiu o mesmo procedimento de ajuste descrito na subseção 8.1 (remoção dos blocos de imagens). Essa analise será realizada observando-se a quantidade total de caracteres de cada imagem, ou seja, será levado em consideração apenas os casos de substituições, ausências e inserções, para cada etapa do fluxograma do ambiente PhotoDoc. O resultado desta analise será apresentado na subseção 9.4 deste Capítulo.

9.2 Legibilidade e subjetividade

Devido ao sistema de reconhecimento visual dos seres humanos ser extremamente complexo, ainda não há consenso no método estabelecido pelo cérebro para identificação de caracteres. Evidências dos trabalhos nos últimos 20 anos indicam que os seres humanos utilizam letras de uma palavra para identificá-la como um todo [24]. Três categorias de modelos de reconhecimento de palavras são mais utilizados: o modelo de contorno das palavras, que sugere que palavras são reconhecidas como unidades completas, o modelo serial, que afirma que palavras são lidas letra-a-letra e o modelo paralelo de reconhecimento de letras, o qual sugere que letras de uma palavra são reconhecidas simultaneamente, e a informação das letras é utilizada para reconhecer a palavra, sendo esse o mais utilizado.

Apesar da complexidade dos algoritmos utilizados por ferramentas de OCR, ainda há muito a ser implementado para que estas sejam aptas a transpor documentos com níveis altos de ruído, distorções, baixa resolução, dentre outros problemas com os quais o olho humano lida com facilidade. Além disso, ferramentas de OCR atuam apenas em documentos binários [36], realizando esta binarização no caso de a entrada ser uma imagem colorida. Comparando com imagens binárias, sabe-se que a escala de cinza melhora a qualidade da imagem ao olho humano [34], portanto, com a popularização de câmeras digitais, busca-se a atuação de ferramentas de OCR em documentos em escala de cinza, de modo a lidar melhor com a baixa resolução [36].

A subjetividade tem papel importante na medição da qualidade de imagens. Em digitalização de acervos bibliográficos é comum a presença de um operador que qualifica a imagem como apta ou não, armazenando ou solicitando nova digitalização [38]. Um experimento realizado pelo autor no ano de 2007 entre alunos do Centro de Informática da Universidade Federal de Pernambuco (CIn-UFPE), envolvendo 80 alunos, classificando 369 imagens, mostra a relação entre o reconhecimento e classificação correta pelos alunos e o reconhecimento por uma ferramenta de OCR (Tabela 9.1). As imagens foram extraídas de fotografias com 4.1 *Mpixels*, e digitalizadas com escaner a 150 dpi e 300 dpi de resolução. Tais imagens contendo caracteres isolados, palavras ou frases, inserindo-se ruído de sal e pimenta (*salt-and-pepper noise)* e/ou borrando a imagem antes de realizar a binarização em

três níveis. A Tabela 9.1 leva à conclusão de que os resultados obtidos no reconhecimento de caracteres por ferramentas comerciais de OCR não podem ser utilizados como medição objetiva de qualidade, mas que seu reconhecimento com alta taxa de acerto é indicador de que o documento provavelmente é de boa qualidade.

Classificação	Reconhecimento OCR
Excelente	38,09%
Boa	19,05%
Regular	33,33%
Ruim	4,76%
Péssimo	0%

Tabela 9.1 - Reconhecimento do OCR em relação à classificação humana de qualidade.

A busca por um índice objetivo de qualidade tem levado pesquisadores a montar modelos de degradação em textos e sugerir valores [3] [4] [20] [47]. Muitos desses modelos são difíceis de serem validados, uma vez que seus referenciais são empíricos para a extração de indicadores de qualidade, havendo poucos métodos para validação [19]. Esta busca por um método quantitativo efetivo de medição de qualidade em documentos digitalizados tem procurado prover transposição mais eficiente por ferramentas de OCR. É comum ter como medição de qualidade a extração de características dos caracteres e ruídos presentes na imagem.

9.3 Pré-processamento e seus resultados

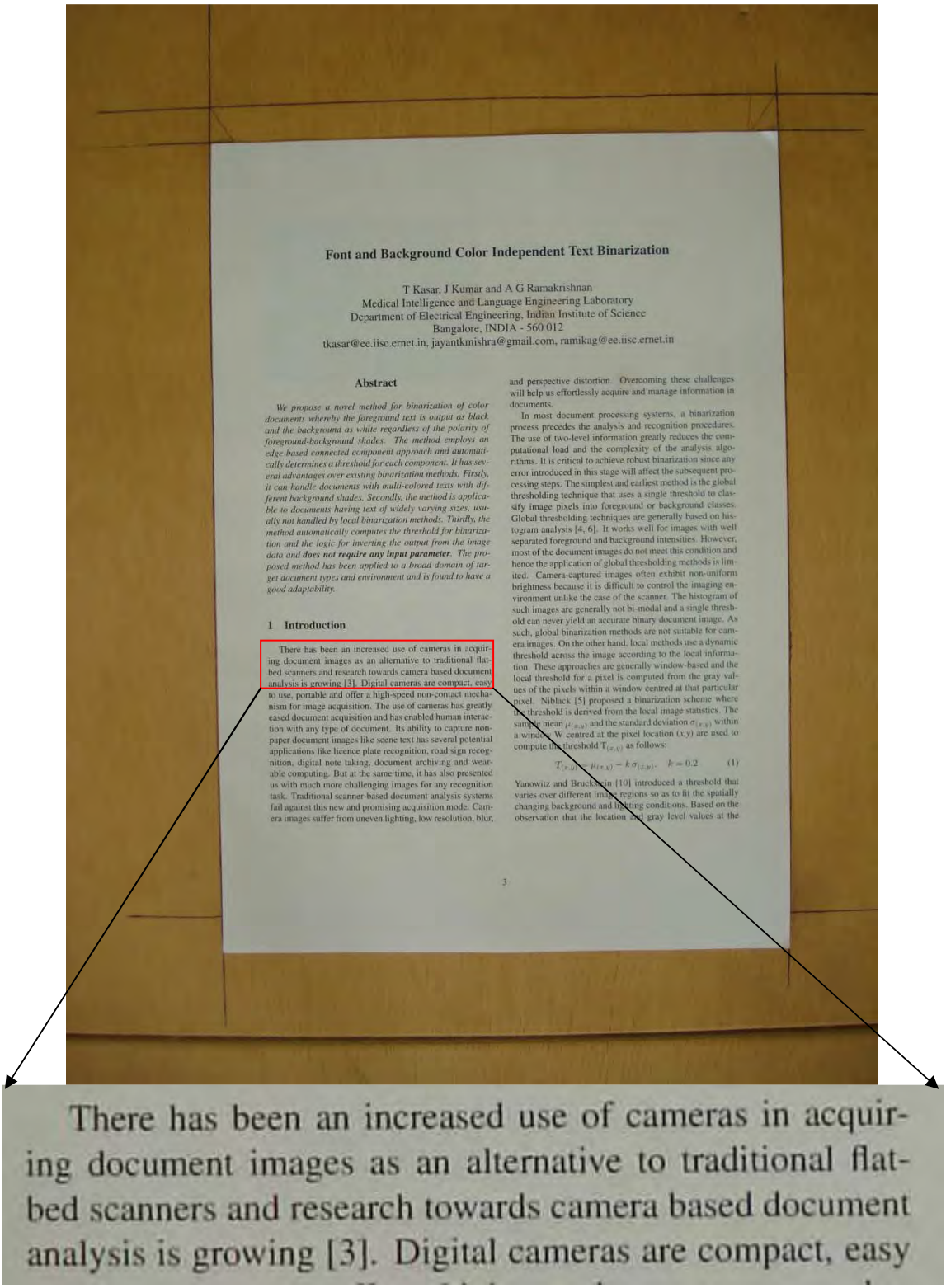
O pré-processamento dos documentos digitalizados por câmeras digitais busca melhorar a legibilidade por humanos, além de prover menor espaço para armazenamento e transmissão via rede de computadores, e melhorar a qualidade da saída em uma transcrição de imagem para texto. Nas subseções que seguem, são feitas medições e comparações, buscando identificar a influência de cada fator na qualidade da imagem.

9.3.1 Flash e iluminação inadequada

Apesar de os documentos obtidos tipicamente mostrarem diferentes regiões de brilho para um observador humano, e que isto impõe dificuldades para segmentação, binarização e remoção de bordas, a ferramenta de OCR utilizada não demonstra maiores erros em regiões de maior ou menor brilho desde que seja preservado contraste entre a fonte do documento e seu papel. Isto leva à conclusão que o algoritmo de binarização utilizado pela ferramenta de OCR é adaptativo.

Experimentos demonstraram que em ambientes com pouca iluminação, onde o efeito do flash torna-se facilmente visível, encontra-se maior quantidade de erros de transcrição, levando à conclusão que embora o algoritmo para binarização utilizado pela ferramenta de OCR seja adaptativo, não é suficientemente eficiente para contornar o problema criado pela iluminação irregular. A região da imagem que apresenta maior atuação do flash apresenta menores erros durante a transcrição, devido a

melhor iluminação, sendo que o flash demonstra ser ineficiente na aplicação ao documento como um todo. A Figura 9.1 ilustra a transposição de um trecho de menor brilho em um documento fotografado.



There has been an increased use uf cameras in acquirÂ»
ing document images as an altemative tc traditional Hatbed scanners and research towards camera based document
analysis is growing [3]. Digital cameras are compact easy

Figura 9.1 - Exemplo de transcrição em região da imagem por meio do Tesseract[53].

9.3.2 Correção da inclinação

Mesmo em documentos digitalizados por *scanners* é comum o surgimento de inclinação do documento devido ao fato de que nem sempre o documento é posto corretamente na superfície do scaner, seja pelo operador ou pela alimentação automática do escaner.

Para seres humanos a rotação de imagens é desagradável e introduz dificuldade na leitura do texto. A inclinação de documentos representa a inserção de diversos elementos prejudiciais à visão computacional, tais como maior espaço para armazenamento e maior captação de erros em reconhecimento e na transcrição de documentos por ferramentas de OCR [45]. Esses elementos fazem com que a correção da inclinação seja uma fase comum ao pré-processamento em qualquer ambiente de processamento de imagens.

No caso de documentos digitalizados por câmeras digitais sem suporte mecânico, este problema é presente em quase todos os documentos. Para essa medição, foram ignoradas as distorções geométricas, introduzidas pelas curvaturas das lentes das câmeras. Os vértices inferiores foram utilizados para traçar-se uma reta, da qual foi possível extrair-se essa inclinação.

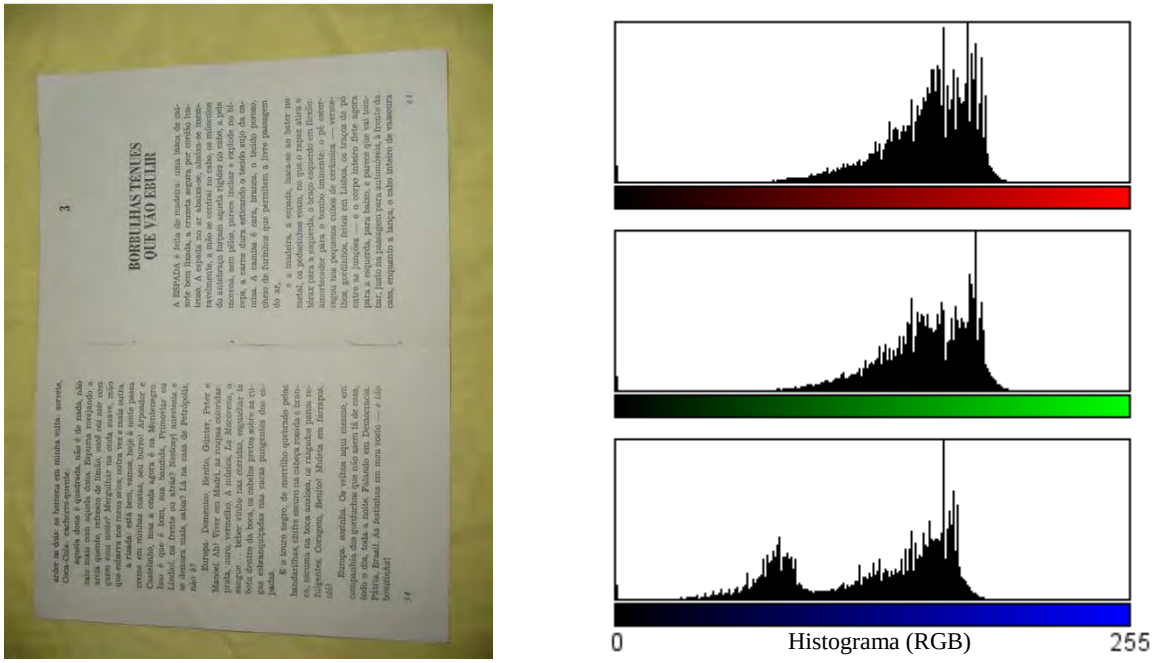
As maiorias das ferramentas comerciais de OCR compensam automaticamente inclinações em até 15 graus [27]. As distorções geométricas ou de perspectiva podem gerar pequenas variações no ângulo de inclinação do texto, em diferentes regiões.

9.3.3 Remoção de bordas

As desvantagens e problemas introduzidos pela presença de bordas em documentos digitalizados são os mesmos para scanners e câmeras digitais, embora a remoção de bordas nesses dois casos não seja relacionada. O fator principal é que o ambiente de digitalização por câmeras digitais é imprevisível.

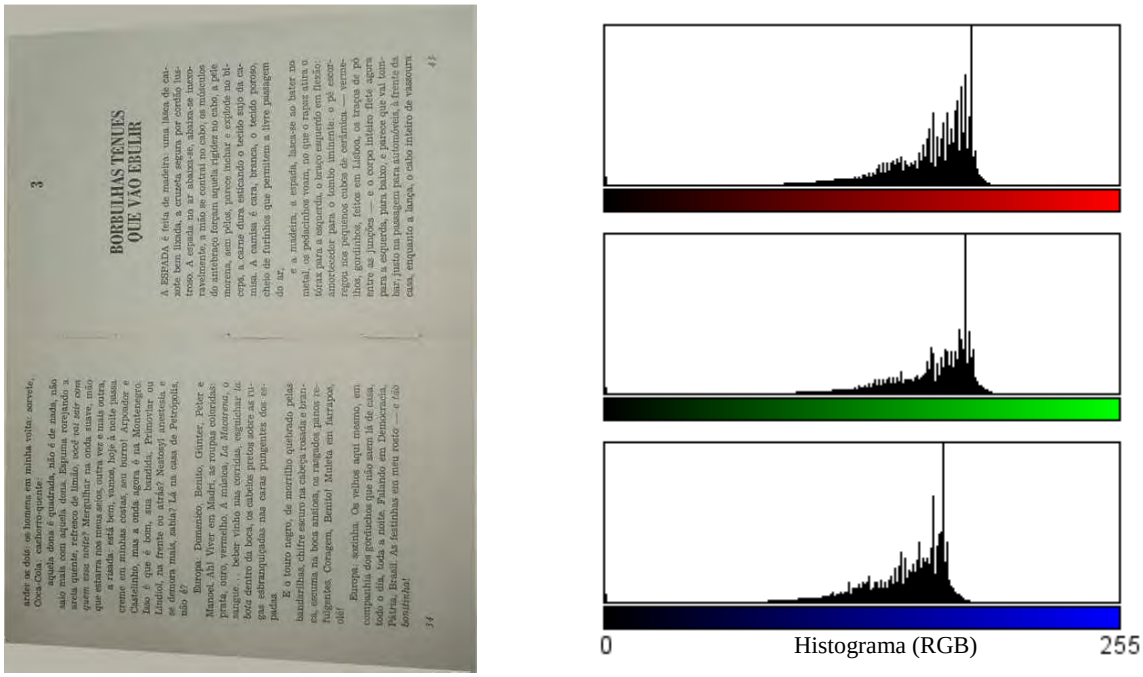
A presença de bordas em documentos adquiridos por câmeras digitais afeta diretamente os algoritmos de segmentação das ferramentas de OCR, gerando aumento na quantidade de erros em palavras e caracteres. Essa quantidade de erros varia dependendo da complexidade da borda. Para melhor análise dos documentos, realizou-se a remoção de bordas dos documentos como fase de pré-processamento. Devido à presença de distorções de perspectiva, o método utilizado para a remoção de bordas faz com que vestígios das bordas permaneçam na imagem. Estas bordas vestigiais foram substituídas pela cor mais freqüente na área central da imagem, conforme ilustrado nas Figuras 9.2 e 9.3. Através do histograma apresentado nas duas imagens percebe-se a redução na quantidade de cores na imagem cuja borda foi removida. A escolha da cor de substituição deu-se em virtude das técnicas conhecidas de binarização de documentos, pois o brilho das cores presentes nas regiões da imagem analisada afetam diretamente o valor do limiar de binarização.

Por exemplo, caso as bordas vestigiais fossem substituídas pela cor branca, o nível de limiar para binarização do documento seria mais alto, podendo levar à degradação dos caracteres próximos.



(Documento fotografado a mão-livre)

Figura 9.2 - Documento digitalizado com borda sem o uso de flash.



(Documento fotografado a mão-livre)

Figura 9.3 - Documento ilustrado na Figura 9.2 com borda remanescente substituída.

9.3.4 Distorções geométricas

O formato esférico das lentes de câmeras digitais e a proximidade do documento no momento da digitalização introduzem distorções em linhas retas que não são observadas em *scanners*. Nos documentos analisados foram verificadas as distâncias entre o limiar do documento e linhas retas traçadas entre as extremidades. As medições levaram à conclusão que a distorção máxima é

correspondente a um valor inferior a 3% do número de pixels em uma linha da imagem, apresentando média de 18 pixels para os documentos fotografados a 5.1 *Mpixels* e de 23 pixels para os documentos fotografados a 7.2 *Mpixels*. Devido ao fato dessas distorções estarem espalhadas ao longo dos documentos, não se observou maiores degradações nas extremidades dos documentos, região onde essas distorções são mais acentuadas. A Figura 9.4 ilustra como foi realizada a medição desta distorção, sendo medida a região destacada em verde.

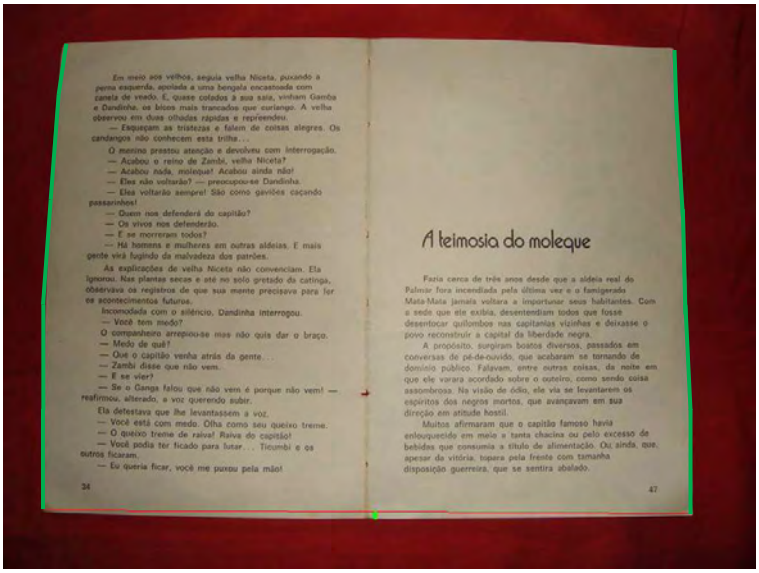


Figura 9.4 - Ilustração da distorção geométrica.

9.3.5 Distorções de perspectiva

A aquisição de documentos por câmeras digitais sem suporte mecânico paralelo ao plano do documento invariavelmente conduz a distorção de perspectiva. O objetivo desta etapa de pré-processamento foi medir a influência desta distorção no resultado provido pela ferramenta de OCR.

9.4 Análise da transcrição automática

Nesta seção são apresentados os resultados da transcrição automática pela ferramenta comercial ABBYY FineReader 9.0 Pro das imagens citadas na seção 9.1.

9.4.1 Alinhamento de seqüências

Para uma melhor análise da transcrição dessas imagens, os textos extraídos dessas imagens foram normalizados, ou seja, retiraram-se os espaços em branco e em seguida alinhados com os respectivos textos extraído dos arquivos *pdfs*. A idéia do alinhamento seguiu as seguintes etapas:

- Tomando-se linha de texto como uma seqüência de DNA onde cada caractere da língua inglesa foi definido como uma base nitrogenada [97], adaptou-se os algoritmos usados em bioinformática [97] para o alinhamento de seqüências de DNA;

- Armazena-se o custo do alinhamento de uma dada linha de texto e a posição dos caracteres alinhados em relação à seqüência obtida do arquivo PDF (original);
- Por fim calcula-se o menor custo da combinação dos diversos alinhamentos em relação à seqüência original.

Um exemplo de alinhamento é apresentado a seguir, onde se pode observar em vermelho os casos de substituição, em verde os casos de ausência e em azul os casos de inserção.

Therehasbeenanincreaseduseofcamerasinacquir-###
Therehasbeenanincreaseduseufcamerasinacquir#Â»

ingdocumentimagesasanalternativetotraditionalflat-
ingdocumentimagesasanalternativetctraditionalHat-

bedscannersandresearchtowardscamerabaseddocument
bedscannersandresearchtowardscamerabaseddocument

analysisisgrowing[3].Digitalcamerasarecompact easy
analysisisgrowing[3].Digitalcamerasarecompacteasy

Exemplo do alinhamento do texto transcrito pelo Tesseract [53] da Figura 9.1.

9.4.2 Análise Cumulativa

Após as análises realizadas no pré-processamento, os erros gerados na transposição das imagens para o formato de texto foram medidos e comparados. As Tabelas 9.3 a 9.8 ilustram os resultados obtidos em valor absoluto da transcrição automática pela ferramenta de OCR ABBYY FineReader 9.0, onde qualquer caractere presente no documento que não conste na transcrição, que seja substituído por outro, ou que seja inserido indevidamente na transcrição é classificado como erro de caractere.

Número de Imagens	100
Número de Caracteres (PDF)	351,382

Tabela 9.2 - Descrição do conjunto de dados.

100 DPI	*JPG*	*PNG*	*TIFF*
Substituição	11,569	11,866	11,557
Ausência	3,244	3,595	3,254
Inserção	10,352	10,070	10,070
200 DPI	*JPG*	*PNG*	*TIFF*
Substituição	7,532	7,501	7,537
Ausência	2,223	2,224	2,224
Inserção	8,784	8,784	8,783
300 DPI	*JPG*	*PNG*	*TIFF*
Substituição	7,327	7,325	7,325
Ausência	2,216	2,216	2,216
Inserção	8,784	8,791	8,785

Tabela 9.3 - Resultado da Transcrição pelo ABBYY FineReader 9.0 sobre as imagens escaneadas true color.

100 DPI	*JPG*	*PNG*	*TIFF*
Substituição	10,421	10,287	10,557
Ausência	3,301	3,291	3,254
Inserção	10,413	10,404	10,070
200 DPI	*JPG*	*PNG*	*TIFF*
Substituição	7,22	7,000	6,873
Ausência	2,281	2,282	2,001
Inserção	8,031	8,918	8,783
300 DPI	*JPG*	*PNG*	*TIFF*
Substituição	7,017	7,005	6,925
Ausência	2,281	2,290	2,916
Inserção	8,003	8,019	8,051

Tabela 9.4 - Resultado da Transcrição pelo ABBYY FineReader 9.0 sobre as imagens escaneadas binarizadas pelo algoritmo de Otsu [34].

	5.1 *Mpixels*		7.2 *Mpixels*	
0° Baixa	*+ Flash*	*-Flash*	*+ Flash*	*-Flash*
Substituição	9,454	9,870	9,163	9,380
Ausência	2,405	2,045	2,084	2,253
Inserção	9,699	10,115	9,144	9,153
0° Alta	*+ Flash*	*-Flash*	*+ Flash*	*-Flash*
Substituição	9,891	62866	54996	63867
Ausência	2,603	2,443	2,105	2,286
Inserção	10,033	10,445	9,401	9,631
15° Sul	*+ Flash*	*-Flash*	*+ Flash*	*-Flash*
Substituição	10,143	11,671	9,163	10,965
Ausência	3,075	3,964	3,050	3,948
Inserção	11,085	11,854	10,027	10,072
30° Sul	*+ Flash*	*-Flash*	*+ Flash*	*-Flash*
Substituição	12,041	13,198	10,129	10,678
Ausência	4,173	3,953	3,128	3,245
Inserção	12,039	12,889	10,882	10,944
15° Oeste	*+ Flash*	*-Flash*	*+ Flash*	*-Flash*
Substituição	13,138	12,342	12,136	11,308
Ausência	4,463	4,494	3,321	3,763
Inserção	11,685	13,254	10,223	10,365
30° Oeste	*+ Flash*	*-Flash*	*+ Flash*	*-Flash*
Substituição	13,854	13,840	12,910	12,113
Ausência	5,389	4,655	3,948	3,819
Inserção	12,429	13,669	10,041	10,716
Mão-Livre	*+ Flash*	*-Flash*	*+ Flash*	*-Flash*
Substituição	11,591	11,866	9,996	10,167
Ausência	2,603	2,911	2,173	2,304
Inserção	10,133	10,445	9,474	9,715

Tabela 9.5 - Resultado da Transcrição pelo ABBYY FineReader 9.0 sobre as imagens fotografadas.

	5.1 *Mpixels*		7.2 *Mpixels*	
0° Baixa	*+ Flash*	*-Flash*	*+ Flash*	*-Flash*
Substituição	9,454	9,870	9,063	9,380
Ausência	2,405	2,045	2,084	2,253
Inserção	8,872	9,037	8,096	8,153
0° Alta	*+ Flash*	*-Flash*	*+ Flash*	*-Flash*
Substituição	9,891	9,966	9,096	9,167
Ausência	2,603	2,443	2,105	2,286
Inserção	8,992	9,573	8,458	8,681
15° Sul	*+ Flash*	*-Flash*	*+ Flash*	*-Flash*
Substituição	10,143	11,671	9,163	10,965
Ausência	3,075	3,964	3,050	3,948
Inserção	9,101	10,244	8,251	8,988
30° Sul	*+ Flash*	*-Flash*	*+ Flash*	*-Flash*
Substituição	12,041	13,198	10,129	10,678
Ausência	4,173	3,953	3,128	3,245
Inserção	11,003	11,469	9,927	9,863
15° Oeste	*+ Flash*	*-Flash*	*+ Flash*	*-Flash*
Substituição	13,138	12,342	12,136	11,308
Ausência	4,463	4,494	3,321	3,763
Inserção	11,052	12,974	9,047	9,182
30° Oeste	*+ Flash*	*-Flash*	*+ Flash*	*-Flash*
Substituição	13,854	13,840	12,910	12,113
Ausência	5,389	4,655	3,948	3,819
Inserção	11,949	12,569	9,873	9,531
Mão-Livre	*+ Flash*	*-Flash*	*+ Flash*	*-Flash*
Substituição	11,591	11,866	9,996	10,167
Ausência	2,603	2,911	2,173	2,304
Inserção	10,133	10,445	9,474	9,715

Tabela 9.6 - Resultado da Transcrição pelo ABBYY FineReader 9.0 sobre as imagens fotografadas após processamento pelo PhotoDoc (Remoção de Bordas).

	5.1 *Mpixels*		7.2 *Mpixels*	
0° Baixa	*+ Flash*	*-Flash*	*+ Flash*	*-Flash*
Substituição	8,311	8,983	8,124	8,431
Ausência	2,302	2,315	2,209	2,219
Inserção	8,930	9,187	8,412	8,571
0° Alta	*+ Flash*	*-Flash*	*+ Flash*	*-Flash*
Substituição	8,871	8,966	8,526	8,777
Ausência	2,655	2,481	2,332	2,305
Inserção	8,992	9,573	8,724	8,847
15° Sul	*+ Flash*	*-Flash*	*+ Flash*	*-Flash*
Substituição	10,143	11,671	9,163	10,965
Ausência	3,075	3,964	3,050	3,948
Inserção	9,101	10,244	8,251	8,988
30° Sul	*+ Flash*	*-Flash*	*+ Flash*	*-Flash*
Substituição	10,157	11,087	9,383	10,011
Ausência	3,469	2,970	2,358	2,617
Inserção	11,151	11,510	10,014	9,993
15° Oeste	*+ Flash*	*-Flash*	*+ Flash*	*-Flash*
Substituição	11,011	11,974	10,241	10,348
Ausência	3,463	3,295	2,412	2,949
Inserção	11,052	12,974	9,047	9,182
30° Oeste	*+ Flash*	*-Flash*	*+ Flash*	*-Flash*
Substituição	11,204	11,874	10,541	10,922
Ausência	3,521	3,610	2,823	2,988
Inserção	12,091	12,694	9,995	9,987
Mão-Livre	*+ Flash*	*-Flash*	*+ Flash*	*-Flash*
Substituição	9,873	9,421	8,975	9,066
Ausência	2,411	2,727	2,012	2,127
Inserção	10,184	10,502	9,528	9,785

Tabela 9.7 - Resultado da Transcrição pelo ABBYY FineReader 9.0 sobre as imagens fotografadas após processamento pelo PhotoDoc (Correção Perspectiva+Remoção de Bordas).

	5.1 *Mpixels*		7.2 *Mpixels*	
0° Baixa	+ Flash	-Flash	+ Flash	-Flash
Substituição	7,521	7,873	7,219	7,533
Ausência	2,238	2,342	2,107	2,126
Inserção	8,943	9,198	8,411	8,595
0° Alta	+ Flash	-Flash	+ Flash	-Flash
Substituição	7,561	7,913	7,312	7,501
Ausência	2,547	2,399	2,101	2,284
Inserção	8,992	9,573	8,724	8,847
15° Sul	+ Flash	-Flash	+ Flash	-Flash
Substituição	9,517	9,891	8,315	9,051
Ausência	3,075	3,964	3,050	3,948
Inserção	9,101	10,244	8,251	8,988
30° Sul	+ Flash	-Flash	+ Flash	-Flash
Substituição	9,941	10,712	9,737	9,914
Ausência	3,469	2,970	2,358	2,617
Inserção	11,151	11,510	10,014	9,993
15° Oeste	+ Flash	-Flash	+ Flash	-Flash
Substituição	9,134	9,897	9,013	9,127
Ausência	3,463	3,295	2,412	2,949
Inserção	11,052	12,974	9,047	9,182
30° Oeste	+ Flash	-Flash	+ Flash	-Flash
Substituição	11,204	11,874	10,541	10,922
Ausência	3,521	3,610	2,823	2,988
Inserção	12,091	12,694	9,995	9,987
Mão-Livre	+ Flash	-Flash	+ Flash	-Flash
Substituição	7,873	7,421	7,875	7,959
Ausência	2,017	2,531	2,012	2,127
Inserção	10,342	10,701	9,915	9,972

Tabela 9.8 - Resultado da Transcrição pelo ABBYY FineReader 9.0 sobre as imagens fotografadas após correção de perspectiva, remoção da borda, realce e binarização.

É possível observar através das Tabelas 9.3 e 9.5 que a quantidade de erros nas transcrições das imagens de documentos adquiridas por câmera digital a 5.1 e 7.2 *Mpixels* foi menor do que as imagens adquiridas por scanner a 100 dpi e que ao longo do processamento dessas imagens pelo PhotoDoc o valor absoluto de erros de transcrição se aproximou do resultados obtidos nas imagens escaneadas a 300 *dpis*. Observando as Tabelas acima se pode constatar que:

- A etapa de remoção de bordas diminui o número de erros de inserção já que o ruído proveniente dela é eliminado.
- Por sua vez a etapa de correção de perspectiva traz um aumento do número de inserções uma vez que durante o processo de correção novos artefatos são gerados (interpolação), por outro lado possibilita o aumento da resolução dos caracteres mais afastados o que leva a uma diminuição no valor absoluto do número de erros de substituição. Pode-se observar através das Tabelas 9.5 e 9.8 o ganho na transcrição dessas imagens. Ainda é possível observar que a distorção de perspectiva é um fator que degrada o reconhecimento dos caracteres, sendo esse aumento de degradação diretamente proporcional ao ângulo formado pela câmera e a normal em relação ao plano que contem o documento, as imagens com ângulo de 0° e a mão livre

(inferiores a 5°) apresentaram melhores resultados de transcrição. Ainda é possível concluir que o aumento da resolução contribui para melhor distribuição dos erros, assim como a sua redução e que o processamento pelo PhotoDoc trouxe um melhor custo benefício em relação ao aumento de resolução da câmera digital (Tabela 9.8).

- Já a etapa de realce+binarização apresentou o melhor resultado absoluto em termos de erros de caracteres, porem em alguns casos houve o aumento do número de inserções geradas pelo processo de binarização.
- Por fim, tomando as Tabelas 9.4 e 9.8 é possível notar a viabilidade do uso do PhotoDoc para o processamento de imagens de documentos, uma vez que em termos absolutos o número de erros de caracteres das imagens fotografadas pelo PhotoDoc apresentam valores próximos (diferença máxima de 8%), o que representa um excelente resultado visto as dificuldades impostas pela digitalização de documentos por câmeras digitais portáteis.

Capítulo 10
Conclusões e trabalhos futuros

A partir da análise dos resultados apresentados no decorrer do capítulo 9 desta dissertação é possível constatar, apesar das adversidades encontradas durante a digitalização de documentos por câmeras fotográficas digitais, que é viável a aquisição e disseminação de informação por meio desses dispositivos. Os efeitos indesejados causados no processo de digitalização de documentos mostraram que sua influência aumenta o erro na transcrição por ferramentas de OCR, entretanto este aumento da quantidade de erros mostrou não inviabilizar essa transcrição, visto que em valores absolutos de erros de caracteres nos documentos fotografados após a filtragem automática pelo PhotoDoc, foi muito próxima aos documentos digitalizados por scanner a 200 e 300 *dpi*. Além da análise dos principais fatores influentes na qualidade de documentos digitalizados por câmeras fotográficas digitais, foram desenvolvidos três algoritmos para o processamento de documentos fotografados coloridos, sendo um para detecção de bordas, outro para correção de perspectiva e por fim um algoritmo de realce o que trouxe um grande avanço na pesquisa apresentados em [96].

O algoritmo de detecção de borda é uma evolução do apresentado em [11], onde existem algumas limitações que foram resolvidas por esta dissertação, tais com: documentos com baixa resolução, cor de fundo próxima a cor do papel e presença de imagens no documento. Todos esses são fatores que degradam de forma acentuada o desempenho do algoritmo proposto em [11]. O mesmo é aplicado ao algoritmo para busca dos vértices dos documentos apresentado em [49], onde estas mesmas limitações apresentadas em [46] existem. Já o algoritmo de realce proposto se mostrou eficiente na melhoria dos resultados da binarização, das imagens de documentos fotográficos, que por sua vez possibilitou um ganho expressivo na extração do OCR dessas imagens realizada por [95].

Ainda há muito a ser desenvolvido, principalmente em documentos com baixa resolução, para que se possa considerar a análise de documentos fotografados um problema resolvido. Os trabalhos futuros incluem o desenvolvimento de filtros de super-resolução [110][114] e reconstrução dos caracteres [76][111] [113], que possivelmente melhoraram a transcrição por ferramentas de OCR.

Referências

[1] N. F. Alves, Estratégias para melhoria do desempenho de ferramentas comerciais de reconhecimento óptico de caracteres, Dissertação de Mestrado em Engenharia Elétrica, Universidade Federal de Pernambuco, Recife, Brasil, 2003.

[2] T. Akiyama and N. Hagita. Automated entry system for printed documents. Pattern Recognition, vol. 23, pp 1141-1154, 1990.

[3] H. S. Baird, H. Bunke, and K. Y. (Eds.), Document image defect models, New York, pp. 546-556, New York: Springer Verlag, 1992.

[4] M. Cannon, J. Hochberg, and P. Kelly, Quality assessment and restoration of typewritten document images. IJDAR - International Journal on Document Analysis and Recognition, vol. 2, no. 2-3, pp. 80-89, 1999.

[5] P. Clark and M. Mirmehdi, Location and recovery of text on oriented surfaces,in SPIE conference on Document Recognition and Retrieval VII. The International Societyfor Optical Engineering, January 2000, pp. 267-277. Available: http://www.cs.bris.ac.uk/Publications/Papers/1000439.pdf.

[6] W. Philips, On the recovery of oriented documents from single images, in Proceedings of the 4th IEEE Advanced Concepts for Intelligent Vision Systems, Ed. Ghent University, September 2002, pp.190-197. Available: http://www.cs.bris.ac.uk/Publications/Papers/1000664.pdf.

[7] J. M. M. da Silva, R. D. Lins, and V. C. da Rocha, Binarizing and filtering historical documents with back-to-front interference, ACM SAC'06: Proceedings of the 2006 ACM Symposium on Applied Computing, pp. 853-858, ACM Press, 2006.

[8] J. M. M. da Silva, Um novo algoritmo baseado em entropia para filtragem de interferência frente-verso, Dissertação de Mestrado em Engenharia Elétrica, Universidade Federal de Pernambuco, Recife, Brasil, 2004.

[9] C. Dance and L. Fan, Color reconstruction in digital cameras: optimization for document images, IJDAR - International Journal on Document Analysis and Recognition, vol. 7, pp. Issue 23, Pages 138 146, Springer Verlag, July 2005.

[10] J. G. James. e L. Velho, Computação Gráfica: Imagem. Sociedade Brasileira de Matemática, 1994.

[11] A. R. G. e Silva and R. D. Lins, Background removal of document images acquired using portable digital cameras, in Proceedings of ICIAR 2005, Lecture Notes in Computer Science, vol. 3656, pp. 278-285, Springer Verlag, 2005.

[12] K.C. Fan, T.R. Lay, and Y.-K. Wang, Marginal noise removal of document images. 6th International Conference on Document Analysis and Recognition (ICDAR 2001), Seattle, WA, USA, September 2001, pp. 317-321, New York : IEEE Press.

[13] R. C. Gonzalez and R. E. Woods, Digital Image Processing, 3nd ed. Prentice Hall, 2008.

[14] R. Hartley and A. Zisserman, Multiple View Geometry in Computer Vision. Cambridge University Press, 2000.

[15] C. Hass, JPEG Quality Comparison, http://www.impulseadventure.com/photo/jpeg-decoder.html. Visitado em 11/08/2009.

[16] L. Jagannathan and C. V. Jawahar, Perspective correction methods for camera based document analysis, First International Workshop on Camera-based Document Analysis and Recognition, CBDAR, pp. 148-154, IAPR Press, 2005.

[17] B. Jähne, Digital Image Processing, 3rd ed. Springer, 1995.

[18] G. Johannsen and J. Bille, A threshold selection method using information measures, ICPR'82: Proceedings 6th International Conference Pattern Recognition, pp. 140-143, 1982.

[19] T. Kanungo, H. Baird, and R. Haralick, Validation and estimation of document degradation models, Proceedings Fourth Annual Symposium. Document Analysis and Information Retrieval, pp. 5-30, April 1995.

[20] T. Kanungo, R. Haralick, and I. Phillips, Global and local document degradation models, 1993. http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.51.359.

[21] J. N. Kapur, P. K. Sahoo, and A. K. C. Wong, A new method for gray-level picture thresholding using the entropy of the histogram, Computer Vision, Graphics and Image Processing, vol. 29(3), pp. 273-285, 1985.

[22] D. Kidner, M. Dorey, and D. Smith, What's the point? interpolation and extrapolation with a regular griddem, Geocomputation 99: Proceedings of the 4th International Conference on GeoComputation, 1999.

[23] J. Kim, Y. Byun, and J. Choi, Background removal of document images acquired using portable digital cameras, in Advances in Multimedia Information Processing - PCM 2004, Lecture Notes in Computer Science, vol. 3333. Springer, 2004, pp. 331-339.

[24] K. Larson, The science of word recognition or how I learned to stop worrying and love the bouma, Microsoft Corporation, 2004, http://www.microsoft.com/typography/ctfonts/WordRecognition.aspx. Visitado em 07/10/2008.

[25] A. Leykin and F. Cutzu, Differences of edge properties in photographs and paintings.in International Conference on Image Processing, ICIP, pp. 541-544, 2003.

[26] J. Liang, D. Doermann, and H. Li, Camera-based analysis of text and documents: A survey, International Journal on Document Analysis and Recognition, vol. 7, no. 2-3, pp. 83-104, Springer Verlag, July 2005.

[27] R. D. Lins and N. F. Alves, A new technique for acessing the performance of OCR, IADIS - International Conference on Computer Applications, vol. 1, pp. 51-56, Algarve, 2005.

[28] S. Lu, B. M. Chen, and C. C. Ko, Perspective rectification of document images using fuzzy set and morphological operations. Image Vision Comput., vol. 23, no. 5, pp. 541-553, 2005.

[29] S. Lu and C. L. Tan, The restoration of camera documents through image segmentation, in Document Analysis Systems, 2006, pp. 484 495.

[30] C. A. B. Mello and R. D. Lins, Image segmentation of historical documents, Proceedings of Visual 2000, pp. 209-216, Mexico City, Mexico, 2000.

[31] C. A. B. Mello and R. D. Lins, Generation of images of historical documents by composition, in DocEng '02: Proceedings of the 2002 ACM symposium on Document engineering, pp. 127-133, ACM Press, 2002.

[32] G. K. Myers, R. C. Bolles, Q.-T. Luong, J. A. Herson, and H. Aradhye, Rectification and recognition of text in 3-d scenes, IJDAR, vol. 7, no. 2-3, pp. 147 158, 2005.

[33] A. C. Naiman, The use of grayscale for improved character presentation, Ph.D. disser- tation, University of Toronto, Toronto, Canada, 1991.

[34] N. Otsu, A threshold selection method from gray level histograms, IEEE Transactions on Systems, Man and Cybernetics. - SMC, vol. 9(1), pp. 62-66, 1979.

[35] K. Popat, Decoding of text lines in grayscale document images, Proceedings of the 2001 International Conference on Acoustics, Speech, and Signal Processing (ICASSP 2001), pp. 1513-1516, Salt Lake City, Utah: IEEE, May 2001.

[36] T. Pun, Entropic thresholding, a new approach, C. Graphics and Image Processing, vol. 16(3), 1981.

[37] J. Reilly and F. Frey, Recommendations for the evaluation of digital images produced from photographic, microphotographic, and various paper formats, Library of Congress National Digital Library Project, Washington, DC, 1996, url: http://www.microsoft.com/typography/ctfonts/WordRecognition.aspx. Visitado em 11/08/2009.

[38] P. L. Rosin and A. D. Marshall, A light-weight text image processing method for handheld embedded cameras, in Proceedings of the British Machine Vision Conference 2002, BMVC 2002. British Machine Vision Association, 2002.

[39] L. S. and C. L. Tan, Camera document restoration for OCR, First International Workshop on Camera-based Document Analysis and Recognition, CBDAR, pp. 17-24, 29 August 2005, Seoul, Korea.

[40] M. Seeger and C. Dance, Binarising camera images for ocr, ICDAR '01: Proceedings of the Sixth International Conference on Document Analysis and Recognition, pp. 54-59, IEEE Computer Society, 2001.

[41] C. Shannon, A mathematical theory of communication, in Bell System Technology Journal, vol. 27, pp. 370-423 and 623-656, 1948.

[42] C. Thillou and B. Gosselin, Color binarization for complex camera-based images, in Videometrics VIII. Proceedings of the SPIE, vol. 5667, pp. 301-308, 2004.

[43] O. D. Trier and T. Taxt, Evaluation of binarization methods for document images, IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 17, no. 3, pp. 312-315, 1995.

[44] B. T. Ávila and R. D. Lins, A new algorithm for removing noisy borders from monochromatic documents, in SAC '04: Proceedings of the 2004 ACM Symposium on Applied Computing, pp. 1219-1225, ACM Press, 2004.

[45] L. U. Wu, M. A. Songde and L. U. Hanqing, An efective entropic thresholding for ultrasonic imaging, ICPR'98: International Conference Patterns Recognition., pp. 1522-1524, 1998.

[46] H. S. Yam and E. H. B. Smith, Estimating degradation model parameters from character images, Proceedings of ICDAR 2003, vol. 02, p. 710-715, 2003.

[47] J. C. Yen, F.-J. Chang, and S. Chang, A new criterion for automatic multilevel thresholding. IEEE Transactions on Image Processing, vol. 4, no. 3, pp. 370-378, 1995.

[48] ACD Systems, http://www.acdsee.com/support. Visitado em 10/09/2009.

[49] R. D. Lins, G.Pereira e Silva and A.R.Gomes e Silva, Assessing and Improving the Quality of Document Images Acquired with Portable Digital Cameras, ICDAR´07, vol 2, pp. 569-573, Curitiba, Brasil, 2007.

[50] R. D. Lins, A. R. G. Silva, and G. F. P. Silva, Enhancing Document Images Acquired Using Portable Digital Cameras. In: International Conference on Image Analysis and Recognition, 2007, Montreal (Canadá). Proceedings of ICIAR 2007. Springer Verlag, 2007. v.LNCS. p. 1229-1241.

[51] G. F. P. Silva, and R. D. Lins, PhotoDoc: A Toolbox for Processing Document Images Acquired Using Portable Digital Cameras, CBDAR´07, pp. 107-114, Curitiba, Brasil, 2007.

[52] ImageJ url: http://rsb.info.nih.gov/ij/ , Visitado em 10/08/2009.

[53] Tesseract url: http://code.google.com/p/tesseract-ocr/ Visitado em 10/08/2009.

[54] K.C.Fan, Y.K.Wang, T.R.Lay, Marginal noise removal of document images, Pattern Recognition. 35, 2593-2611, 2002.

[55] D. M. Oliveira and R. D. Lins, Processing Teaching-board Images Acquired with Portable Digital Cameras, 2nd InternationalWorkshop on Camara Based Document Analysis and Recognition, CBDAR 2007, pp. 79-86, Curitiba, Brazil.

[56] M. Sezgin and B. Sankur. Survey over Image Thresholding Techniques and Quantitative Performance Evaluation. Journal of Eletronic Imaging, 13(1), pp 145-165, January, 2004;

[57] Z. Liu; Z. Zhang; H. Li-Wei; 2007. Whiteboard scaning and image enhancement. Digital Signal Processing, Volume 17, Issue 2, p. 414-432.

[58] Nokia, 2008. Device Details -- Nokia 6120 classic. Disponível em: http://www.forum.nokia.com/devices/6120_classic. Visitado em 04/04/2009.

[59] A. R. G. Silva, 2006. Análise e Melhoria da Qualidade de Documentos Fotografados. Dissertação de Mestrado. Universidade Federal de Pernambuco, PPGEE, Pernambuco,Recife, 2006.

[60] Söbel, I.E.; 1970. Camera Models and Machine Perception. Ph.D. dissertation. Stanford University University. Palo Alto. Calif.

[61] Theuwissen, Albert J.P.; 2007. CMOS image sensors: Stateof-the-art and future perspectives. Em: 37th European Solid State Device Research.

[62] X. C. Yin; J. Sun;Y. Fujii; K. Fujimoto; S. Naoi; 2007. Perspective Rectification for Mobile Phone Camera-Based Documents Using a Hybrid Approach to Vanishing Point Detection. Proceedings of Second International IAPR Workshop on Camera-Based Document Analysis and Recognition, p.37-44, IAPR.

[63] A. Masalovitch., L. Mestetskiy, 2007. Usage of Continuous Skeletal Image Representation for Document Images De-warping. Proceedings of Second International IAPR Workshop on Camera-Based Document Analysis and Recognition, p.45-52, IAPR.

[64] S. Uchida; M. Sakai; M. Iwamura; S. Omachi; K. Kise; 2007. Instance-Based Skew Estimation of Document Images by a Combination of Variant and Invariant. Proceedings of Second International IAPR Workshop on Camera-Based Document Analysis and Recognition, p.79-86, IAPR.

[65] N. Abramson, "Information Theory and Coding", McGraw-Hill Book Co, 1963.

[66] T.Morris, “Computer Vision And Image Processing”, Palgrave Macmillan, 2003.

[67] A.P.A. Castro. Detecção de Bordas e Navegação Autônoma Utilizando Redes Neurais Artificiais. São José dos Campos - SP, 2003. 154p. Dissertação de Mestrado - Instituto Nacional de Pesquisas Espaciais, INPE.

[68] S. Haykin, Redes Neurais: Princípios e Prática. 3nd ed, Porto Alegre: Bookman, 2006. 900 p.

[69] L. Fausett, Fundamentals of Neural Networks: Architectures, Algorithms, and Applications. New Jersey: Prentice Hall, 1994. pp.461- 471.

[70] J. F. Canny, A Computational Approach to Edge Detection. IEEE Trans. Pattern Analysis and Machine Intelligence, v. PAMI, n. 8, p. 679-698, 1986.

[71] B. Jähne, H. HauBecker, and P. GeiBler. Handbook on Computer Vision and Apllications, volume 2. Academic Press, 1999.

[72] M. Hanmandlu, J. See, and S. Vasikarla. Fuzzy edge detector using entropy optmization. In IEEE Computer Society, editor, Proceedings of the International Conference on Technology: Coding and Computing, pages 665–670, 2004.

[73] P.V.C. Hough; Methods and means for recognizing complex patterns. U.S. Patent 3.069.654, 1962.

[74] D. M. Oliveira, 2007. Tableau um Ambiente para Processamento de Imagens de Quadros Didáticos. Dissertação de Mestrado. Universidade Federal de Pernambuco, PPGEE, Pernambuco,Recife, 2008.

[75] G. Brown, 2008. How autofocus cameras works. Disponível em: http://electronics.howstuffworks.com/autofocus.htm/printable. Visitado em 10/08/2009.

[76] S. Borman and R. Stevenson. Spatial resolution enhancement of low-resolution image sequences - a comprehensive review with directions for future research. Technical Report, University of Notre Dame, 1998.

[77] H. Baird, Document image defect models and their uses, in International Conference on Document Analysis and Recognition, ICDAR93, Tsukuba, Japan, pp. 62 67, 2003, New York : IEEE Press.

[78] R. D. Lins. A Taxonomy for Noise Detection in Images of Paper Documents - The Physical Noises. International Conference on Image Analysis and Recognition, 2009, Halifax (Canada). Proceedings of ICIAR 2009. Heidelberg : Springer Verlag, 2009. v. 5627. pp. 844-854.

[79] G. F. P. Silva; R. D. Lins; B. Miro; S.J. Simske; M Thielo. Automatically Deciding if a Document was Scanned or Photographed. Journal of Universal Computer Science, vol. 15, no. 18 (2009), pp.3364-3375.

[80] Lins, R. D ; Silva, G. F. P ; Simske, S.J. ; Fan, Jian ; Shaw, Mark ; Sá, P ; Thielo, M . Image Classification to Improve Printing Quality of Mixed-Type Documents. In: International Conference on Document Analysis and Recognition, 2009, Barcelona. Proceedings of ICDAR 2009. New York : IEEE Press, 2009. p. 1106-1110.

[81] Adobe Acrobat 8.0 Pro url: http://www.adobe.com, Visitado em 10/08/2009.

[82] H.Frigui and R.Krishnapuram. Clustering by competitive agglomeration. Pattern Recognition, v.30(7), pp. 1109-1119, 2001.

[83] M.A.Hearst and J.O.Pedersen. Reexamining the Cluster Hypotesis: Scatter Gathet on Retrieval Results, SIGIR, 1996.

[84] S.Krishnamachari and M.Abdel-Mottaleb. Image Browsing using Hierarchical Clustering, IEEE Symposium on Computers and Communications, ISCC'99, July 99.

[85] P.Scheunders. Comparison of Clustering Algorithms Applied to Color Image Quantization, Pattern Recognition Letters, v18(11-13):1379-1384, 1997.

[86] G.Park, Y.Baek and L.Heung-Kyu. A Ranking Algorithm Using Dynamic Clustering for Content-Based Image Retrieval. CIVR'2002, pp.328-337, LNCS 2383, Springer Verlag, 2002.

[87] L. Breiman, "Random Forests", Machine Learning, 45(1), pp.5-32, 2001.

[88] R.D.Lins and D.S.A.Machado, A Comparative Study of File Formats for Image Storage and Trans., v13(1):175-183, Journal of Electronic Imaging, 2004.

[89] S.J. Simske, "Low-resolution photo/drawing classification: metrics, method and archiving optimization," Proceedings IEEE ICIP, IEEE, Genoa, Italy, pp. 534-537, 2005.

[90] Weka 3: Data Mining Software in Java, url: http://www.cs.waikato.ac.nz/ml/weka/, visitado em 10/08/2009.

[91] R. G. Maya. N. P. Jacobson, E. K. Garcia, OCR binarization and image pre-processing for searching historical documents, Pattern Recognition 40 (2007) 389 – 397, Elsevier.

[92] B. Gatos, K. Ntirogiannis and I. Pratikakis, Document Image Binarization Contest (DIBCO 2009), ICDAR 2009, pp. 1375-1382, Barcelona, Spain.

[93] D.M Oliveira and R.D. Lins. A New Method for Shading Removal and Binarization of Documents Acquired with Portable Digital Cameras. In: International Workshop on Camera-Based Document Analysis and Recognition, 2009, Barcelona. Proceedings of CBDAR 2009. New York : IAPR Press, 2009. pp. 98-105.

[94] Java API PDFBOX, url: http://incubator.apache.org/pdfbox/, Visitado em 10/08/2009.

[95] ABBYY FineReader 9.0 Pro, url: http://www.abbyy.com/, Visitado em 10/08/2009.

[96] G.F.P. Silva, Análise e melhoria da qualidade de documentos fotografados, Tabalho de Graduação, Centro de Informatica Universidade Federal de Pernambuco, 2007 , Setembro, Recife.

[97] K. S. Guimarães; T. M. Przytycka. Domain-domain and Protein-protein Interaction. Protein-protein interactions and networks: Identification, Analysis and Prediction, Springer Verlag, 2008, v.21 , p. 83-99.

[98] M.Cheriet and R.F.Moghaddam. DIAR: Advances in Degradation Modeling and Processing, ICIAR 2008, LNCS(5112):1-10, Springer Verlag, 2008.

[99] J.R.C. Pinto et al. Underline Removal on Old Documents. ICIAR 2004, LNCS(3212), v(2):226-233, 2004.

[100] E. Kavallieratou and H. Antonopoulou, Cleaning and Enhancing Historical Document Images, Intelligent Vision Systems, LNCS 3708:pp. 681-688, Springer-Verlag, 2005.

[101] J. Bernsen. "Dynamic thresholding of gray level images". ICPR'86: Proc. Intl. Conf. Patt. Recog., pp. 1251-1255, 1986.

[102] A. Khashman and B. Sekeroglu. "A Novel Thresholding Method for Text Separation and Document Enhancement", Proceedings of the 11th Panhellenic Conference on Informatics (PCI 2007), Patras, Greece, 18-20 May 2007.

[103] P. W. Palumbo, P. Swaminathan, and S. N. Srihari. "Document image binarization: Evaluation of algorithms". Proc. SPIE 697, 278-286, 1986.

[104] T. W. Ridler and S. Calvard, "Picture thresholding using an iterative selection method," IEEE Trans. Syst. Man Cybern. SMC-8, 630-632, 1978.

[105] J. Kittler and J. Illingworth, "Minimum error thresholding," Pattern Recogn. 19, 41-47, 1986.

[106] W. Niblack. "An Introduction to Image Processing". pp. 115-116, Prentice-Hall,, 1986.

[107] J. Sauvola and M. Pietaksinen. "Adaptive document image binarization". Pattern Recognition. 33, 225-236, 2000.

[108] J. M. White and G. D. Rohrer. "Image thresholding for optical character recognition and other applications requiring character image extraction". IBM J. Res. Dev. 27(4), 400-411, 1983.

[109] L. O'Gorman, Experimental comparisons of binarization and multithresholding Methods on document images, in: Proceedings of the IAPR International Conference on Pattern Recognition, vol. 2, IEEE, 1994, pp. 395-398.

[110] S. Baker and T. Kanade. Limits on super-resolution and how to break them. IEEE Trans. Pattern Anal. Mach. Intell., 24(9), pp. 1167-1183, 2002.

[111] E. Borenstein and S. Ullman. Combined top-down/bottom-up segmentation. IEEE Trans. Pattern Anal. Mach. Intell., 30(12), pp. 2109-2125, 2008.

[112] F. Drira. Towards restoring historic documents degraded over time. In DIAL '06, pages 350.357. IEEE Computer Society, 2006.

[113] C. Liu, R. Szeliski, S. B. Kang, C. L. Zitnick, and W. T. Freeman. Automatic estimation and removal of noise from a single image. IEEE Trans. Pattern Anal. Mach. Intell., 30(2), pp.299-314, 2008.

[114] H. Q. Luong and W. Philips. Robust reconstruction of low-resolution document images by exploiting repetitive character behaviour. International Journal of Document Analysis and Recognition, 11(1), pp.39.51, 2008.

Apêndice A
PhotoDoc Manual

1 Introdução

PhotoDoc foi desenvolvido na linguagem Java o que o torna uma ferramenta versátil que pode ser utilizada pelos principais Sistemas Operacionais (Windows, Linux e MacOS). Sempre que o usuário descarrega suas fotos ele poderá utilizar as funcionalidades deste ambiente antes de armazená-las, imprimir ou enviar através das redes as imagens dos documentos. Os algoritmos e as funcionalidades básicas do PhotoDoc podem ser incorporados em um dispositivo portátil tal como um PDA ou mesmo uma câmera digital, tal possibilidade será desenvolvida futuramente. Devido à simplicidade e à portabilidade a versão atual de PhotoDoc foi desenvolvida como um *plug-in* do ImageJ. ImageJ é um ambiente *open source* para o processamento de imagens desenvolvido por Wayne Rasband na linguagem Java, no National Institute of Mental Health, Bethesda, Maryland, EUA. A Figura A1.1 mostra apresenta a tela do PhotoDoc que está sendo ativado no ImageJ.

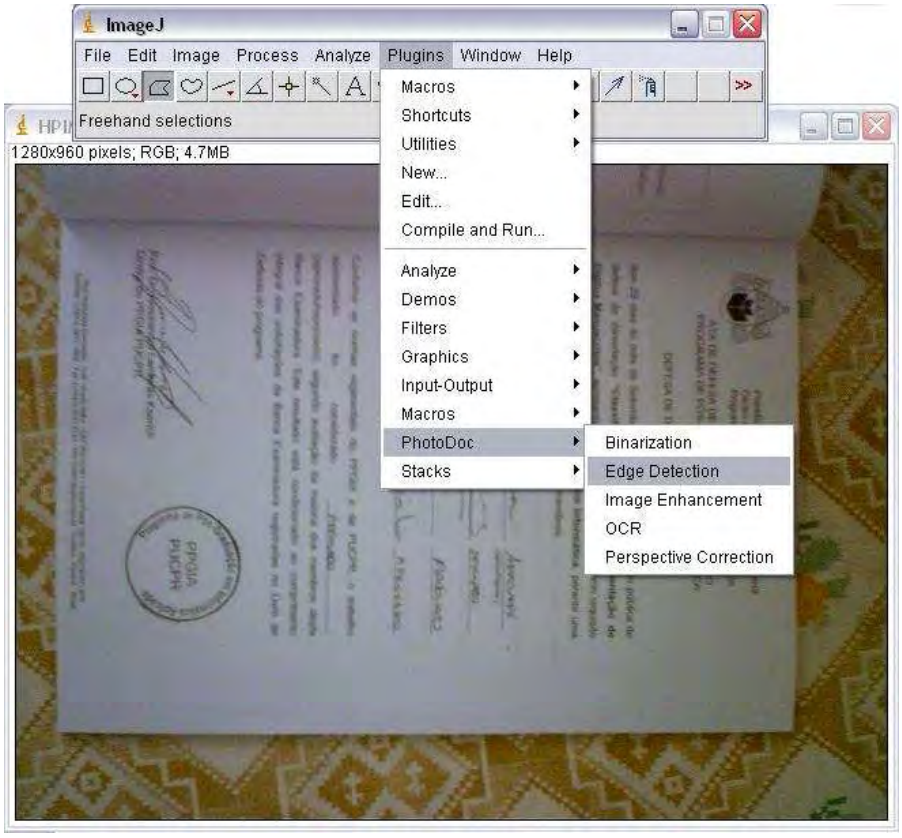


Figura A1.1 - Imagem da ferramenta PhotoDoc ativada no ImageJ.

Pode-se observar na Figura 1.1, a versão atual do plug-in do PhotoDoc oferece cinco funcionalidades distintas, que aparecem na seguinte ordem:

1. *Binarization*;
2. *Edge Detection*;
3. *Image Enhancement*;
4. OCR;
5. *Perspective Correction and Crop*.

O fato do PhotoDoc está presente no ImageJ como um plug-in também permite que o usuário experimente as diferentes funcionalidades e filtros dos atuais plug-ins já incorporados pelo ImageJ. É importante comentar que a biblioteca do ImageJ é *open source* o que permite o uso em separado do PhotoDoc. As funcionalidades do PhotoDoc serão descritas nas a seguir.

2 Detecção de Bordas

Uma das primeiras etapas a serem abordas em processamento de imagem de documentos no PhotoDoc é detectar os limites físicos reais do documento original. O algoritmo para detecção de bordas acoplado ao PhotoDoc foi projetado mais especificamente para imagens adquiridas por câmeras digitais, porém se mostrou útil para o processamento de documentos escaneados. O resultado da ativação da tecla da “Edge Detection” em um documento no PhotoDoc fornece uma imagem tal como essa apresentada na figura A2.1.

Figura A2.1 - Imagem de documento com as bordas (amarela) detectadas automaticamente pelo PhotoDoc.

A funcionalidade de detecção da borda no PhotoDoc permita que o usuário ajuste as bordas arrastando ao longo dos quatro cantos da imagem do documento. Isto pode ser de grande ajuda sempre que a foto do documento não é cercada completamente por uma borda ou quando as bordas não forem detectadas automaticamente.

3 Correção de Perspectiva e Recorte

PhotoDoc incorpora uma funcionalidade para corrigir a distorção e o enviesamento da imagem introduzida durante sua aquisição. O resultado da ativação da tecla "*Perspective Correction and Crop*" sobre a imagem A2.1 no PhotoDoc resulta na imagem apresentada pela Figura A3.1.


Pontifícia Universidade Católica do Paraná
Centro de Ciências Exatas e de Tecnologia
Programa de Pós-Graduação em Informática

ATA DE DEFESA DE DISSERTAÇÃO DE MESTRADO
PROGRAMA DE PÓS-GRADUAÇÃO EM INFORMÁTICA

DEFESA DE DISSERTAÇÃO Nº 09/2006

Aos 29 dias do mês de Setembro de 2006 realizou-se a sessão pública de defesa da dissertação "**Classificação de Pontos de Segmentação de Dígitos Manuscritos**", apresentada por **Eduardo Vellasques** como requisito parcial para a obtenção do título de **Mestre em Informática**, perante uma Banca Examinadora composta pelos seguintes membros:

Prof. Dr. Luiz Eduardo Soares de Oliveira — PUCPR (Orientador) — APROVADO
Prof. Dr. Alceu de Souza Britto Junior — PUCPR — APROVADO
Prof. Dr. Flávio Bortolozzi — PUCPR — APROVADO
Prof. Dr. Rafael Dueire Lins — UFPE — APROVADO

Conforme as normas regimentais do PPGIA e da PUCPR, o trabalho apresentado foi considerado APROVADO (aprovado/reprovado), segundo avaliação da maioria dos membros desta Banca Examinadora. Este resultado está condicionado ao cumprimento integral das solicitações da Banca Examinadora registradas no Livro de Defesas do programa.

Prof. Dr. Alessandro Lameiras Koerich
Diretor do PPGIA/PUCPR

PPGIA PUCPR

Figura A3.1 - Imagem de documento após a correção de perspectiva pelo PhotoDoc.

4 Realce

PhotoDoc incorpora uma funcionalidade para corrigir a distorção causadas pela iluminação irregular provenientes da captura de imagens de documentos por câmeras fotográficas digitais. O resultado da ativação da tecla "*Image Enhancement*" sobre a imagem 3.1 resulta na imagem apresentada pela Figura A4.1.

Figura A4.1 - Imagem de documento depois de realçada pelo PhotoDoc.

5 Binarização

O PhotoDoc incorpora alguns dos algoritmos de binarização. Os algoritmos de binarização incorporados ao PhotoDoc nesta versão são todos globais:

1. Algoritmo 1 – SLR_Improved_Threshold;
2. Algoritmo 2 – MelloLins_Threshold ;
3. Algoritmo 3 – Pun_Threshold;
4. Algoritmo 4 – KapurSahooWong_Threshold ;
5. Algoritmo 5 – Wu_Songde_Hanqing_Threshold ;
6. Algoritmo 6 – Otsu_Threshold ;
7. Algoritmo 7 – Yen_ChangChang_Threshold ;
8. Algoritmo 8 – Johannsen-Bille [18].

6 OCR

Outra funcionalidade presente no PhotoDoc é interação com a ferramenta de reconhecimento de caracteres ópticos (OCR) Tesseract. O Tesseract foi desenvolvido nos laboratórios da HP entre 1985 e 1995. Era uma das três melhores ferramentas de OCR presentes no UNLV de 1995. Desde então, poucos avanços foram realizados, mas é provavelmente uma das melhores ferramentas de OCR *open source* disponível hoje. A sua limitação é a respeito da imagem de entrada (aceita apenas imagens binárias sem compressão) e não tratar o *layout* do documento.

Ao ativar a opção “OCR” do PhotoDoc ele ativará o Tesseract. Os testes preliminares tais como o apresentado na figura A6.1 mostra o bom resultado da transcrição das imagens pelo Tesseract.

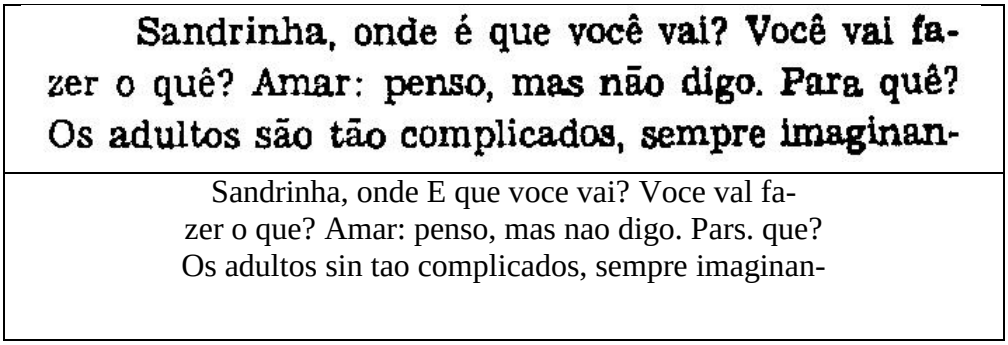


Figura A6.1 - Imagem de um segmento de texto e a respectiva transcrição feita pelo Tesseract.

Apêndice B
Publicações

A1. R. D. Lins, G. F. P. Silva and A. R. G. Silva, Assessing and Improving the Quality of Document Images Acquired with Portable Digital Cameras, ICDAR´07, Curitiba, Brasil, 2007.

A2. R. D. Lins, A. R. G. Silva, and G. F. P. Silva, Enhancing Document Images Acquired Using Portable Digital Cameras. In: International Conference on Image Analysis and Recognition, 2007, Montreal (Canadá). Proceedings of ICIAR 2007. Springer Verlag, 2007. v.LNCS. p. 1229-1241.

A3. G. F. P. Silva, and R. D. Lins, PhotoDoc: A Toolbox for Processing Document Images Acquired Using Portable Digital Cameras, CBDAR´07, Curitiba, Brasil, 2007.

A4. G. F. P. Silva; R. D. Lins; B. Miro; S.J. Simske; M Thielo. Scaned or Photographed?Automatically Deciding How a Document was Digitized. In: International Workshop on Camera-Based Document Analysis and Recognition, 2009, Barcelona. Proceedings of CBDAR 2009. New York: IAPR Press, 2009. p. 1-10.

A5. R. D. Lins, G. F. P Silva, S. J. Simske, J. Fan, M. Shaw, P. Sá, M. Thielo. Image Classification to Improve Printing Quality of Mixed-Type Documents. In: International Conference on Document Analysis and Recognition, 2009, Barcelona. Proceedings of ICDAR 2009. New York: IEEE Press, 2009. p. 1106-1110.

A6. R. D. Lins; B. Miro; G. F. P. Silva. An OCR Assessment of the Quality of Document Images Acquired with Portable Digital Cameras. In: (ICDAR 2009/CBDAR)Third International Workshop on Camera-Based Document Analysis and Recognition, 2009, Barcelona. Proceedings of the Third International Workshop on Camera-Based Document Analysis and Recognition. New York : IEEE Press, 2009. v. 1. p. 106-111.

A7. J. M. M. da Silva; R. D. Lins; G. F. P. Silva. Melhorando A Qualidade de Documentos Coloridos com Interferência Frente-Verso. In: XXVI Simpósio Brasileiro de Telecomunicações (SBrT 2009), 2008, Rio de Janeiro. Anais do XXVI Simpósio Brasileiro de Telecomunicações. Rio de Janeiro : SBRT, 2008.

A8. R. D. Lins, G. T. Silva; G. F. P. Silva. Content Recognition and Indexing in the Livememory Platform. In: Eighth IAPR International Workshop on Graphics RECognition - GREC 2009, 2009, La Rochelle. Proceedings of GREC 2009. Heidelberg : Springer Verlag, 2009. v. 1. p. 224-230.

A9. J. M. M. da Silva; R. D. Lins; G. F. P. Silva. Enhancing the Quality of Color Documents with Back-to-Front Interference. In: International Conference on Image Analysis and Recognition, 2009, Halifax. Proceedings of ICIAR 2009 - Lecture Notes in Computer Science. Heidelberg : Springer Verlag, 2009. v. 5627. p. 875-885.

Assessing and Improving the Quality of Document Images Acquired with Portable Digital Cameras

Rafael Dueire Lins, Gabriel Pereira e Silva, André Ricardson Gomes e Silva
Departamento de Eletrônica e Sistemas – Universidade Federal de Pernambuco - Brazil
rdl@ufpe.br, gfps@cin.ufpe.br, andrericardson@yahoo.com.br

Abstract

Professionals and students of many different areas start to use portable digital cameras to take photos of documents, instead of photocopying them. This article analyses the quality of such documents for Optical Character Recognition and proposes ways of improving their transcription and readability.

1. Motivation

The fast growth on image quality of portable digital cameras together with a drastic price reduction was widened its applicability into unforeseen domains. One of them is using portable digital cameras for digitalizing documents. Students and professionals of many different areas now use those devices as a fast way to acquire document images, taking advantage of their low weight, portability, low cost, small dimensions, etc. This new research area [1][2] is evolving fast in many different directions and claims for new algorithms, tools and processing environments that are able to provide users in general with simple ways of visualizing, printing, transcribing, compressing, storing and transmitting through networks such images. Reference [3] points out some particular problems that arise in this document digitalization process: the first of all is background removal. Very often the document photograph goes beyond the document size and incorporates parts of the area that served as mechanical support for taking the photo of the document. The second problem is due to the skew often found in the image in relation to the photograph axes, as documents have no fixed mechanical support very often there is some degree of inclination in the document image. The third problem is non-frontal perspective, due to the same reasons that give rise to skew. A fourth problem is caused by the distortion of the lens of the camera. This means that the perspective distortion is not a straight line but a convex line, depending on the quality of the lens and the relative position of the camera and the document. The fifth difficulty in processing document images acquired with portable cameras is due to non-uniform illumination. This paper focuses on assessing the output of a commercially OCR (Optical Character Recognition) software for such documents and follows the steps pointed out in [3] to improve their transcription and readability.

2. Assessment methodology

Assessing image quality is a task of uttermost complexity. Although the human visual-neural system is an extremely sophisticated, subjectivity plays an import role in image recognition and choice of quality, thus it should be avoided by every means. In this paper the assessment methodology was limited to analyze the performance of commercial OCR tools. Omnipage Professional 15.0 from Nuance [4] was used, because it is possibly the best general purpose available today. This study compares the results obtained by transcribing a batch of 50 documents which were scanned with a HP scanner (model 3200c) with 100, 150, and 300 dpi resolution in true color with the results obtained by transcribing the same documents with the cameras of 3.2 and 4.1 Mega pixels.
These results are later used to assess the gains obtained with the documents after each processing step. On its turn, analyzing the results of OCRs is far from being a trivial task. The methodology presented in reference [5] which takes into account the nature of the errors in transcription was adopted here.

The errors were classified according to:

1. Character errors (character replacement)
2. Missing characters.
3. Character insertion.
4. Graphical accents errors.
5. Word errors (number of incorrect words)
6. Word missing (complete words not transcribed)
7. Punctuation errors.

W*ords* are lexemes with at least three characters, avoiding stopwords and isolated characters.

3. Test images features

The test document images used here were pages extracted from books and other documents, printed on translucent paper in such a way that back-to-front interference was not observed [6]. No glossy paper was tested. Documents range in size from A5 to Legal, with predominance of size around A4. Some of them may include black-and-white or colour photograph. The only restriction imposed to documents is that there is at least a 2-pixel separation frame between the document background (paper) and document information. Figures 01 and 02 exemplify some of the document images tested. Documents were obtained in true-color, under different illumination conditions, with and without the inbuilt camera flash, using two different models of portable cameras manufactured by Sony Corp. (models DSC-P52 and DSC-P40) of 3.2 and 4.1 Mega pixels, respectively.

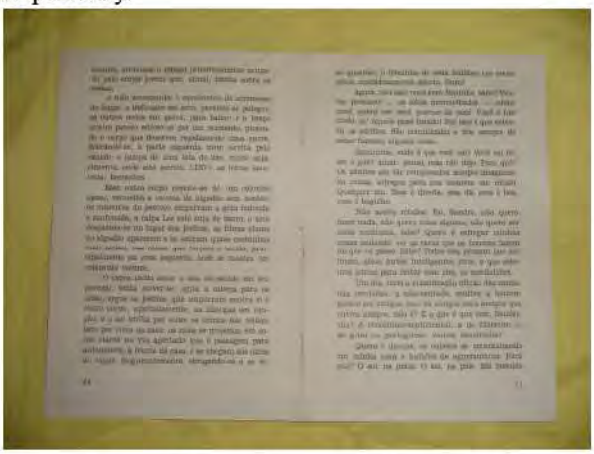


Figure 01 Image photographed in 3 Mpixels

The cameras were set into "auto-focus" mode, i.e. the user leaves to the device the automatic setting of the focus. This is consistent with the expected knowledge of the end user. Whenever the embedded strobe flash of the camera was not used the resulting document photograph "looked" slightly blurred (out-of-focus). Most possibly, this is related with the fact that the diaphragm of the objective stayed open for much longer to compensate the loss in illumination. As no mechanical support was used to stabilize the camera, chances are that the photographer moved briefly during the shot. Some other times, a slight inclination of the camera in a clear environment may be enough to the luminosity sensor to assume that there is no need for the camera to provide extra illumination, canceling the flash activation. Thus, in order to minimize these factors and also to meet the real use of portable digital cameras for acquiring document images all the photo documents analyzed in this research were:

1. Taken with the embedded flash of the camera set as "on", forcing its activation regardless of the luminosity of the environment.
2. Obtained indoors.

These restrictions increased the quality of the images and yielded a better recognition rate by the OCR.

Table I presents the results of the average of errors found in images scanned with 100, 150 and 300 dpi as well as the result of the transcription of the images of the same documents acquired using 3.2 and 4.1 Mega pixels portable digital cameras. A batch of 50 documents was studied totaling 13,532 words and 2,095,596 characters. These are reference values to assess the gains obtained by the processing steps below.

TABLE I
ORIGINAL CHARACTER AND WORD ERRORS FOUND IN IMAGES

		100	150	300	3Mp	4Mp
Character	Replacement	74	38	24	101	85
	Punctuation	123	3	1	79	18
	G. Accents	22	4	10	107	38
	Missing	38	27	5	15	10
	Insertion	43	92	19	39	31
Word	Errors	53	19	20	150	68
	Exclusions	-	-	-	2	1

4. Skew Analysis

In the case of scanned documents, very often documents are not always correctly placed on the flat-bet scanner either manually by operators or by the automatic feeding device. Skewed images are unpleasant for human visualization, introduce extra difficulty in text reading, claim extra space for storage, degrade OCR performance, etc. In the case of documents acquired with portable digital cameras with no mechanical support this problem arises in almost every document. The measure of the skew angle of the bottom line of the 50 documents analyzed was less than 2°. One should observe that the horizontal axis of the document, or bottom line, is taken as the reference. The background support frame as well as perspective distortion were ignored. Most commercial OCR automatically compensates skew angles up to 5 degrees, a range often found in scanned documents [5]. The addition of perspective and lens distortion yielded small variation in the skew angle of text lines throughout the document. This skew variation in the

batch of 50 documents studied was of less than 0.2° and did not degrade OCR response whenever compared the global document with line-by-line transcription.

5 Border Removal

An automatic background border removal of images of documents obtained with portable digital cameras should impose as few restrictions as possible, because users tend to acquire those document images in non-

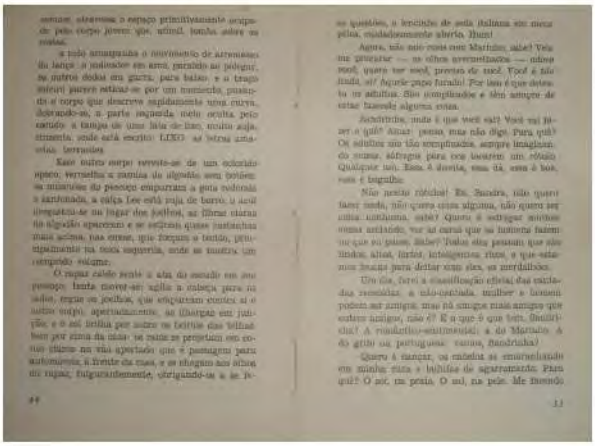


Figure 02 Document image from Figure 01 with border removed and skew corrected with vestigial border painted with the most frequent color in the document

ideal conditions of colour, texture, illumination of the surface the document is placed on for digitalisation, perspective camera-document, etc. As may be observed in Figure 01, the document image is surrounded by a yellow background area of no value in terms of information. This area not only drops the quality of the resulting image for CRT screen visualization, but also consumes space for storage and large amounts of toner for printing, alters the segmentation algorithm of the OCR and thus affects the response obtained in the number of characters and words correctly transcribed. Several papers in the literature address this problem in different applications [7, 8, 9, 10]. Removing such frame manually is not practical due to the need of a specialized user and time consumed in the operation. The algorithm presented in reference [9] was used here to automatically remove such border as an OCR pre-processing stage. It assumes that the background may be of any kind of colour or texture, provided that there is a colour difference of at least 32 levels between the image background and at least one of the RGB components of the most frequent colour of the document background (paper).

Due to the distortion caused by perspective, the border removal method leaves some vestigial border which was painted with the most frequent color in the document. Other possibilities for painting the vestigial border such as painting with the colors of contiguous areas were not analyzed in this study. One may conjecture that they may make the resulting document look more "natural" to the human reader but yields no effect in OCR response.

6 Analysis of lens distortion

The spherical shape of the lenses of cameras introduces a distortion of lines in photographed documents that is not seen in scanned ones. References [11] and [12] propose methods for the compensation of such distortion. Although in this experiment the spherical lens distortion was not compensated, the batch of documents digitized had its lens distortion evaluated. For this purpose, a straight line was drawn taking as extremities the vertices of each document. The maximum distance in pixels between this line and the document edge corresponds to the maximum lens distortion. The average chordal distance for the 3.2 Mpixel camera Sony DSC-P52 was 10 pixels, while for the 4.1 Mpixels DSC-S40 an average of 25 pixels was obtained. The lenses are Carl Zeiss S=5.0mm, 1:2.8, Multipoint A. The OCR tool showed little sensitivity to lens distortion. No especial degradation in performance was felt in the outer parts of documents, where lens distortion is felt stronger.

TABLE II
CHARACTER ERRORS FOUND IN IMAGES AFTER PROCESSING

		100	150	300	3Mp	4Mp
Character	Replacement	74	38	24	104	84
	Punctuation	123	3	1	40	12
	G. Accents	22	4	10	83	22
	Missing	38	27	5	17	16
	Insertion	43	92	19	25	19
Word	Errors	53	19	20	106	67
	Exclusions	-	-	-	5	1

Table II presents the OCR character and word errors found for the transcription of the 50 test images after undergoing border removal and skew correction, with vestigial border painted with the document most frequent color. Figure 02 present the test image from Figure 01 at this stage. The analysis of Table II shows that the use of portable cameras for document digitalization is a viable way of acquiring information even for automatic text transcription provided some simple pre-processing is performed. The number of

Missing and Inserted characters obtained by cameras was less than the one for the scanned documents.

7 Perspective correction

The freedom allowed in acquiring document images with portable digital cameras without mechanical support invariably leads to perspective distortion. Several algorithms in the literature address this problem [3, 12, 13, 14]. The correction of perspective distortion has border detection as a first step to find the polygon that margins the image and getting the four corner points that will serve as reference for the linear transformation. The image of the four corner points serve to crop the perspective corrected image and automatically performs skew correction. On the other hand, perspective distortion opens a number of alternatives which cause different effects in the quality of the image produced both in terms of visualization and OCR response. In general, as already mentioned when image skew was addressed in this paper, the skew angle was small (less than 2°), thus this means that the image tends to exhibit a trapezoidal shape. Two alternatives for correction arise: either to narrow the opening edges or to widen the closing edges. The latter alternative was discarded because the general trend is to disconnect contiguous areas, which has a serious degrading effect on OCR response. Thus, the better alternative is former. Three interpolation methods are commonly presented in the literature: closest neighbor, bilinear and bi-cubic. Figure 04 shows the effect each of such methods has whenever applied to a straight line.

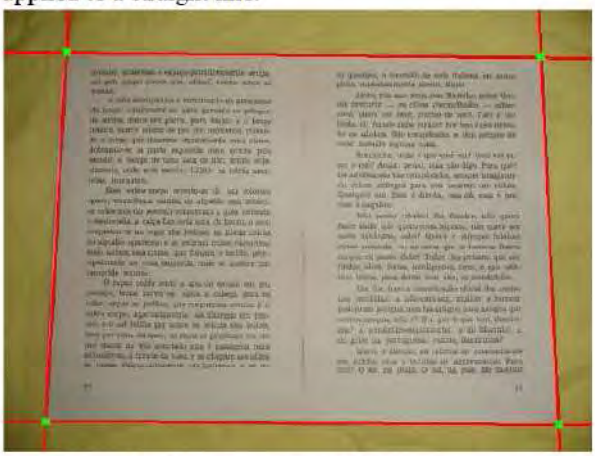


Figure 03 Image from Figure 01 showing perspective correction reference points and edges.

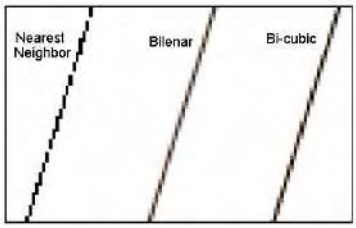


Figure 04 Three interpolation methods: nearest neighbor, bilinear and bi-cubic

The image obtained after perspective correction and cropping as shown in Figure 05 looks far more pleasant for the human reader. Figure 06 zooms at the result of different interpolation techniques on a word.

Figure 05 Image from Figure 01 after perspective correction with bi-cubic interpolation and cropping.

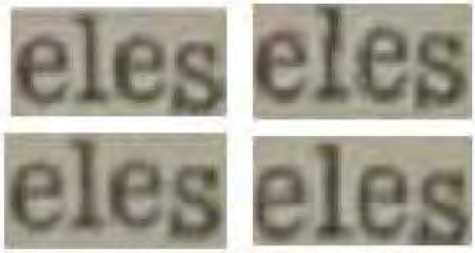


Figure 06 Zoom into a word of Figure 01.
Top: original, nearest neighbor,
Bottom: bilinear, and bi-cubic interpolations

Tables III and IV present the OCR response for documents, after the three interpolation techniques.

9. Conclusions and lines for further works

This paper provides a comparative analysis of the quality of documents acquired through 3.2 and 4.1 Mpixels Sony portable digital cameras in comparison with their scanned version with three different resolutions (100, 150 and 300 dpi). A batch of 50 documents was studied totaling 13,532 words and

2,095,596 characters. The quantitative analysis performed herein allows to conclude that portable digital cameras not only provide a simple way to digitalize documents to be read by humans, but the quality of documents allows means for image-to-text transcription using commercial OCRs. Today, portable 6 Mpixels cameras are often found yielding an image resolution closer to the one offered by scanners. Most possibly such cameras provide better OCR response, but this is still to be borne out by experiments. On the other end of camera technology one finds portable phones and PDA's whose cameras are reaching the 3.2 Mpixel barrier.

Several challenges are faced to improve OCR performance. Illumination and compensating the effect of the embedded strobe flash are two of the most important ones as they pose difficulties to image binarization.

10. References

[1] D.Doermann, J.Liang, H. Li, "Progress in C.-Based Document Image Analysis," ICDAR'03, Vol(1): 606, 2003.

[2] J. Liang, D. Doermann and H. Li. Camera-Based Analysis of Text and Documents: A Survey. International Journal on Document Analysis and Recognition, 2005.

[3] R.D.Lins, A.R.Gomes e Silva and G.Pereira e Silva, Enhancing Document Images Acquired Using Portable Digital Cameras, ICIAR'07, LNCS, Springer-Verlag, 2007.

[4] Nuance Corp. http://www.nuance.com/omnipage/professional

[5] R.D.Lins and N.F.Alves. A New Technique for Assessing the Performance of OCRs. IADIS – Int. Conf. on Comp. Applications, IADIS Press, v. 1, p. 51-56, 2005.

[6] J. M. M. da Silva *et al.* Binarizing and Filtering Historical Documents with Back-to-Front Interference, ACM-SAC 2006, Nancy, April 2006.

[7] K.C.Fan, Y.K.Wang, T.R.Lay, Marginal noise removal of document images, Patt.Recognition. 35, 2593-2611, 2002.

[8] Lu S and C L Tan, Camera document restoration for OCR, CBDAR 2005/ICDAR 2005, Seoul, Korea.

[9] R. Gomes e Silva and R. D.Lins. Background Removal of Document Images Acquired Using Portable Digital Cameras. LNCS 3656, p.278-285, 2005.

[10]H.S.Baird, Document image defect models and their uses, ICDAR'93, Japan, IEEE Comp. Soc., pp. 62-67, 1993.

[11]L.G.Shapiro and G.C.Stockman, Computer Vision, March 2000. http://www.cse.msu.edu/~stockman/Book/book.html.

[12]L. Jagannathan and C. V. Jawahar, "Perspective correction methods for camera based document analysis," pp. 148–154, CBDAR 2005, Seoul, Korea. 2005.

[13]P. Clark, M. Mirmehdi, "Recognizing Text in Real Scenes", IJDAR, Vol. 4, No. 4, pp. 243-257, 2002.

[14]Clark, M. Mirmehdi, "On the Recovery of Oriented Docs. from Single Images", CSTR-01-004, Bristol, 2001.

[15]M. Seeger, C. Dance, "Binarising Camera Images for OCR", Proc. of Sixth ICDAR, p. 0054-0059, Sept. 2001.

[16]Lu S, Tan CL, The restoration of camera documents through image segmentation, LNCS 3872: 484-495 2006.

[17]P. Clark, M. Mirmehdi, "Location and Recovery of Text on Oriented Surfaces", SPIE CDRR VII, pp. 267-277, 2000.

[18]S. J. Lu, *et al*, "Perspective rectification of document etc.," Image and Vision Computing, V(23):541-553, 2005.

[19]T.Kanungo, R.M.Haralick, I.Phillips, Global and local document degradation models, ICDAR, pp. 730-734, 1993.

TABLE III		No correction		Nearest neighbor	
		3Mp	4Mp	3Mp	4Mp
Character	Replacement	104	84	130	81
	Punctuation	40	12	74	21
	G. Accents	83	22	67	24
	Missing	17	16	90	10
	Insertion	25	19	33	28
Word	Errors	106	67	151	83
	Exclusions	5	1	12	1

TABLE IV		Bilinear		Bi-cubic	
		3Mp	4Mp	3Mp	4Mp
Character	Replacement	123	57	136	55
	Punctuation	77	40	76	39
	G. Accents	70	27	66	27
	Missing	92	4	94	4
	Insertion	33	30	43	19
Word	Errors	154	67	157	65
	Exclusions	12	0	12	0

PhotoDoc: A Toolbox for Processing Document Images Acquired Using Portable Digital Cameras

Gabriel Pereira e Silva and Rafael Dueire Lins,
Departamento de Eletrônica e Sistemas – Universidade Federal de Pernambuco - Brazil
gfps@cin.ufpe.br, rdl@ufpe.br

Abstract

This paper introduces PhotoDoc a software toolbox designed to process document images acquired with portable digital cameras. PhotoDoc was developed as an ImageJ plug-in. It performs border removal, perspective and skew correction, and image binarization. PhotoDoc interfaces with Tesseract, an open source Optical Character Recognizer originally developed by HP and distributed by Google.

1. Introduction

Portable digital cameras are omnipresent in many ways of life today. They are not only an electronic device on their own right but have been embedded into many other devices such as portable phones and palmtops. Such availability has widened the range of applications, some of them originally unforeseen by their developers. One of such applications is using portable digital cameras to acquire images of documents as a practical and portable way to digitize documents saving time and the burden of having either to scan or photocopy documents. Figures 01 to 04 present different documents digitized using different camera models.

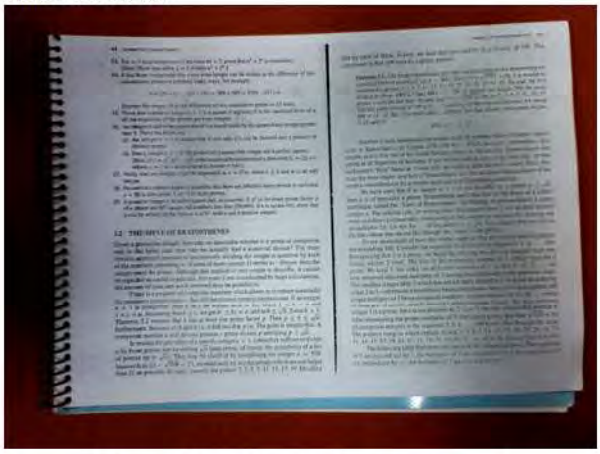


Figure 01 – Document image acquired with the camera of a LG cell phone KE-970 – (1600x1200 pixels) 409KB, without strobe flash

Figure 02 - Document image acquired with the camera of HP iPaq rx3700 (1280x960 pixels) 162KB, without strobe flash

This new use of portable digital cameras gave birth to a new research area [1][2] that is evolving fast in many different directions.

Figure 03 - Document image acquired with camera Sony DSC-P40 (4.1 Mpixels)

Figure 04 - Document image acquired with camera Sony DSC-P52 (3.2 Mpixels)

New algorithms, tools and processing environments are needed to provide users in general with simple ways of visualizing, printing, transcribing, compressing, storing and transmitting through networks such images. PhotoDoc is a processing environment developed to meet such needs.

The test documents used here were obtained in true-color, under different illumination conditions, with and without the inbuilt camera strobe flash, using a portable cell phone manufactured by LG KE-970 – (1600x1200 pixels) 409KB without strobe flash, a HP iPaq rx3700 (1280x960 pixels) 162KB without strobe flash, and two different models of portable cameras manufactured by Sony Corp. (models DSC-P52 and DSC-P40) of 3.2 and 4.1 Mega pixels, respectively. All cameras were set into "auto-focus" mode, i.e. the user leaves to the device the automatic setting of the focus.

Several specific problems arise in this digitalization process and must be addressed to provide a more readable document image, which also claims less toner to print, less storage space and consumes less bandwidth whenever transmitted through networks. The first of all is background removal as document photograph goes beyond the document size and incorporates parts of the area that surrounds it. The absence of mechanical support for taking the photo yields a non-frontal perspective that distorts and skews the document image. The distortion of the lenses of the cameras makes the perspective distortion not being a straight but a convex line, depending on the quality of the lens and the relative position of the camera and the document. Non-uniform illumination of the environment and strobe flash, whenever available in the device adds difficulties in image enhancement and binarization.

2. The PhotoDoc Environment

PhotoDoc was conceived as a device independent software tool to run on PCs. Whenever the user unloads his photos he will be able to run the tool prior to storing, printing or sending through networks the document images. The algorithms and the basic functionality of PhotoDoc may be incorporated to run on a device such as a PDA or even a camera itself. Such a possibility is not considered further herein. Due to implementation simplicity and portability the current version of PhotoDoc was implemented as an ImageJ [20] Plug-in. ImageJ is an open source image processing environment in Java developed by Wayne Rasband, is at the Research Services Branch, National Institute of Mental Health, Bethesda, Maryland, USA. Figure 05 shows a screen shot of PhotoDoc being activated from ImageJ.

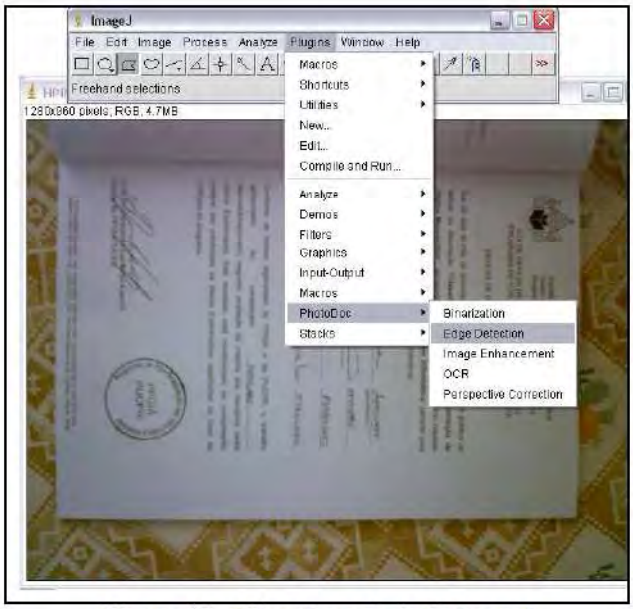


Figure 05 – PhotoDoc plugin in ImageJ

As one may observe in Figure 05, the present version of the PhotoDoc plug-in offers five different filters, which appear in alphabetical order:

- Binarization
- Edge Detection
- Image Enhancement
- OCR, and
- Perspective Correction and Crop

The fact that PhotoDoc is now in ImageJ also allows the user to experiment with the different filters and other plug-ins already present in ImageJ. It is important to stress, however, that ImageJ as an open code library allows the developer to extract from it only the needed functionality in such a way that the developer may provide to ordinary user a PhotoDoc interface that looks independent from ImageJ. At present, the authors of this paper consider such possibility premature. Such tool particularization seems to be more adequate is coupled with a particular device, which allows also a better fine tuning of the algorithms developed for and implemented in PhotoDoc. In what follows the PhotoDoc filter operations are described.

3. A New Border Detection Algorithm

The very first step to perform in processing a document image in PhotoDoc is to detect the actual physical limits of the original document [3]. The algorithm presented in reference [7] was developed based on images acquired by 3 and 4 Mpixel cameras. Unfortunately, its performance in lower resolution cameras has shown to be inadequate.
A new edge detection algorithm, based on ImageJ filters, was developed and is presented herein. This new algorithm behaved suitably on a wide range of images with different kinds of paper, including glossy ones. The new algorithm was obtained by composing existing filters in ImageJ. The steps of the new border removal algorithm are:

1. Process + Enhance Contrast.
2. Process + Find Edges.
3. Image Type 8 bits.
4. Image Adjust Threshold – Black and White value 122.
5. Process Binary + Erode.

At this point the resulting image provides well defined borders that allow finding the document edges. Figure 06 presents the result of applying the steps of the algorithm above to the document image presented in Figure 01, which appears in the top-left corner, until reaching the resulting image in the bottom-right one.

Figure 06 – Step by step filtering of image for border edge detection using ImageJ

The result of the activation of button "Edge Detection" on a document in PhotoDoc yields an image such as the one presented on Figure 07.

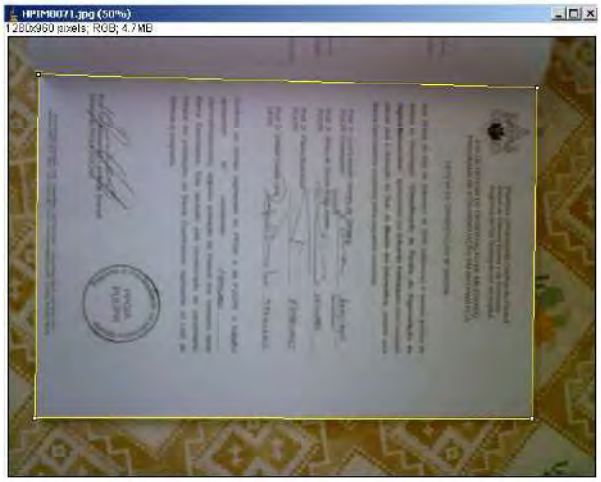


Figure 07 - Document image with edges (in yellow) automatically detected by PhotoDoc

Although the algorithm presented correctly detected edges for all the tested documents, the Edge Detection filter in PhotoDoc allows the user to adjust the edges by dragging along the four corners of the document image. This may also be of help whenever the document photo is not completely surrounded by a border of whenever strong uneven illumination causes edges not to be detected.

4. Perspective Correction and Crop

PhotoDoc incorporates a filter to correct the distortion and skew of the image introduced by the non-frontal position of the camera in relation to the document. A pinhole model for the camera was adopted [5, 6, 9] and provides a simple way to solve the problem at first. The experiments reported in [8] point at, narrowing edges and using bi-cubic interpolation as the rule-of-thumb to yield images more pleasant for the human reader and also with less transcription errors in OCR response. The difficulty inherent to such transformation is finding the four corners of the original image that will be taken as the basis for the correction. Edge or border detection, as explained above, is the first step the image should undergo. Once the document edges are determined as shown in Figure 07 the "Perspective Correction" filter in PhotoDoc may be called as shown in Figure 08.

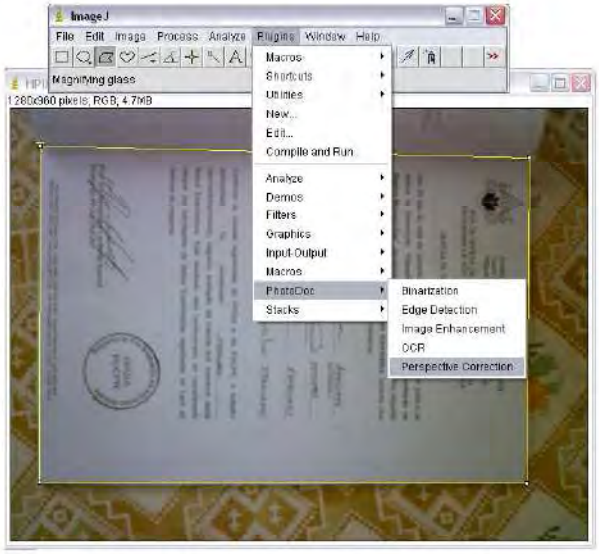


Figure 08 – Activating Perspective Correction in edge detected image from Figure 07 in PhotoDoc

PhotoDoc will automatically perform a perspective, skew, and orientation correction and then crop the resulting image, yielding a document as the one shown in Figure 09. One should observe that the cropped document whenever visualized in a screen display provides a far better image to the human reader, if printed saves toner and yields a more readable document, and whenever stored or transmitted through computer networks saves on average 25% of space [7].

Figure 09 – Cropped document after perspective, skew and orientation correction by PhotoDoc

The perspective correction algorithm in PhotoDoc was implemented using the JAI (Java Advanced Imaging) library [23].

5. Image Enhancement

There is a paramount number of possibilities to improve the quality of document images depending of the features of the camera, such as its resolution, lens distortion, the quality of embedded algorithms, environment illumination, quality and color of the paper, etc. Devices have many different technical characteristics, thus it is impossible to find a general solution that would suit all of them, overall with their fast technological evolution pace. However, in the case of devices with embedded flash, if it was not used the resulting document photograph "looked" slightly blurred (out-of-focus). Most possibly, this is related with the fact that the diaphragm of the objective stayed open for much longer to compensate the loss in illumination. As no mechanical support was used to stabilize the camera, chances are that the photographer moved briefly during the shot. Some other times, a slight inclination of the camera in a clear environment may be enough to the luminosity sensor to assume that there is no need for the camera to provide extra

illumination, canceling the flash activation. Thus, in order to minimize these factors one may recommend that document photos are:

1. Taken with the embedded flash of the camera set as "on", forcing its activation regardless of the luminosity of the environment.
2. Obtained indoors.

Several of the ImageJ image enhancement algorithms were tested. Non-uniform illumination brings a high degree of difficulty to the problem. A general algorithm that provided gains in all the images studied weakening the illumination problem was provided by the "Enhance Contrast" filter in ImageJ, as may be observed in the image presented in Figure 10.

Figure 10 – PhotoDoc enhanced version of Figure 09.

The "Enhance Contrast" filter in ImageJ performs histogram stretching. The *Saturated Pixels* value determines the number of pixels in the image that are allowed to become saturated. Increasing this value will increase contrast. This value should be greater than zero to prevent a few outlying pixels from causing the histogram stretch to not work as intended. For the case of PhotoDoc the best results were obtained with 0.5% of saturated pixels.

6. Document Binarization

Monochromatic images are the alternative of choice for most documents with no iconographic or artistic value saving storage space and bandwidth in network transmission. Most OCR tools pre-process their input images into grayscale or binary before character recognition. Reference [8] reports on the binarization of documents acquired with portable digital cameras. Fifty test images obtained with 3.2 and 4.1 Mpixel cameras (Sony DSC-P52 and DSC-P40, respectively) had their borders removed and were perspective and skew corrected before binarization, both globally and also splitting the images into 3, 6, 9 and 18 regions [16]. The following algorithms were tested:

1. Algorithm 1 – da Silva *et al.* [10];
2. Algorithm 2 – Mello *et al* [10];
3. Algorithm 3 – Pun [11];
4. Algorithm 4 – Kapur-Sahoo-Wong [12];
5. Algorithm 5 – Wu-Songde-Hanqing [13];
6. Algorithm 6 – Otsu [14];
7. Algorithm 7 – Yen-Chang-Chang [15];
8. Algorithm 8 – Johannsen-Bille [16].

According to [8], the global algorithm that yielded the best results both for visual inspection and in OCR response was the entropy based algorithm by da_Silva *et al* [10] (Figure 11). The best results obtained by applying the binarization algorithms in regions of the documents were provided by the algorithm by Kapur-Sahoo-Wong [12] with 18 regions (Figure 12).

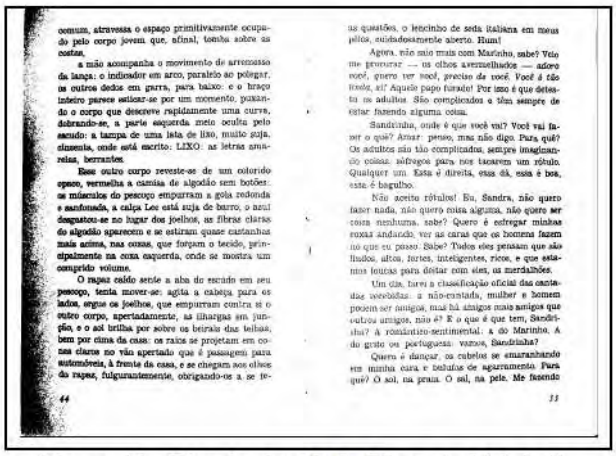


Figure 11. Binarized with da Silva *et al.*(global)

Some new algorithms were tested herein including Sauvola [18], Niblack [17] and MROtsu [14]. The results obtained may be found in Figures 13 to 15. Unfortunately, the results obtained for binarizing images captured by the devices without embedded strobe flash were unsatisfactory. Further analysis and pre-processing must be studied for such images.

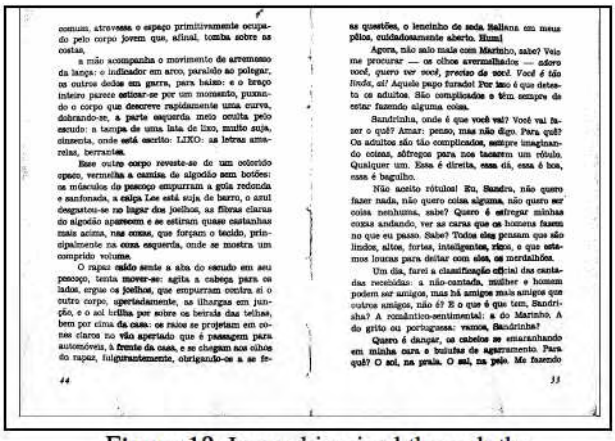


Figure 12. Image binarized through the Kapur-Sahoo-Wong algorithm (18 regions).

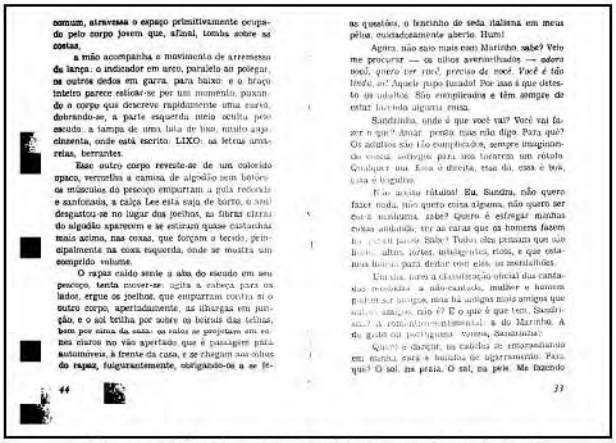


Figure 13.Binarized with Sauvola (global).

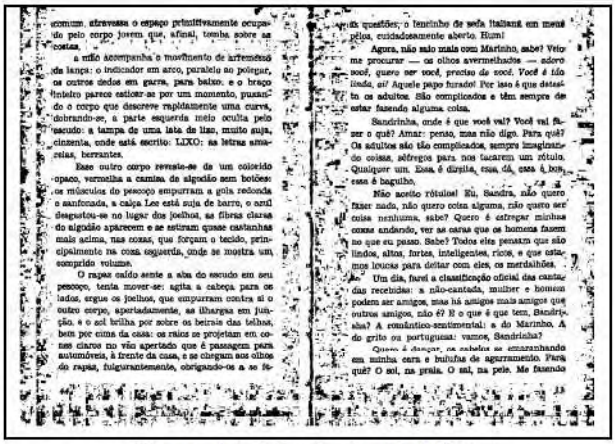


Figure 14. Binarized with Niblack (local, window_size = 50, k = -0.02)

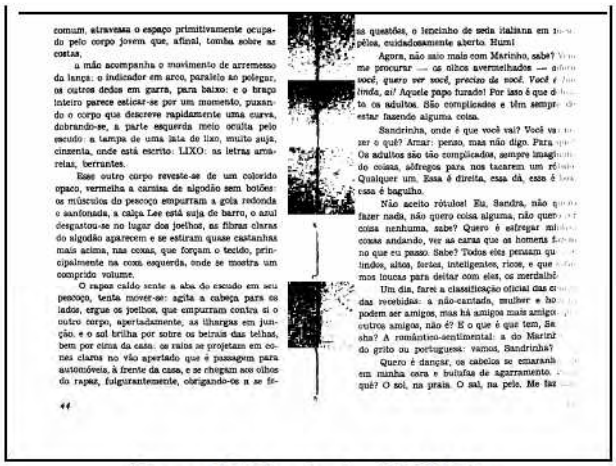


Figure 15.Binarized with MROtsu

PhotoDoc provides to the user two different ways of performing image binarization. The first one is "Automatic" and the second one opens a menu with all the algorithms above that may work either in global or image segmented mode with 2, 4, 8 and 16 regions. Local algorithm such as Niblack allows the user to define the parameters. The current implementation of "Binarization + Automatic" is set to run the algorithm by da Silva *et al.*(global) [10]. Later implementations may encompass a statistical analysis of images to choose the most suitable filter to a given image.

7. OCR

Optical Character Recognition is one of the key functionalities in any modern document processing environment [4]. Even when the recognition output is poor it may provide enough information for document indexing end keyword search. Reference [8] assesses the quality of the transcription of document images acquired with portable digital cameras with Omnipage Professional Edition version 15.0 [21]. It shows that the performance of the transcription of 50 of such documents obtained with the same models of Sony cameras used herein is close to the performance of the 100 dpi scanned counterparts.

ImageJ allows the call of executable code from within. The Tesseract [22] is an OCR Engine developed at HP Labs between 1985 and 1995. It was one of the top 3 engines in the 1995 UNLV Accuracy test. Since then, it has had little work done on it, but it is probably one of the most accurate open source OCR engines available today. The source code reads a binary, grey or color image and output text. PhotoDoc OCR whenever chosen activates the Tesseract OCR Engine. Preliminary tests such as the one made by submitting

the image in the top part of Figure 15 provided as output the text in the bottom part of Figure 15.

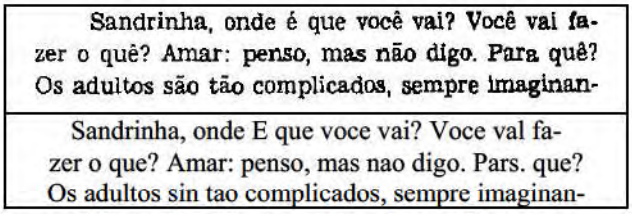


Figure 15 – Top: segment of textual image
Bottom: Transcribed text by Tesseract OCR

One should remark that better transcription could possibly be obtained if a Portuguese dictionary were used in conjunction with the Tesseract OCR. According to the measure of OCR performance presented in [19] the transcription above reached the figures presented on Table I.

TABLE I
ORIGINAL CHARACTER AND WORD ERRORS FOUND IN IMAGES

Character	Replacement	1	Word	Errors	9
	Punctuation	1		Exclusions	0
	G. Accents	8			
	Missing	0			
	Insertion	0			

From Table I one may see that the transcription errors were simple to be corrected as neither words nor characters are missing and there is no character insertion in the text. Besides that, word errors appeared in isolation and no word presented more than one character error in it. Most errors were due to the absence of Graphical Accents.

Conclusions and Lines for Further Works

PhotoDoc is a user friendly tool for processing document images acquired using portable digital cameras. Its current version was developed as a plug-in in ImageJ an open source portable Java library. PhotoDoc runs on users' PC and is device and manufacturer independent, working suitably from low end images acquired using cameras embedded in cell phones and palmtops to better models such as medium range, 3 and 4 Mpixels cameras, or even the state-of-the-art 6 Mpixels devices.

Several lines may be followed to provide further improvements to PhotoDoc filters. Most possibly, the most important of them is studying ways of compensating uneven illumination in documents. In the case of cameras with embedded strobe flash this may be simpler because the light source emanates from a specific point. In the case of devices that have no embedded flash illumination is provided by the environment and may come from different sources in position and power, increasing the complexity of its compensation. On the OCR front, much may be done ranging from providing dictionary help to Tesseract to performing post-processing in the textual output. This feature is already part of the newest (Jul/2007) version of Tesseract. Other open source OCR´s may also be analyzed, tested and incorporated to PhotoDoc. Preliminary testing with the open-source OCR engine OCRopus [24] showed that it was outperformed by the Tesseract OCR. For conclusive results, further testing is needed.

The PhotoDoc code is freely available at: http://www.telematica.ee.ufpe.br/sources/PhotoDoc

Acknowledgements

The work reported herein was partly sponsored by CNPq – Conselho Nacional de Desenvolvimento Científico e Tecnológico, Brazilian Government.

References

[1] D.Doermann, J.Liang, H. Li, "Progress in Camera-Based Document Image Analysis," ICDAR'03, V(1): 606, 2003.
[2] J. Liang, D. Doermann and H. Li. Camera-Based Analysis of Text and Documents: A Survey. International Journal on Document Analysis and Recognition, 2005.
[3] K.C.Fan, Y.K.Wang, T.R.Lay, Marginal noise removal of document images, Patt.Recognition. 35, 2593-2611, 2002.
[4] Lu S and C L Tan, Camera document restoration for OCR, CBDAR 2005/ICDAR 2005, Seoul, Korea.
[5] L.G.Shapiro and G.C.Stockman, Computer Vision, March 2000. http://www.cse.msu.edu/~stockman/Book/book.html.
[6] L. Jagannathan and C. V. Jawahar, "Perspective correction methods for camera based document analysis," pp. 148–154, CBDAR 2005/ICDAR 2005, Seoul, Korea. 2005.
[7] R. Gomes e Silva and R. D.Lins. Background Removal of Document Images Acquired Using Portable Digital Cameras. LNCS 3656, p.278-285, 2005.
[8] R.D.Lins, A.R.Gomes e Silva and G.Pereira e Silva, Assessing and Improving the Quality of Document Images Acquired with Portable Digital Cameras, ICDAR'07, Curitiba, Brasil, 2007.
[9] H.S.Baird, Document image defect models and their uses, ICDAR'93, Japan, IEEE Comp. Soc., pp. 62-67, 1993.
[10]J. M. M. da Silva *et al.* Binarizing and Filtering Historical Documents with Back-to-Front Interference, ACM-SAC 2006, Nancy, April 2006.
[11]T. Pun, Entropic Thresholding, A New Approach, C. Graphics and Image Processing, 16(3), 1981.

[12] J. N. Kapur, P. K. Sahoo and A. K. C. Wong. A New Method for Gray-Level Picture Thresholding using the Entropy of the Histogram, Computer Vision, Graphics and Image Processing, 29(3), 1985.
[13] L. U. Wu, M. A. Songde, and L. U. Hanqing, An effective entropic thresholding for ultrasonic imaging, ICPR'98: Intl. Conf. Patt. Recog., pp. 1522–1524 (1998).
[14] M.R .Gupta, N.P. Jacobson and E.K. Garcia. OCR binarization and image pre-processing for searching historical documents. Pattern Recognition 40 (2007): 389-397 (2007).
[15] J. C. Yen,*et al.* A new criterion for automatic multilevel thresholding. IEEE T. Image Process. IP-4, 370–378 (1995).
[16] G. Johannsen and J. Bille. A threshold selection method using information measures. ICPR'82: 140–143 (1982).
[17] W.Niblack, "An Introduction to Image Processing" pp.115-116, Prentice-Hall,Englewood Cliffs,NJ(1986).
[18] J. Sauvola, M. Pietikainen, Adaptive document image binarization, Pattern Recognition 33 (2) (2000) 225–236.
[19] R.D.Lins and N.F.Alves. A New Technique for Assessing the Performance of OCRs. IADIS – Int. Conf. on Comp. Applications, IADIS Press, v. 1, p. 51-56, 2005.
[20] ImageJ http://rsb.info.nih.gov/ij/
[21] Nuance Corp. http://www.nuance.com/omnipage/professional
[22] Tesseract http://code.google.com/p/tesseract-ocr/
[23] JAI (Java Advanced Imaging). https://jai.dev.java.net.
[24] OCRopus .http://www.ocropus.org.

Scanned or Photographed?
Automatically Deciding How a Document was Digitized

Gabriel Pereira e Silva,
Rafael Dueire Lins,
Brenno Miro
UFPE, Recife, Brazil
gfps@cin.ufpe.br, rdl@ufpe.br

Steven J.Simske
HP Labs, Fort Collins, USA
steven.simske@hp.com

Marcelo Thielo
HP Labs, Porto Alegre, Brazil
marcelo.resende.thielo@hp.com

Abstract

Portable digital cameras are being used widely by students and professionals in different fields as a practical way to digitize documents. Tools such as PhotoDoc enable the automatic processing of such documents, performing border removal and perspective correction. A PhotoDoc processed document and a scanned one look very similar to the human eye if both are in true color. However, if one tries to automatically binarize a batch of documents digitized from portable cameras compared to scanners, they have different features. The knowledge of their source is fundamental for successful processing. This paper presents a classification strategy to distinguish between scanned and photographed documents. Over 16,000 documents were tested with a correct classification rate of over 99.96%.

1. Introduction

Portable digital cameras are ubiquitous. Either in standalone versions, or incorporated in cell phones, the quality of the images has risen at a fast pace while their price has dropped drastically. Such pervasiveness has given rise to unforeseen applications such as using portable digital cameras for digitalizing documents by users of many different professional areas. For instance, students and professionals are taking photos of writing boards instead of taking notes; lawyers are taking photos of legal processes instead of going through a difficult bureaucratic path to take documents out of court to photocopy them, etc. This new research area [1][4] is evolving fast in many directions. General users, non-specialized in image processing, want new algorithms, tools and processing environments to be able to provide simple and user-friendly ways of visualizing, printing, transcribing, compressing, storing and transmitting document images. Figure 1 presents an example of a document acquired with a portable digital camera. Reference [6] points out some particular problems that arise in this document digitalization process: the first is background removal. Very often the document photograph goes beyond the document size and incorporates parts of the area that served as mechanical support for taking the photo of the document. The second problem is due to the skew often found in the image in relation to the photograph axes. As portable cameras have no fixed mechanical support, often there is some inclination in the document image. The third problem is non-frontal perspective, due to the same reasons that give rise to skew. A fourth problem is caused by the distortion of the lens of the camera. This means that the perspective distortion is not a straight line, but a convex arc, depending on the quality of the lens and the relative position of the camera and the document. The fifth difficulty in processing document images acquired with portable cameras is non-uniform illumination.

Figure 1. Example of a photo document

Figure 2. PhotoDoc processed photo document

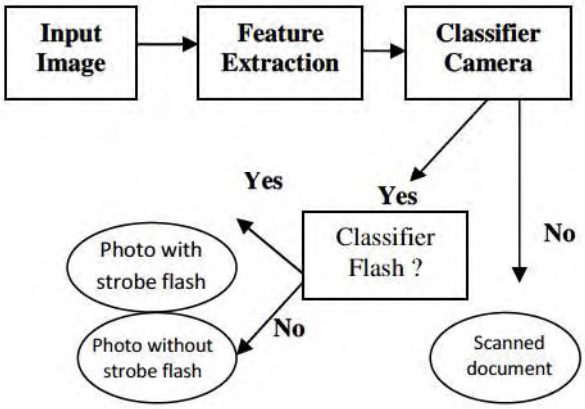
Reference [3] presents PhotoDoc, a freely available toolbox for processing document images acquired with portable digital cameras, which is implemented as a plugin in ImageJ [9]. Figure 2 presents an example of a photo document processed with PhotoDoc, which is implemented as a Plugin in ImageJ [11].

Figure 3. Binarization of a photo document using a global algorithm (Otsu)

Illumination is less uniform for documents captured with digital cameras in comparison to scanned images. It may not be easy for a person to differentiate between a document processed using PhotoDoc and the same document captured with a scanner. Distinguishing between them is important in the case, for example, of image binarization. The irregular illumination in general tends to provide shaded black areas in the direct binarization of a photo document as shown in Figure 3. Color images such as the one in Figure 4, both scanned and photographed, are also present in the test set used here.

Figure 4. PhotoDoc processed color photo document

This paper focuses on a classification strategy to distinguish, in a batch of documents, the scanned documents from documents acquired with portable digital cameras. Camera documents are further classified based on whether a strobe flash was used, as shown below:



The classification strategy depends on the following:

- The choice of the set of features to be extracted. The features selected must provide enough elements to distinguish between the clusters of interest. Feature extraction has also impact in classification time.
- The choice of the classifier. Some classifiers are able to perform better than some others depending on the nature and class of the problem, the representativeness of the features selected, etc.
- The quality and size of the training set used for the classifier. The training set must be carefully chosen to encompass the whole diversity of the universe of objects to be classified, with as less redundancy as possible.

This paper shows that the classifier presented in reference [7] presents excellent performance for distinguishing between documents obtained from scanners and portable digital cameras with or without the strobe flash on. The results obtained are compared with the classification strategy in reference [8]. The new classifier not only reached a higher correct classification rate, but besides that, elapsed much less time for feature extraction and classification. The classifier presented herein was implemented using Weka [10, 12], an excellent, user friendly and open-source platform developed at the University of Waikato. The test set encompassed 17,781 documents of which only 3 documents were misclassified, yielding a correct classification rate of 99.98%.

2. Experiments Performed

The starting point for this work was collecting images that are representative of the two different clusters of interest: scanned and photographed documents. The photographed documents were split in two sub-clusters: images acquired with and without the strobe flash on.

The test set for the photo document cluster is formed by 9,573 documents acquired with a Sony Cybershot digital camera DSC-W55 in 5 and 7.2 Mpixels, with and without mechanical support, in-built strobe flash on and off. In the camera set there are also 404 photos taken with a portable camera Sony DSC-S40 and 60 photos from a cell phone LG Shine ME970, both without any mechanical support. All photo documents were processed with PhotoDoc [3] a photo document processing tool that crops the framing border and corrects perspective and skew, should be classified as "document".

The 6,444 of the scanned documents were digitized with a Ricoh Affício 1075 flatbed scanner in 100, 200 and 300 dpi saved into four different file formats: bmp (uncompressed), jpg (1% losses), png (lossless), and tiff (uncompressed), using the software provided by the scanner manufacturer. Although the jpeg file format may be seen as unsuitable for such kind of image it is often used by people in general [5]. In addition, 300 images were acquired with a scanner HP 5300c in 300 dpi, true color, stored in tiff (uncompressed) and 1000 jpeg images in different resolutions were collected from the Internet.

Table 1 shows the numbers of images per file format in the test set.

	JPG	PNG	TIFF	BMP	Total
Photo	10,037	****	****	****	**10,037**
Scanned	2,611	1,611	2,522	1,000	**7,744**
Total	**12,648**	**1,611**	**2,522**	**1,000**	**17,821**

Table 1 – Images per file formats

2.1 Features Tested

The choice of the features to be extracted and tested is the key to the success and performance of the classification. Image entropy is often used as the key for classification [8]. It has a large computational cost, however. Entropy calculation demands a scan in the image to calculate the relative frequency of a given color, for instance, which is than multiplied for its logarithm and added up. The classifier described in reference [9] is based on the binary classification approach, and assumes a Gaussian distribution for each of the features. Its performance degrades in proportion to the non-Gaussian nature of the data. We designate this the entropy-based classifier, as the set of features chosen herein has entropy calculation as its key.

The work presented in reference [7] proposes a new classification strategy that assumes that decreasing the gamut of an image, analyzed together with its grey scale and monochromatic equivalents would provide enough elements for a fast and efficient image classification. The features tested are:

- Palette (true-color/grayscale)
- Gamut
- Conversion into Grayscale
- Gamut in Grayscale (if RGB)
- Conversion into Binary (Otsu)
- Number of black pixels in binary image.
- (#Black_pixels/Total_#_pixels)*100%
- (Gamut/Palette)*100% (true-color/grayscale)

Image binarization is performed by using Otsu [8] algorithm. The data above are extracted for each image and placed in a vector of features.

The classification strategy adopted herein follows the feature set proposed in reference [9]. The training set used had size of about 8% of the test set and was selected from within the images of Sony Cybershot digital camera DSC-W55 in 5 and 7.2 Mpixel and the Ricoh Affício 1075 flatbed scanner in 100, 200 and 300 dpi. The images in the training set were not part of the test set. The entropy-based classifier [9] was used to compare the results obtained. Both classifiers used the same training and test sets.

2.2 Sub-sampling

Image sub-sampling may be used as a way to reduce the time elapsed in feature extraction of images to be classified. The key points in image sub-sampling are:

1- The larger the image file, the richer in data redundancy; thus, if the redundant data are thrown away the efficiency both in feature-collection time and classification may rise.
2- The selection of points to be analyzed for feature collection should not be random. It should somehow provide a "reduced" version of the original image (although in some cases it may be distorted by unequal scaling!).

Twenty different sub sampling strategies were evaluated on the images of this study. The cascaded subsampling strategy consisted of removing more points from the larger image files and provided the best overall accuracy of any classification schema, while simultaneously significantly improving the performance of the feature extractor, as shown in the next section. The cascaded subsampler performs the following operations:

size = height*width

- If size ≤ 300,000 break;
- If 300,000< size ≤ 500,000:
  remove even lines or columns
  (whatever the larger);
- If 500,000 < size ≤ 700,000:
  remove even lines and columns;
- If 700,000 < size ≤ 900,000:
  remove 2 lines in every 3 lines and even columns,
  (if height>width)
  remove even lines and
  2 columns in every 3 columns, otherwise;
- If 900,000 < size remove 2 lines and 2 columns
  in every 3 lines and columns;

Code for the "cascaded" sub-sampler

3. Results

The results of classification are presented in two steps. The group of results was obtained with 16,017 images digitized with the Sony Cybershot digital camera DSC-W55 in 5 and 7.2 Mpixel, and the Ricoh Affício 1075 flatbed scanner in 100, 200 and 300 dpi. Several different classifiers implement in Weka [10, 12] were tested. Random forests provided the best classification results amongst the statistical classifiers. A Multi-layer Perceptron (MLP) neural classifier was also tested and the best results obtained for eight neurons on two layers. The confusion matrices obtained by the classifiers that used the proposed set of features are shown in Table 2. The entry "Photo +sf" stands for the document images photographed with the strobe flash on, while "Photo –sf" denotes it off.

Classifier		Photo +sf	Photo -sf	Scanned	Accuracy %
Random Forest 5-trees	Photo +sf	4029	0	0	100
	Photo -sf	4	5534	6	99,81962
	Scanned	0	0	6444	100
Random Forest 10-trees	Photo +sf	4,029	0	0	100
	Photo -sf	4	5,537	3	99.8737
	Scanned	0	0	6,444	100
Random Forest 15-trees	Photo +sf	4029	0	0	100
	Photo -sf	7	5535	2	99,83766
	Scanned	0	0	6444	100
Random Forest 20-trees	Photo +sf	4029	0	0	100
	Photo -sf	8	5534	2	99,81962
	Scanned	0	0	6444	100
Random Forest 100-trees	Photo +sf	4029	0	0	100
	Photo -sf	7	5535	2	99,83766
	Scanned	0	0	6444	100
MLP	Photo +sf	4029	0	0	100
	Photo -sf	13	5531	0	99,76551
	Scanned	0	1	6443	99,98448

Table 2 – Confusion matrix of the **proposed** classifier with 16,017 original images

Table 2 points out that the Random Forest statistical classifier [1] with 10 trees presented the best classification results.

Table 3 shows the results obtained for the same set of classifiers trained and tested with subsampled images.

Classifier		Photo +sf	Photo -sf	Scanned	Accuracy %
Random Forest 5-trees	Photo +sf	4029	0	0	100
	Photo -sf	2	5525	17	99,65729
	Scanned	0	0	6444	100
Random Forest 10-trees	Photo +sf	4,029	0	0	100
	Photo -sf	0	5,540	4	99.9278
	Scanned	0	0	6,444	100
Random Forest 15-trees	Photo +sf	4029	0	0	100
	Photo -sf	2	5539	3	99,90981
	Scanned	0	0	6444	100
Random Forest 20-trees	Photo +sf	4029	0	0	100
	Photo -sf	2	5539	3	99,90981
	Scanned	0	0	6444	100
Random Forest 100-trees	Photo +sf	4029	0	0	100
	Photo -sf	3	5539	2	99,90981
	Scanned	0	0	6444	100
MLP	Photo +sf	4029	0	0	100
	Photo -sf	10	5531	3	99,76551
	Scanned	0	0	6444	100

Table 3 – Confusion matrix of the **proposed** classifiers with 16,017 subsampled images

Using subsampling, the relative performance of the classifiers was stable. Again, Random-forests with 10 trees provided the best results. Curiously, subsampling, besides speeding-up the feature extraction time, increased correct classification rate. One important point worth noting is that the misclassified documents, when binarized using a global algorithm, performed satisfactorily. Having the strobe flash off may resemble a scanned document, provided there is enough uniform illumination from the environment. Then, the misclassification errors in this case do not cause serious problems to the overall process.

Now, the entropy-based set of features for classification proposed by reference [9] was tested on the original data and the results obtained are presented on Table 4.

Proposed Classifier	Photo +sf	Photo -sf	Scanned	Accuracy
Photo +sf	3402	272	355	84.4378 %
Photo -sf	71	4466	1007	80.5555 %
Scanned	32	152	6260	97.1446 %

Table 4 – Confusion matrix of the **entropy-based** classifier with original images

The results obtained for entropy based classifier with subsampled images are shown on Table 5.

Proposed Classifier	Photo +sf	Photo -sf	Scanned	Accuracy
Photo +sf	3402	270	357	84.4378 %
Photo -sf	69	4562	913	82.2871 %
Scanned	24	158	6262	97.1756 %

Table 5 – Confusion matrix of the **entropy-based** classifier with subsampled images

The comparison between the entropy-based and the new one proposed here shows that the new one is about 10% better than the previous one.

The classification of the 404 photos taken with a portable camera Sony DSC-S40 and 60 photos from a cell phone LG Shine ME970, both without any mechanical support, and the images obtained with scanner HP 5300c and the images collected from the Internet did not bring any misclassification at all.

4. Time Performance

Table 6 presents the feature extraction and classification times along with the programming language used for implementation. Besides classification accuracy per cluster, the average feature extraction and classification times are presented. One should also remark that there is a difference in time scale between feature extraction and classification.

	Feature extraction		Classification	
	Time (s)	Language	Time (ms)	Language
Original	**0.4174**	**C++**	**0.12**	**C#**
Subsampled	**0.1470**	**C++**	**0.12**	**C#**
Original	**0.4174**	**C++**	**0.10**	**C++**
Subsampled	**0.1470**	**C++**	**0.10**	**C++**
Entropy Or.	**1.4576**	**C#**	**6.13**	**C#**
Entropy Ss.	**0.497**	**C#**	**6.13**	**C#**

Table 6 – Feature extraction and classification times

Table 6 shows that the set of features used for image classification based on image palette conversion outperforms the entropy-based classifier by a factor of four for feature extraction and by a factor of fifty for image classification. ("Entropy Or." stands for the Entropy-based classifier [9] with the original images,

while "Entropy Ss." corresponds to the Entropy-based classifier with subsampled images).

The figures of the relative performance of the classifiers for the proposed set of features varying the number of trees and the MLP implemented in Weka (Java) are shown on Table 7.

Proposed Classifier	Time (ms)
Random Forest 5-trees	5.4
Random Forest 10-trees	6.1
Random Forest 15-trees	6.7
Random Forest 20-trees	7.9
Random Forest 100-trees	9.5
MLP	6.8

Table 07 – Classification times in Weka (Java)

One may observe that the Random-forests classifier reaches the best trade-off classification and time efficiency.

5. Conclusions

Weka [10, 12] has shown to be an excellent test bed for statistical analysis. The choice for a Random tree classifier was made after performing several experiments with the large number of alternatives offered by Weka, although results did not vary widely. Amongst them a preliminary comparison between the new statistical classifier proposed here and a MLP neural classifier provided worse results (around 94% of accuracy).

The choice of the images in the training set is of paramount importance to the performance of the classifier. They must be representative of the whole universe of images in a cluster.

The classification scheme presented in this paper increased the correct classification rate by more than 10%. This automatic classification allows distinguishing scanned from photographed document images yielding better ways to suitably process document images.

Acknowledgements

Research presented herein was partly sponsored by CNPq – Conselho Nacional de Desenvolvimento Científico e Tecnológico, and HP – UFPE Project TechDoc sponsored by MCT, both of the Brazilian Government.

6. References

[1] L. Breiman, "Random Forests", Machine Learning, 45(1), pp. 5-32, 2001.

[2] D. Doermann, J. Liang, H. Li, "Progress in Camera-Based Document Image Analysis," ICDAR'03, Volume (1): 606, 2003.

[3] G. Pereira e Silva and R.D. Lins. PhotoDoc: A Toolbox for Processing Document Images Acquired Using Portable Digital Cameras. CBDAR 2007, pp.107-114, 2007.

[4] J. Liang, D. Doermann and H. Li. Camera-Based Analysis of Text and Documents: A Survey. International Journal on Document Analysis and Recognition, 2005.

[5] R.D. Lins and D.S.A. Machado, A Comparative Study of File Formats for Image Storage and Transmission, vol. 13(1):175-183, Journal of Electronic Imaging, 2004.

[6] R.D. Lins, A.R. Gomes e Silva and G. Pereira e Silva, Enhancing Document Images Acquired Using Portable Digital Cameras, ICIAR'07, LNCS, Springer-Verlag, 2007.

[7] R.D. Lins, G.F. Pereira e Silva, S.J. Simske, J. Fan, M.S. Shaw, P. Sá, M.R. Thiello, Image Classification to Improve Printing Quality of Mixed-Type Documents. ICDAR 2009, IAPR Press, Barcelona, 2009.

[8] N. Otsu. "A threshold selection method from gray level histograms". IEEETrans.Syst.Man Cybern. Vol. (9):62-66, 1979.

[9] S.J. Simske, "Low-resolution photo/drawing classification: metrics, method and archiving optimization," *Proceedings IEEE ICIP*, IEEE, Genoa, Italy, pp. 534-537, 2005.

[10] I.H. Witten, E. Frank. Data Mining: Practical Machine Learning Tools and Techniques (2nd Edition) –- Morgan Kaufmann, June 2005. ISBN 0-12-088407-0.

[11] ImageJ http://rsb.info.nih.gov/ij/

[12] Weka 3: Data Mining Software in Java, website http://www.cs.waikato.ac.nz/ml/weka/.

2009 10th International Conference on Document Analysis and Recognition

Image Classification to Improve Printing Quality of Mixed-Type Documents

Rafael Dueire Lins, Gabriel Pereira e Silva
UFPE, Recife, Brazil
rdl@ufpe.br, gfps@cin.ufpe.br

Steven J.Simske, Jian Fan, Mark Shaw,
HP Labs., Palo Alto, USA
{steven.simske, jian.fan, mark.q.shaw}@hp.com

Paulo Sá, Marcelo Thielo
HP Labs., Porto Alegre, Brazil
paulo.sa@hp.com
marcelo.resende.thielo@hp.com

Abstract

Functional image classification is the assignment of different image types to separate classes to optimize their rendering for reading or other specific end task, and is an important area of research in the publishing and multi-Average industries. This paper presents recent research on optimizing the simultaneous classification of documents, photos and logos. Each of these is handled during printing with a class-specific pipeline of image transformation algorithms, and misclassification results in pejorative imaging effects. This paper reports on replacing an existing classifier with a Weka-based classifier that simultaneously improves accuracy (from 85.3% to 90.8%) and performance (from 1458 msec to 418 msec/image). Generic subsampling of the images further improved the performance (to 199 msec/image) with only a modest impact on accuracy (to 90.4%). A staggered subsampling approach, finally, improved both accuracy (to 96.4%) and performance (to 147 msec/image) for the Weka-base classifier. This approach did not appreciable benefit the HP classifier (85.4% accuracy, 497 msec/image). These data indicate staggered subsampling using the optimized Weka classifier substantially improves the classification accuracy and performance without resulting in additional "egregious" misclassifications (assigning photos or logos to the "document" class).

1. Introduction

Image clustering has been researched by the database community since the early 1980's aiming to make efficient information retrieval in image databases [1][2]. In that kind of application one image is used to search the database looking for either the same or similar images. The basic idea is to try to organise the images in the database using some "common" features [3][2]. The same "features" are used to analyse the image that will serve as the "search-key". Instead of stepping through the whole database image-by-image, the retrieval process tries to match the properties of the search-key image, also known as *query image*, with the different image clusters in the database. This largely reduces the search-space making the retrieval process far more efficient. One of the features that has presented greater success in image retrieval was the analysis and clustering by the colour histogram [4][5]. The semantics of images have also been used as a clustering method [6] in database retrieval. Images that have similar "motifs" are most likely to have properties that are common to each other forming clusters. On the other hand, images whose theme completely uncorrelated should exhibit very different properties.

Image classification is used in all-in-one and multi-functional devices to differentially render images belonging to different clusters. In particular, document, photo and logo images require widely different imaging pipelines to optimize their appearance when copied or printed. Documents (text, tables), for example, require sharpening that would damage the appearance of photos and logos. Logos use a palette that would "posterize" photos. Photos, in turn, can be rendered with a lower resolution (but greater bit depth) than either documents or logos. Figure 01 shows examples of typical representatives of the three classes of interest herein.

Photo **Document** **Logo**

Figure 01. Example images of the clusters of interest

This paper improves the results of the classifier described in reference [7] and presents a new, Weka-

978-0-7695-3725-2/09 $25.00 © 2009 IEEE
DOI 10.1109/ICDAR.2009.167

based classifier [8] used to distinguish between these three types of images. Part 2 describes the experiments performed. Part 3 summarizes the results. We conclude with a discussion of the results.

2. Experiments Performed

The starting point for this work was collecting images that are representative of the different clusters of interest. Images were classified one-by-one by one person and the data-set was checked by three other people to avoid misclassifications and repetitions of the same image. Sometimes the "same" image appears in the test set in different file formats, for instance an image may appear in jpg, tiff, and bmp, as their features (palette, gamut, size) change from a format to another. Figure 01 shows an example of each of the classes of image of interest for this work. Images that do not belong to any of those classes are classified as "Don´t know".

The "Photo" cluster encompassed many different sorts of photos, which ranged from people, landscapes, objects and even documents. Most photos were true-color although there were grey scale ones. The resolution also varied widely from VGA (480x640 pixels) to 7.2 Mpixels. The photos were collected from family albums of the people linked to the authors to ones obtained from the Internet. As professionals of many different areas start to use portable digital cameras to acquire images of documents, such images were included in this study, bringing an extra level of difficulty: a document acquired with a camera is classified as a photo or a document? The answer to that question is not straightforward and may puzzle even humans. The criterion adopted here was that if the image encompasses only the document it is classified as "document" if parts of the surroundings are included it is classified as "photo". That means that if the document image in the rightmost part of Figure 02 is classified as "photo" and the same image after being processed in a tool such as PhotoDoc [9]. The "photo" test set has 7,968 photos of people and landscape and 500 photos of documents.

Grey-level	Landscape	Document

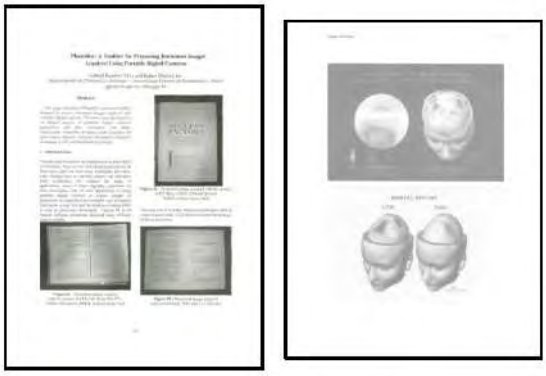
Figure 02. Example of photos

The 3,051 logos in the test set were collected from the Internet and from many different sources. Logos tend to exhibit a palette with a small number of colors, although very often they are saved in jpeg file format, introducing hues not originally intended. Figure 03 presents some examples of images classified as logos.

Figure 03. Examples of logos

The "Document" cluster was formed by 3,856 images of documents acquired from several different ways. Five hundred documents were photographed with a Sony Cybershot digital camera DSC-W55 in 5 and 7.2 Mpixels, with and without mechanical support, in-built strobe flash on and off, and then processed with PhotoDoc [9] that crops the framing border and corrects perspective and skew, should be classified as "document". About half of the remaining documents were scanned documents with different resolutions (from 100 to 300 dpi) and saved in bmp, tiff, and jpeg, which although not suitable for such kind of image is often used by people in general [10]. The remaining photos were obtained by saving Adobe pdf documents into tiff and jpeg.

Figure 04 presents some examples of document images used in this work.

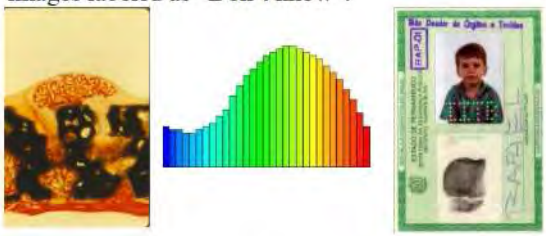


PhotoDoc processed **Scanned document**

Figure 04. Examples of documents

The last cluster of images in the test set is the "Don´t Know" images. These images were included as to increase the possibility of misclassifications. They are images that appear in the "real world" and range widely in nature from biological images, to vector graphics (obtained by softwares such as Excell®,

Powerpoint®, etc.). Figure 05 shows examples of images labeled as "Don´t know".



Biomedical **Vector graphics** **Document**

Figure 05. Examples of "Don´t Know" images

Table 01 shows the numbers of images per file format in the test set.

	JPG	TIFF	BMP	Total
Photo	7,476	35	457	**7,968**
Logo	2,984	0	67	**3,051**
Doc	3,048	808	0	**3,856**
Don´t know	202	0	327	**529**
Total	**13,710**	**843**	**851**	**15,404**

Table 01 – Images per file formats

2.1 Features Tested

The choice of the features to be extracted and tested is the key to the success and performance of the classification. Image entropy is often used as the key for classification [7]. It has a large computational cost, however. Entropy calculation demands a scan in the image to calculate the relative frequency of a given color, for instance, which is than multiplied for its logarithm and added up. The existing classifier is based on the binary classification approach originally described in [7]. It assumes a Gaussian distribution for each of the features, and its performance degrades in proportion to the non-Gaussian nature of the data.

This work assumed that decreasing the gamut of an image, analyzed together with its grey scale and monochromatic equivalents would provide enough elements for a fast and efficient image classification. The features tested are:

- Palette (true-color/grayscale)
- Gamut
- Conversion into Grayscale (if RGB)
- Gamut in Grayscale (if RGB)
- Conversion into Binary (Otsu)
- Number of black pixels in binary image.
- (#Black_pixels/Total_#_pixels)*100%
- (Gamut/Palette)*100% (true-color/grayscale)

Image binarization is performed by using Otsu [11] algorithm. The data above are extracted for each image and placed in a vector of features.

2.2 Training and test sets

Table 02 summarizes some of the features of the images in the test set. The height and width stand for the number of pixels in the image. RGB size stands for the true color size of the image (if a color image). 8-bits size is either the size of the original image if in gray scale or the size of the grey-scale converted from true-color. #B_pixels stands for the number of black pixels in the monochromatic converted image.

Photo	**Average**	**Median**	**Variance**	**Deviation**
height	1138	1104	387968	622
width	1274	1132	548014	740
RGB size	1.98MB	1.49MB	460646	214627
8-bits size	667KB	578KB	675668	7548028
gamut RGB	13641	9956	171547287	5884
gamut gray	231	247	1288	0,707
#B_pixels	1373240	755150	23511	153332
Logo	**Average**	**Median**	**Variance**	**Deviation**
height	232	180	19324	139
width	253	221	13066	114
RGB size	15KB	7KB	45800	214010
8-bits size	9KB	5KB	50658	609718
gamut RGB	5324	3848	37149265	6095
gamut gray	230	241	1102	33
#B_pixels	38785	6579	59580	244095
Doc	**Average**	**Median**	**Variance**	**Deviation**
height	1896	1734	325997	570
width	1437	1328	201513	448
RGB size	1.10MB	879KB	111962	334581
8-bits size	674KB	568KB	207357	143999
gamut RGB	2700	2097	16317414	4039
gamut gray	181	211	5502	74
#B_pixels	1310386	749770	157416	3967564
Don´t Know	**Average**	**Median**	**Variance**	**Deviation**
height	477	363	180173	424
width	574	490	208282	456
RGB size	1.55MB	1.33MB	129678	3387135
8-bits size	520KB	386KB	323578	172717
gamut RGB	3954	2934	207165	5514
gamut gray	217	198	959	28
#B_pixels	101896	54475	82345	3169822

Table 02 – Main features on the images in the test set

The training set was carefully selected to guarantee the diversity of the images in the test set, having in mind that quality matters more than size. Table 03 presents the relative size of the training and test sets.

	Test	Training	%
Photo	**7,968**	**668**	**8.34**
Logo	**3,051**	**412**	**10.22**
Doc	**3,856**	**276**	**4.70**
Don´t know	**529**	**0**	**0**
Total	**15,404**	**1,356**	**8.80**

Table 03 – Sizes of Training x Test sets

The Weka [8] classification strategy used was the Random Forests (number of trees equal to 10) [12].

2.3 Sub-sampling

The factors for the feature extractor to improve were:

1- The larger the file - the richer in data redundancy - thus if the redundant data are thrown away the efficiency both in time and classification.

2- The selection of points should not be random. It should somehow provide a "reduced" version of the original image (although in some cases it may be distorted by unequal scaling!).

Twenty different sub sampling strategies were evaluated on the images. Two are presented here. The first, designated the "simple" subsampling technique, consisted of splitting the image in blocks of 4x4 pixels and averaging their values. This sub sampler provided the best overall improvement in performance for a subsampling approach that did not involve a decision tree. The second, the cascaded subsampling strategy, consisted of removing more points from the larger image files and provided the best overall accuracy of any classification schema, while simultaneously significantly improving performance, as shown in the next section. The cascaded subsampler performs the following operations:

```
size = height*width

• If size ≤ 300,000  break;
• If 300,000< size ≤ 500,000:
     remove 1 line or 1 column (whatever the larger);
• If 500,000 < size ≤ 700,000:
     remove 1 line and 1 column;
• If 700,000 < size ≤ 900,000:
     remove 2 lines  and 1 column, (if height>width)
     remove 1 line and 2 columns, otherwise;
• If 900,000 < size  remove 2 lines and 2 columns;
```

Code for the "cascaded" sub-sampler

3. Results

This section presents the results of classification of the images in the test set comparing the current classifier [7] and the one proposed herein. Both classifiers had the same training and test sets. The results are divided into three groups: original, simple (averaging), and cascaded subsampling.

3.1 Original Data

The results of classification are presented by the confusion matrices obtained. Table 04 presents the results for the current classifier, while Table 05 shows the results obtained with the new classifier.

Current	Photo	Logo	Document	DK	A
Photo	7280	620	12	56	0.914
Logo	429	2104	96	422	0.690
Document	206	351	3299	0	0.856
Don´t Know	70	225	0	234	0.442

Table 04 – Confusion matrix of the **current** classifier with original images

The mean accuracy of the current classifier (A) for the Photo, Logo and Document images was 12638/14875 = 85.3%.

New	Photo	Logo	Document	DK	A
Photo	7554	363	14	37	0.948
Logo	282	2730	23	16	0.894
Document	277	266	3314	0	0.859
Don´t Know	151	309	17	52	0.098

Table 05 – Confusion matrix of the **new** classifier with original images

The mean accuracy (A) for the Photo, Logo and Document images was 12893/14875 = 86.3%, which shows that the proposed classifier is slightly better than the current one.

3.2 Simple Subsampler Performance

The results for the 4x4-pixel averaging subsampler are shown in Tables 06 and 07.

Current	Photo	Logo	Document	DK	A
Photo	5009	2209	586	164	0.628
Logo	1280	1005	115	651	0.329
Document	1245	376	2183	52	0.566
Don´t Know	101	154	15	259	0.489

Table 06 – Confusion matrix of the **current** classifier with subsampled images (simple)

The mean accuracy (A) for Photo, Logo and Document for the simple subsampler with the current classifier was 8197/14875=55.1%.

New	Photo	Logo	Document	DK	A
Photo	6674	1043	138	113	0.837
Logo	263	2751	20	17	0.901
Document	381	61	3414	0	0.885
Don´t Know	124	330	26	49	0.092

Table 07 – Confusion matrix of the **new** classifier with subsampled images (simple)

The mean accuracy (A) for the Photo, Logo and Document images was 12839/14875 = 86.3%.

The simple subsampler performed as well as the original version of the new classifier, but drastically degraded the accuracy of the current one. The simple subsampler halved the time for feature extraction of the new classifier as is shown in section 4.

3.3 Cascaded Subsampler Performance

The results for the cascaded subsampler are found in tables 08 and 09.

Current	Photo	Logo	Document	DK	A
Photo	7603	275	18	72	0.954
Logo	385	1929	135	602	0.632
Document	311	373	3167	5	0.821
DK	77	174	128	150	0.283

Table 08 – Confusion matrix of the **current** classifier with **cascaded** subsampled images

The mean accuracy (A) for the Photo, Logo and Document images for the cascaded subsampler using the current classifier is 12699/14875 = 85.3%

New	Photo	Logo	Document	DK	A
Photo	7740	164	34	30	0.971
Logo	258	2761	11	21	0.904
Document	93	41	3722	0	0.965
DK	110	300	30	89	0.343

Table 09 – Confusion matrix of the **new** classifier with **cascaded** subsampled images

The mean accuracy (A) for the Photo, Logo and Document images was 14223/14875 = 95.6%. As one may observe the cascaded subsampler largely improved the performance of the new classifier and has a positive effect in performance, as shown below.

4. Time Performance

Table 10 presents the feature extraction and classification times together with information about the language those procedures were implemented into. Besides classification accuracy per cluster, the average feature extraction and classification times are presented. The entries with an "S" superscript denote the "simple" subsampler, while de "C" superscript stand for the "cascaded" subsampler. One should also remark that there is a difference in time scale between feature extraction and classification.

	Feature extraction		Classification	
	Time (s)	Language	Time (ms)	Language
Current	**1.4576**	**C#**	**6.13**	**C#**
New	**0.4174**	**C++**	**0.12**	**C#**
Current^{S}	**0.719**	**C#**	**6.13**	**Java**
New^{S}	**0.199**	**C++**	**0.12**	**C#**
Current^{C}	**0.497**	**C#**	**6.13**	**Java**
New^{C}	**0.1470**	**C++**	**0.12**	**C#**

Table 10 – Feature extraction and classification times

5. Discussion and Conclusions

Weka has shown to be an excellent test bed for statistical analysis. The choice for a Random tree classifier was made after performing several experiments with the large number of alternatives offered by Weka, although results did not vary widely. Amongst them a preliminary comparison between the new statistical classifier proposed here and a MLP neural classifier provided worse results (Photos 91.37%, Logos 85.48%, and Documents 94.54%).

The choice of the images in the training set is of paramount importance to the performance of the classifier. Quality has proved more important than size. For some reason not fully understood, the current classifier seems to be more sensitive to the quality of the training set than the one proposed herein. Enlarging the training set with incorrectly recognized data has proved efficient, but should be used with parsimony. Very small images seem to pose a higher degree of difficulty for classification, as they were more often misclassified.

The test set used here attempted to be representative of the universe of images of interest and incorporated two other test sets developed by Steven Simske and Mark Shaw that proved effective in the tuning of the current classifier. Every effort was made in the correct labeling of images and to avoid image duplication.

The new classification scheme provided here decreased the error rate by a factor of 3.2 (from 14.6% to 4.4%) while simultaneously improving performance by a factor of ten (from 1458 to 147 msec/image processing time) compared to the current scheme based on [7]. The increased accuracy improves the appearance of the printed output while the greatly improved performance frees up computing resources for additional printing tasks.

5. References

[1] H.Frigui and R.Krishnapuram. Clustering by competitive agglomeration. *P. Recognition*, 30(7), 2001.

[2] M.A.Hearst and J.O.Pedersen. Reexamining the Cluster Hypotesis: Scatter Gathet on Retrieval Results, SIGIR, 1996.

[3] S.Krishnamachari and M.Abdel-Mottaleb. Image Browsing using Hierarchical Clustering, IEEE Symposium on Computers and Communications, ISCC'99, July 99.

[4] P.Scheunders. Comparison of Clustering Algorithms Applied to Color Image Quantization, Patt. Recog. Letters, v18(11-13):1379-1384, 1997.

[5] G.Park, Y.Baek and L.Heung-Kyu. A Ranking Algorithm Using Dynamic Clustering for Content-Based Image Retrieval. CIVR'2002, pp.328—337, LNCS 2383, Springer Verlag, 2002.

[6] K.Barnard and D.Forsyth. Learning the Semantics of Words and Pictures, Inter. Conf. C. Vision, 2001.

[7] S.J. Simske, "Low-resolution photo/drawing classification: metrics, method and archiving optimization," *Proceedings IEEE ICIP*, IEEE, Genoa, Italy, pp. 534-537, 2005.

[8] Weka 3: Data Mining Software in Java, website http://www.cs.waikato.ac.nz/ml/weka/.

[9] G.Pereira e Silva and R.D.Lins. PhotoDoc: A Toolbox for Processing Document Images Acquired Using Portable Digital Cameras. CBDAR'2007, pp.107-114, 2007.

[10] Lins, R.D. and D.S.A. Machado, A Comparative Study of File Formats for Image Storage and Trans., v13(1):175-183, Journal of Electronic Imaging, 2004.

[11] N. Otsu. "A threshold selection method from gray level histograms". IEEETrans.Syst.Man Cybern. v(9):62-66, 1979.

[12] L. Breiman, "Random Forests", Machine Learning, 45(1), pp. 5-32, 2001.

An OCR Assessment of the Quality of Document Images Acquired with Portable Digital Cameras

Rafael Dueire Lins, Brenno Miro, Gabriel Pereira e Silva
Departamento de Eletrônica e Sistemas – Universidade Federal de Pernambuco - Brazil
rdl@ufpe.br, gfps@cin.ufpe.br

Abstract

This article analyses the quality of documents acquired with portable digital cameras for Optical Character Recognition. The results obtained are compared with same documents after border removal, perspective and skew correction and their scanned equivalent with different resolutions and saved into distinct file formats.

1. Motivation

Students and professionals of many different areas now use portable digital cameras for digitalizing documents, taking advantage of their low weight, portability, low cost, small dimensions, etc. This new research area [1][2] is evolving fast in many different directions and claims for new algorithms, tools and processing environments that are able to provide users in general with simple ways of visualizing, printing, transcribing, compressing, storing and transmitting through networks such images. Reference [3] points out some particular problems that arise in this document digitalization process: the first of all is background removal. Very often the document photograph goes beyond the document size and incorporates parts of the area that served as mechanical support for taking the photo of the document. The second problem is due to the skew often found in the image in relation to the photograph axes, as documents have no fixed mechanical support very often there is some degree of inclination in the document image. The third problem is non-frontal perspective, due to the same reasons that give rise to skew. A fourth problem is caused by the distortion of the lens of the camera. This means that the perspective distortion is not a straight line but a convex line, depending on the quality of the lens and the relative position of the camera and the document. The fifth difficulty in processing document images acquired with portable cameras is due to non-uniform illumination. This paper focuses on assessing the output of a commercially OCR (Optical Character Recognition) software for such documents. The results obtained are compared with the results obtained of processing the same batch of documents with PhotoDoc [4] a freely available software environment for processing document images acquired with portable cameras. The results of unprocessed and PhotoDoc Processed camera images are compared with the transcription obtained for the scanned version of the same documents. This work besides updating the results presented in [5] to more modern camera models of current use today, it applies a much better assessment methodology.

2. The assessment methodology

Assessing image quality in general is a complex subjective task. A quantitative assessment that avoids such subjectivity is of great importance. Similarly to the experiments reported in [5], in this paper the assessment methodology was limited to analyze the performance of commercial OCR tools. ABBYY FineReader Professional Edition 9 [6] was used, because it is possibly the best general purpose tool available today, able to process formatted texts.

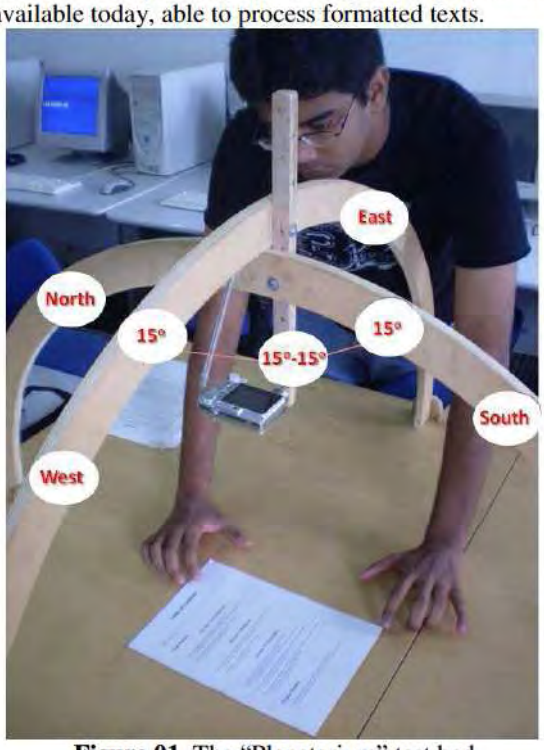


Figure 01. The "Planetarium" test bed

The "Planetarium" test bed shown in Figure 01 allows a controlled way to measure the angles and height of the camera to verify in extreme cases the effects of the perspective into the document transcription. These results are later used to assess the gains obtained with the documents after each processing step. On its turn, analyzing the results of OCRs is far from being a trivial task. The methodology presented in reference [7] which takes into account the nature of the errors in transcription was adopted here.

The errors were classified according to:

1. Character replacement.
2. Missing characters.
3. Character insertion.
4. Punctuation errors.

3. Test images features

The 168 pages of the proceedings of CBDAR 2007 were used as test document images for this work. Several pages include photographs, graphs, tables and other illustrations. They are printed in black in opaque white paper, where negligible back-to-front [8] interference was observed.


PhotoDoc: A Toolbox for Processing Document Images Acquired Using Portable Digital Cameras

Abstract

1. Introduction

Figure 02. Scanned page of the Proceedings of CBDAR´07.

Table I presents the results of the total of errors found in the OCR transcription of all document images digitized with a Ricoh Affício 1075 flatbed scanner in 100, 200 and 300 dpi saved into four different file formats: bmp (uncompressed), jpg (1% losses), png (lossless), and tiff (uncompressed), using the software provided by the scanner manufacturer. Figure 02 shows an example of the test images used in this work.

TABLE I
CHARACTER ERRORS FOUND IN SCANNED DOCUMENT IMAGES

100 DPI	BMP	JPG	PNG	TIF
replacement	55590	63113	63646	66894
punctuation	3851	4174	4057	4977
missing	7907	8095	8852	8454
insertion	96078	95729	96126	96126
SIZE	498MB	21.6MB	75.2MB	498MB
200 DPI	**BMP**	**JPG**	**PNG**	**TIF**
replacement	49837	41194	39912	43270
punctuation	2575	2787	2591	2738
missing	4274	5565	5259	5345
insertion	71467	79298	81402	79904
SIZE	1.95GB	65.1MB	222MB	1.95GB
300 DPI	**BMP**	**JPG**	**PNG**	**TIF**
replacement	38529	58681	58742	58647
punctuation	2398	3376	3429	3397
missing	4478	5883	5940	6447
insertion	81909	91877	93803	92045
SIZE	4.42GB	118MB	395MB	4.42GB

Analyzing the data in Table I one may observe that different file formats yield varying error rate even amongst lossless ones (bmp, png, and tiff). One may possibly say that the best trade-off between space and OCR correct recognition rate is reached in the 200 dpi scanning saved in PNG. It is interesting to note that a higher resolution tended to drastically increase the insertion noise.

4. Unprocessed Image Transcription

The camera used in this work is a Sony Cyber-shot 7.2 Mega Pixels model DSC-W55, with lenses Carl-Zeiss Vario-Tessar 2.8-5.2/6.3-18,9. Figure 03 exemplifies the test image obtained in the "Planetarium" test bed. Documents were obtained in true-color, in 7.0 and 5.0 Mpixels, with and without the inbuilt camera strobe flash. The camera was set into "auto-focus" mode, i.e. the user leaves to the device the automatic setting of the focus. This is consistent with the expected knowledge of the end user. In the case of documents acquired with portable digital cameras with no mechanical support very often images have perspective distortion. Skewed images are unpleasant for human visualization, introduce extra difficulty in text reading, claim extra space for storage, degrade OCR performance, etc. this problem arises in almost all documents. This work used two strategies to measure

perspective degradation the first one was using a device developed by the authors in which the position of the camera was set and in the second strategy the photo was taken "free hand", without mechanical support. Figure 03 shows a document photo taken with the Planetarium.

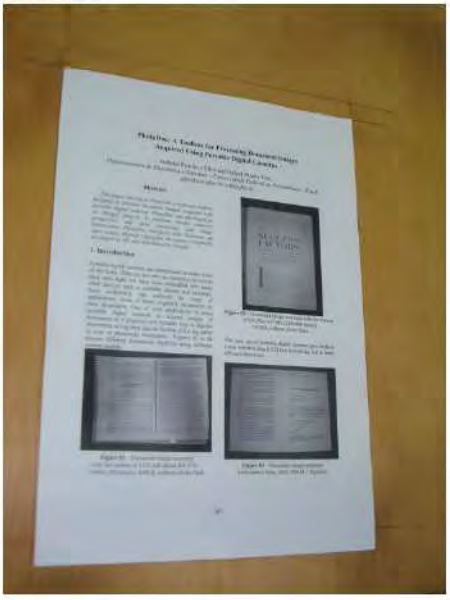


Figure 03. Photo taken in the Planetarium with 15°S-15°W.

The photos taken in the Planetarium are surrounded by part of the board in wood that serves as mechanical support to it. The photos were taken indoors with artificial illumination provided by fluorescent lights sufficiently high-up to light the document surface evenly.

Table II presents the results of the transcription of the documents under different inclinations and heights, with and without strobe flash. One may observe that the position of the camera in relation to the document causes a wide variation in results. As one expects, the greater the angle in relation to the frontal plane of the document, the larger the OCR transcription degradation. In general, the images acquired with 7.2 Mpixels yielded better results than with 5.0 Mpixels, but this was not the case when the perspective distortion happened in the West and South directions simultaneously (15°E – 15° S).

At the height of 37 cm parallel with the plane (0°E – 0°S) the inbuilt strobe flash of the camera yielded an uneven illumination not perceptible visually, but that degraded OCR response both at 5.0 and 7.2 Mpixels.

At the height of 45 cm the contrary phenomenon was observed and the flash yielded better OCR transcription results. When no mechanical support was used (Free-hand) the use of the strobe flash in 7.2 Mpixels provided the best results. These results are close to the ones obtained by using the height of 37 cm without any inclination angle. The experiments performed in reference [5] report a measure of the skew angle of the bottom line of the 50 documents analyzed was around 2° in each direction. That value was also observed in the experiments performed herein. One should remark, however that users in general tend to be less careful and tend to take photos with a higher perspective distortion. Experiments performed with several people that meet that profile showed that the perspective distortion does not exceed 10 degrees in each direction simultaneously and that the document image is seldom chopped off.

TABLE II
CHARACTER ERRORS FOUND IN CAMERA DOCUMENT IMAGES

	5.0 Mpixels		7.2 Mpixels	
15° South	**+ flash**	**no flash**	**+ flash**	**no flash**
replacement	62349	58258	59900	62466
punctuation	3787	3694	3665	3769
missing	6381	6477	6010	6658
insertion	94282	93440	95943	94149
15° West	**+ flash**	**no flash**	**+ flash**	**no flash**
replacement	70138	72342	68136	69308
punctuation	5463	5494	5321	5763
missing	7685	6254	6223	6365
insertion	80154	85100	80498	80375
30° South	**+ flash**	**no flash**	**+ flash**	**no flash**
replacement	64499	68242	62004	57720
punctuation	3850	4513	3846	3935
missing	6154	6885	5963	6683
insertion	93420	88841	92783	86607
30° West	**+ flash**	**no flash**	**+ flash**	**no flash**
replacement	85854	89840	88910	85113
punctuation	8389	8655	8948	8819
missing	8429	8669	8041	9716
insertion	50630	53798	47855	49014
15° W - 15° S	**+ flash**	**no flash**	**+ flash**	**no flash**
replacement	82182	84307	86141	83036
punctuation	6781	7010	7210	7098
missing	8553	9455	9044	9167
insertion	69204	72238	71451	71213
height= 45 cm	**+ flash**	**no flash**	**+ flash**	**no flash**
replacement	60454	62047	63674	60961
punctuation	3651	3829	3706	3673
missing	7352	7486	7133	7973
insertion	93319	92799	94593	94748
height= 37 cm	**+ flash**	**no flash**	**+ flash**	**no flash**
replacement	71954	70470	63563	62380
punctuation	5537	5785	3552	3748
missing	8405	8045	7284	7253
insertion	79699	80515	95144	95153
Free hand	**+ flash**	**no flash**	**+ flash**	**no flash**
replacement	58891	62866	54996	63867
punctuation	5603	3911	5073	4304
missing	10033	7445	10474	9715
insertion	82741	93774	90232	90271

5. PhotoDoc Processing

As mentioned in the introduction of this paper, the motivation for research reported herein is to meet the needs of ordinary people such as students and professionals that acquire document images using portable digital cameras. As shown in Figure 03 such documents are framed by the place that serve as mechanical support the photo to be taken and as a perspective distortion. PhotoDoc is a simple processing environment developed with the aim to help those non-expert users that digitize documents with portable digital cameras. A brief overview of PhotoDoc processing capabilities are shown here as it was used to process all the documents that are also OCR transcribed. Experiments in PhotoDoc showed that lens distortion has negligible effects if compared with border removal and perspective and skew correction. Thus, it is ignored.

5.1 Border Removal

PhotoDoc performs automatic background border removal of images of documents obtained with portable digital cameras imposing as few restrictions as possible, because users tend to acquire those document images in non-ideal conditions of, illumination of the surface the document is placed on for digitalisation, perspective camera-document, etc.

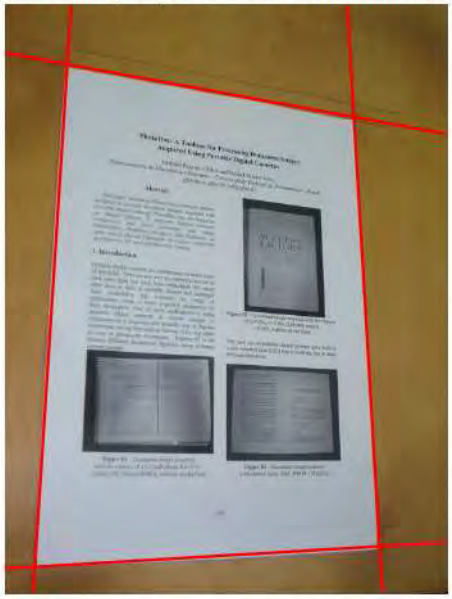


Figure 04. Image from Figure 03 showing perspective correction reference points and edges.

As may be observed in Figure 03, the document image is surrounded by a wooden background area of no value in terms of information. This area not only drops the quality of the resulting image for CRT screen visualization, but also consumes space for storage and large amounts of toner for printing, alters the segmentation algorithm of the OCR and thus affects the response obtained in the number of characters and words correctly transcribed, as shown later on in this paper. Several papers in the literature address this problem in different applications [9, 10, 11, 12]. Removing such frame manually is not practical due to the need of a specialized user and time consumed in the operation. The algorithm presented in reference [11] is used in PhotoDoc to automatically remove such border as an OCR pre-processing stage. It assumes that the background may be of any kind of colour or texture, provided that there is a colour difference of at least 32 levels between the image background and at least one of the RGB components of the most frequent colour of the document background (paper).

5.2 Perspective and skew correction

The freedom allowed in acquiring document images with portable digital cameras without mechanical support invariably leads to perspective distortion. Several algorithms in the literature address this problem [3, 14, 15, 16]. The correction of perspective distortion has border detection as a first step to find the polygon that margins the image and getting the four corner points that will serve as reference for the linear transformation. The image of the four corner points serve to crop the perspective corrected image and automatically performs skew correction. On the other hand, perspective distortion opens a number of alternatives which cause different effects in the quality of the image produced both in terms of visualization and OCR response. In general, the skew angle was small (less than 2^{o}), thus this means that the image tends to exhibit a trapezoidal shape. Two alternatives for correction arise: either to narrow the opening edges or to widen the closing edges. The latter alternative was discarded in PhotoDoc because the general trend is to disconnect contiguous areas, which has a serious degrading effect on OCR response. The interpolation methods applied in PhotoDoc is closest neighbor. The image obtained after perspective correction and cropping closely resembles the scanned one.

Table III presents the OCR response for documents, after PhotoDoc processing. Comparing the results obtained in Tables I, II, and III one may observe that Photodoc processing largely improved the OCR recognition rate of documents, yielding better OCR response than images scanned in 100 dpi resolution, in general. The only exceptions are in the case of very strong perspective distortion 30^{o} South and in the case of insertion errors when photos were taken at a height

of 37 cm both with and without the strobe flash. Insertion errors are harder to be corrected with the help of dictionaries than the other errors.

TABLE III
CHARACTER ERRORS FOUND IN DOCUMENT IMAGES AFTER PHOTODOC PROCESSING

	5.0 Mpixels		7.2 Mpixels	
15° South	+ flash	no flash	+ flash	no flash
replacement	51435	59671	43163	41965
punctuation	2075	2964	2050	2948
missing	5085	5854	5927	5072
insertion	79835	76183	75844	80559
15° West	+ flash	no flash	+ flash	no flash
replacement	53025	51241	41716	42365
punctuation	2169	2432	2881	2880
missing	5744	5550	4713	4875
insertion	77884	75605	79517	78545
30° South	+ flash	no flash	+ flash	no flash
replacement	69041	76198	62129	62678
punctuation	3173	2953	3128	3245
missing	6039	6889	5882	5544
insertion	77543	78036	78078	71426
30° West	+ flash	no flash	+ flash	no flash
replacement	63709	62699	63509	65477
punctuation	3118	4096	4390	4499
missing	6820	7190	7728	8941
insertion	82663	92792	94419	93170
15° -15° S	+ flash	no flash	+ flash	no flash
replacement	63545	62588	62506	61783
punctuation	4397	4017	3888	3949
missing	7939	7118	7604	7835
insertion	95203	92656	96927	97765
height= 45 cm	+ flash	no flash	+ flash	no flash
replacement	50412	59689	40127	43567
punctuation	2048	2071	1862	1898
missing	7588	7386	7232	7795
insertion	71430	77838	84096	81263
height= 37 cm	+ flash	no flash	+ flash	no flash
replacement	52212	53935	41483	43019
punctuation	1671	1794	1559	1755
missing	7047	7617	6507	6583
insertion	95435	97526	97942	97685
Free hand	+ flash	no flash	+ flash	no flash
replacement	54269	60584	56880	71226
punctuation	4589	4823	5538	5597
missing	6225	6872	6946	5795
insertion	77711	84598	70237	92331

5. Conclusions

This paper provides a comparative analysis of the quality of documents acquired through 5.0 and 7.2 Mpixels Sony portable digital camera in comparison with their scanned version with three different resolutions (100, 200 and 300 dpi). A batch of 168 documents was studied totaling 479,154 characters. The quantitative analysis performed herein allows to conclude that portable digital cameras not only provide a simple way to digitalize documents to be read by humans, but the quality of documents allows means for image-to-text transcription using commercial OCRs. The OCR performance improves if the document is processed in an image processing environment such as PhotoDoc that removes the borders introduced during document photographing and is perspective and skew corrected.

Several challenges are faced to improve OCR performance. Illumination and compensating the effect of the embedded strobe flash are two of the most important ones as they pose difficulties to image binarization.

7. References

[1] D.Doermann, J.Liang, H. Li, "Progress in C.-Based Document Image Analysis," ICDAR'03, Vol(1): 606, 2003.

[2] J. Liang, D. Doermann and H. Li. Camera-Based Analysis of Text and Documents: A Survey. International Journal on Document Analysis and Recognition, 2005.

[3] R.D.Lins, A.R.Gomes e Silva and G.Pereira e Silva, Enhancing Document Images Acquired Using Portable Digital Cameras, ICIAR´07, LNCS, Springer-Verlag, 2007.

[4] G.Pereira e Silva and R.D.Lins. PhotoDoc: A Toolbox for Processing Document Images Acquired Using Portable Digital Cameras. CBDAR´2007, pp.107-114, 2007.

[5] R.D.Lins, *et al.* Assessing and Improving the Quality of Document Images Acquired with Portable Digital Cameras, ICDAR´2007, pp.569-573, IEEE Press, 2007.

[6] ABBYY FineReader Professional Edition 9 - http://www.abbyy.com/

[7] R.D.Lins and N.F.Alves. A New Technique for Assessing the Performance of OCRs. IADIS – Int. Conf. on Comp. Applications, IADIS Press, v. 1, p. 51-56, 2005.

[8] J. M. M. da Silva *et al.* Binarizing and Filtering Historical Documents with Back-to-Front Interference, ACM-SAC 2006, Nancy, April 2006.

[9] K.C.Fan, Y.K.Wang, T.R.Lay, Marginal noise removal of document images, Patt.Recognition. 35, 2593-2611, 2002.

[10]Lu S and C L Tan, Camera document restoration for OCR, CBDAR 2005/ICDAR 2005, Seoul, Korea.

[11]R. Gomes e Silva and R. D.Lins. Background Removal of Document Images Acquired Using Portable Digital Cameras. LNCS 3656, p.278-285, 2005.

[12]H.S.Baird, Document image defect models and their uses, ICDAR'93, Japan, IEEE Comp. Soc., pp. 62-67, 1993.

[13]L.G.Shapiro and G.C.Stockman, Computer Vision, March 2000. http://www.cse.msu.edu/~stockman/Book/book.html.

[14]L. Jagannathan and C. V. Jawahar, "Perspective correction methods for camera based document analysis," pp. 148–154, CBDAR 2005, Seoul, Korea. 2005.

[15]P. Clark, M. Mirmehdi, "Recognizing Text in Real Scenes", IJDAR, Vol. 4, No. 4, pp. 243-257, 2002.

[16]Clark, M. Mirmehdi, "On the Recovery of Oriented Docs. from Single Images", CSTR-01-004, Bristol, 2001.

Melhorando A Qualidade de Documentos Coloridos com Interferência Frente-Verso

J. M. M. da Silva, Rafael Dueire Lins e G. F. P. e Silva

Resumo*— *A interferência frente-verso ocorre em documentos escritos (ou impressos) em ambos os lados de papel translúcido. Tal interferência, dificulta sua transcrição automática e binarização. Este artigo apresenta uma nova técnica de filtragem de documentos coloridos com interferência frente-verso que objetiva a melhoraria da legibilidade do mesmo.

Palavras-Chave*— *Interferência frente-verso, documentos históricos, segmentação, interpolação.

Abstract*— *The back-to-front interference occurs on documents written (or printed) on both sides of a translucent paper. Such interference makes more difficult their transcription and binarization. This paper presents a new technique to filter out such interference in color documents, enhancing readability.

Keywords*— *Back-to-front interference, bleeding, show-through, historical documents, segmentation, interpolation.

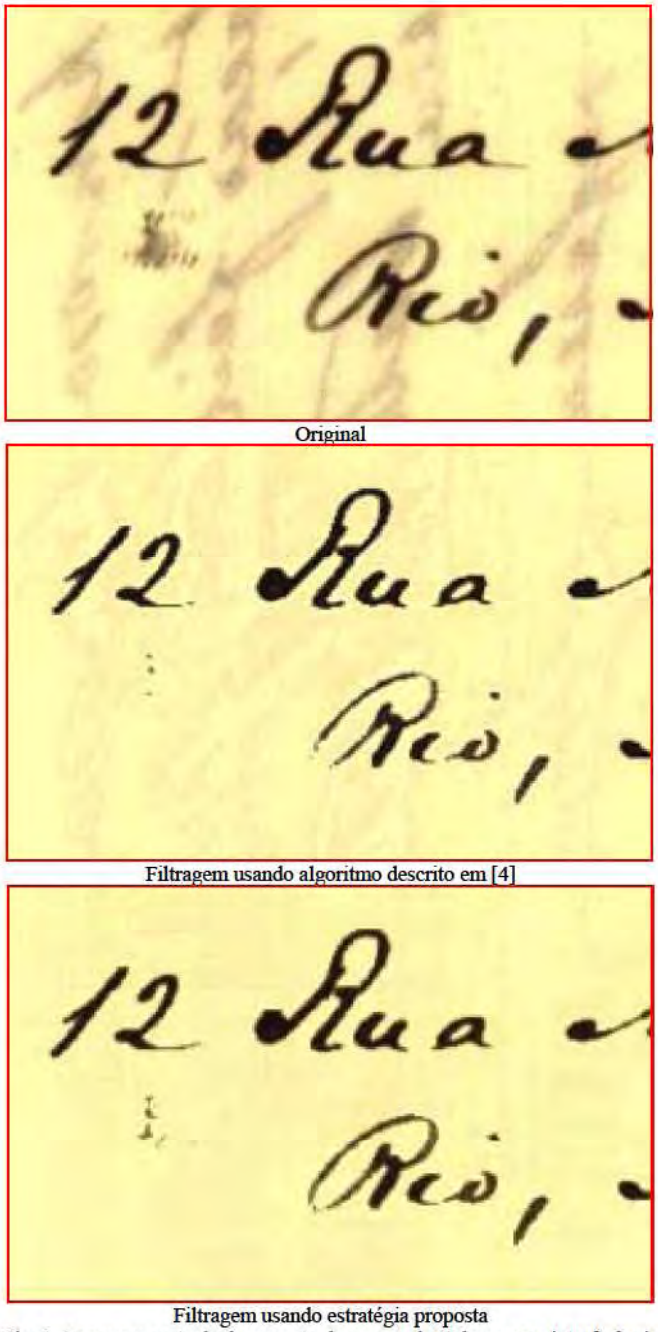
I. Introdução

No início dos anos 90, processou-se a digitalização do acervo de correspondências de Joaquim Nabuco, através de trabalho conjunto realizado entre a Fundação Joaquim Nabuco e a Universidade Federal de Pernambuco [1]. Do rico acervo de aproximadamente 6.500 cartas, cerca de 10% das imagens dos documentos digitalizados apresentavam uma característica não anteriormente descrita na literatura e que passou a ser conhecida como interferência frente-verso (do inglês *back-to-front interference*) [1]. Posteriormente, outros autores utilizaram os nomes de *bleeding* [2] e *show-through* [3] para este mesmo efeito.

A interferência frente-verso ocorre quando o conteúdo da face do verso de um documento se faz presente na face frontal. Para que tal interferência apareça em um documento é necessário que este seja escrito (ou impresso) em ambos os lados de papel translúcido (vide primeira imagem da Figura 1). Essa interferência em documentos degrada seus processos de transcrição automática e binarização. Em documentos históricos, o envelhecimento do papel é mais um fator de dificuldade, pois o seu escurecimento diminui o "grau de separação" entre a tinta de cada um dos lados e o papel.

Este artigo apresenta uma nova estratégia de filtragem da interferência frente-verso em imagens de documentos coloridos, tendo melhores resultados do que os apresentados em [4]. Nas duas estratégias, a idéia é discriminar a área interferente e realizar um preenchimento sobre a mesma.

João Marcelo Monte da Silva, Rafael Dueire Lins e Gabriel de França Pereira e Silva, Departamento de Eletrônica e Sistemas, Centro de Tecnologia e Geociências, Universidade Federal de Pernambuco, Recife, Brasil, E-mails: joao.mmsilva@ufpe.br, rdl@ufpe.br, gfps@cin.ufpe.br.



Original

Filtragem usando algoritmo descrito em [4]

Filtragem usando estratégia proposta

Fig. 1. Zoom em parte de documento do acervo de Nabuco com interferência frente-verso.

A principal diferença entre a nova estratégia e a descrita em [4] está no processo de preenchimento da área interferente. A anterior preenche a área destacada aleatoriamente, utilizando *pixels* previamente classificados

como papel, enquanto a nova realiza uma interpolação "linear". Para que se possa ter uma idéia prévia da melhoria implantada, é mostrado na Figura 1 um mesmo trecho ampliado das imagens de um documento. Percebe-se que o preenchimento promovido pela nova estratégia (terceira imagem) possui um aspecto mais natural. Além disso, o "contorno da interferência" remanescente no resultado da filtragem usando o algoritmo anterior (vide segunda imagem), não mais aparece no da nova.

Na seção II, é feita a descrição desse novo sistema de filtragem. Os resultados e as análises estão apresentados na seção III. Finalmente, na seção IV, são apresentadas as conclusões e linhas para trabalhos futuros.

II. Sistema de Filtragem

Esta seção apresenta a nova estratégia para remover a interferência frente-verso de imagens de documentos coloridos, melhorando os resultados apresentados em [4].

A idéia básica é discriminar a área correspondente à interferência frente-verso (primeira etapa) e preenchê-la com *pixels* cujas cores mais se assemelham às cores do papel (segunda etapa). A melhoria proposta aqui, se dá nas duas etapas mencionadas, as quais serão tratadas separadamente.

A. Discriminação dos Pixels da Interferência

Para se encontrar a área interferente, o algoritmo de segmentação Silva-Lins-Rocha [5] é utilizado duas vezes: a primeira, para separar o texto do resto do documento; e a segunda, para destacar a interferência do papel. Em linhas gerais, as características das distribuições do texto e da interferência são distintas, sendo a da segunda mais dispersa.

O *fator de perda* (α) é um parâmetro do algoritmo de segmentação utilizado que deve garantir um melhor ajuste estatístico entre as distribuições das imagens original e binarizada, baseado na entropia de Shannon [6]. Para a segunda aplicação, propõe-se uma pequena alteração nesse fator, que passa a ser constante (α=1), garantindo uma melhor separação entre a interferência e o papel.

Em suma, para detectar a área interferente através desta estratégia:

1. aplica-se o algoritmo de segmentação [5] para separar a tinta da frente do resto do documento (vide Figuras 2a e 2b); e
2. aplica-se novamente o algoritmo, agora com a alteração do *fator de perda* acoplada, para separar a tinta interferente do papel (vide Figuras 2c e 2d).

Agora, têm-se identificados os *pixels* interferentes.

Para uma melhor visualização de como é feita a segmentação, a Figura 3 apresenta o histograma da versão em níveis de cinza da imagem de um documento com interferência frente-verso. O primeiro limiar T_L é obtido na primeira aplicação do algoritmo e o segundo T_H a partir da segunda. Os *pixels* cujo valor de nível de cinza é inferior a T_L são classificados como tinta da face frontal, os superiores a T_H são ditos pertencer ao papel e os maiores que T_L e menores que T_H são discriminados como interferência.

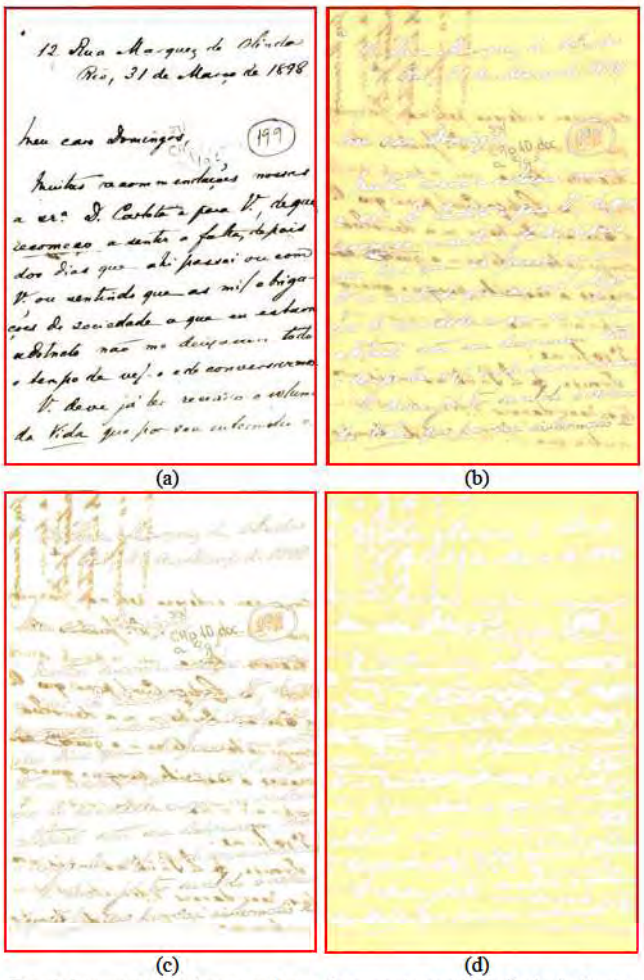


Fig. 2. Segmentos da imagem de um documento com interferência frente-verso: (a) tinta da frente e (b) papel com interferência. Segmentos da imagem da Figura 2b: (c) interferência e (d) papel.

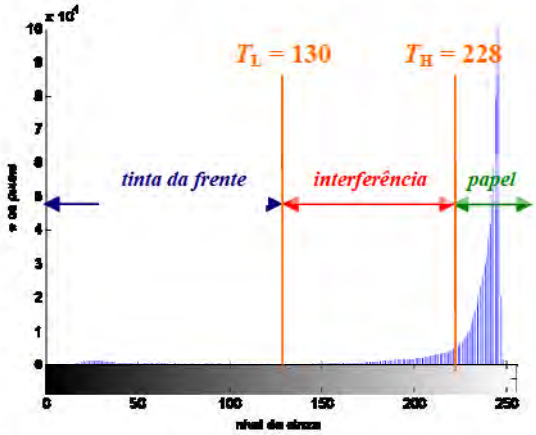
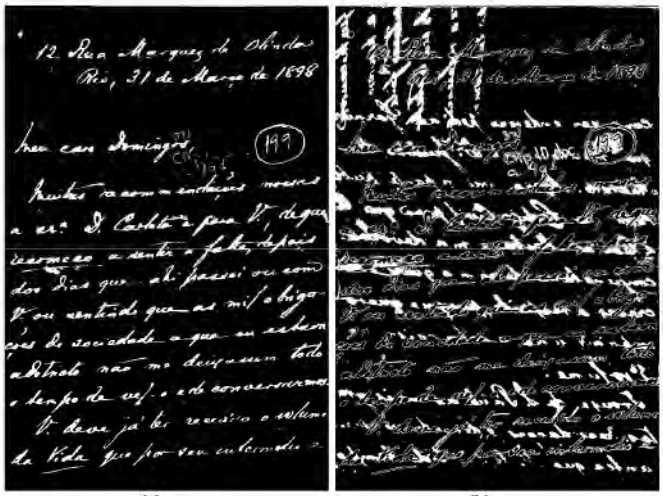


Fig. 3. Histograma da imagem de um documento com interferência frente-verso – detalhes da segmentação.

B. Preenchimento da Área Interferente

O preenchimento da área interferente aqui proposto utiliza uma interpolação "linear" para preencher a área interferente, diferentemente do algoritmo apresentado em [4] que preenche tal área com *pixels* pertencentes ao papel escolhidos aleatoriamente.



(a) (b)

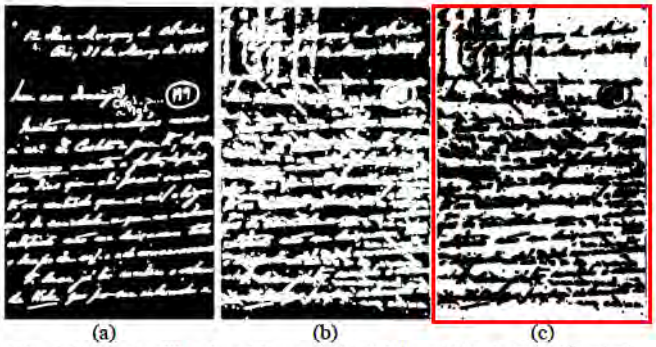
Fig. 4. Máscaras que identificam (a) o texto e (b) a interferência.

O novo processo define duas máscaras binárias: TEXTO e INTERF. A primeira identifica os *pixels* provenientes do texto da frente (vide Figura 4a), a segunda destaca a área interferente (vide Figura 4b). Pode-se supor que apenas a máscara INTERF seja suficiente para o processo de preenchimento, pois os *pixels* que se deseja substituir "já estão discriminados". Contudo, há o aparecimento de algumas dificuldades.

A idéia é substituir as cores dos *pixels* da interferência por cores o mais próximo possível do papel naquela região. Isto é conseguido através de uma interpolação, a qual faz uso das cores dos *pixels* que estão "no contorno" da área a ser interpolada. Se o texto estiver muito próximo (ou seja, na periferia) da área interferente, seus *pixels* participarão do processo de interpolação, o que trará, novamente, cores relativamente escuras para a área interferente, isso vai contra o objetivo de tornar tal área "o mais próximo possível do papel". Dessa forma, antes de interpolar deve-se também retirar o texto. Isso justifica a utilização da máscara TEXTO.

Outro problema que surge após a segmentação é a permanência (no "papel") do contorno do texto e da interferência (vide Figuras 2d e 4b). Este fato certamente torna o processo de interpolação ineficiente, pois as cores dos pixels interferentes seriam substituídas pelas presentes no contorno do texto e da interferência. Para solucionar esse problema, deve-se aplicar uma operação morfológica de dilatação nas máscaras, isso faz com que os contornos do texto e da interferência sejam corretamente classificados como "texto" e "interferência", respectivamente (vide Figuras 5a e 5b).

Como mencionado anteriormente os *pixels* que farão parte do processo de interpolação são os que contornam a área interferente e pertencem ao papel. Dessa forma, cria-se uma nova máscara que destaca os *pixels* pertencentes apenas ao papel. Tal máscara PAPEL pode ser adquirida pelo complemento da resultante da operação lógica OU entre as máscaras TEXTO e INTERF já dilatadas (vide Figura 5c).



(a) (b) (c)

Fig. 5. Máscaras dilatadas: (a) texto (T) e (b) interferência (I). (c) $\overline{\mathrm{T\,OR\,I}}$.

Os *pixels* que serão utilizados no processo de interpolação estão destacados na máscara PAPEL (Figura 5c); e os que serão interpolados são os que aparecem na máscara INTERF dilatada (Figura 5b), mas não estão presentes na máscara TEXTO (Figura 5a). Essa condição do último uso da máscara de TEXTO é imposta para que o processo de interpolação não altere os pixels classificados como texto. Se tal condição não fosse imposta, parte do texto seria "apagada".

Agora, será apresentado o processo de interpolação. Para um melhor entendimento, deve-se observar a imagem da Figura 6.

Sejam as coordenadas:

- (x_0,y_0) de um ponto P do intervalo a ser interpolado;
- (x_0,y_1) do ponto P_N – primeiro ponto ao norte de P;
- (x_0,y_2) do ponto P_S – primeiro ponto ao sul de P;
- (x_1,y_0) do ponto P_O – primeiro ponto à oeste de P;
- (x_2,y_0) do ponto P_L – primeiro ponto à leste de P.

Seja $i_C(x,y)$ a intensidade da componente C (R, G ou B) do *pixel* nas coordenadas (x,y). A intensidade do *pixel* (P) interpolado é dada por

$$i_C(x_0,y_0)=\frac{d_4\cdot i_1+d_3\cdot i_2+d_2\cdot i_3+d_1\cdot i_4}{d_4+d_3+d_2+d_1}, \qquad (1)$$

onde os valores d_k's e i_k's (k=1,...,4) representam as intensidades e as distâncias dos pontos definidos – P_N, P_S, P_O e P_L – ao ponto P, seguindo uma ordem crescente em relação às distâncias. Por exemplo, o ponto mais próximo de P tem distância d_1 e intensidade i_1, o segundo mais próximo tem distância d_2 e intensidade i_2, e assim por diante.

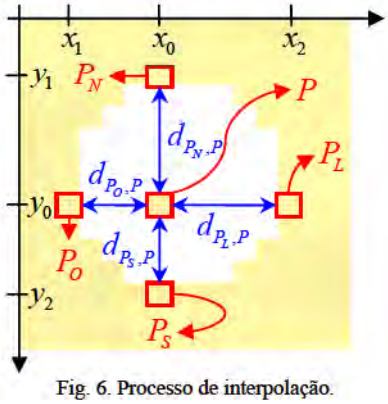
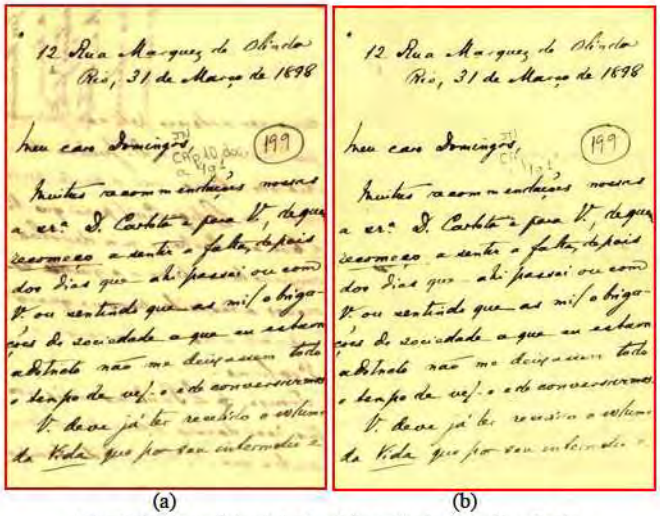


Fig. 6. Processo de interpolação.



(a) (b)

Fig. 7. Imagens (a) original e (b) resultado da aplicação da nova estratégia de filtragem proposta.

A distância entre dois pontos quaisquer P_a e P_b com coordenadas (x_a,y_a) e (x_b,y_b), respectivamente, é definida por

$$d_{Pa,Pb} = \sqrt{(x_a - x_b)^2 + (y_a - b_b)^2}. \quad (2)$$

A Equação 1, efetua uma média ponderada, onde a intensidade do *pixel* mais próximo do ponto P tem um peso maior. Isto é perfeitamente razoável visto que em uma "pequena" vizinhança, geralmente, quanto mais próximo um ponto está de outro, mais próximo são seus valores de intensidade. O resultado da aplicação desta estratégia de filtragem para a imagem da Figura 7a é apresentado na Figura 7b.

III. Análises e Resultados

O algoritmo proposto foi testado em 260 imagens do acervo de documentos digitalizados de Joaquim Nabuco [7] trazendo resultados de melhor qualidade que o algoritmo em [4].

Do total testado, três resultados são mostrados nas Figuras 8, 9 e 10. As imagens apresentadas representam um mesmo trecho ampliado das imagens originais, e resultantes da filtragem com os algoritmos: anterior e proposto. Pode-se constatar a superioridade da nova técnica de filtragem, pois a detecção da interferência e a qualidade do preenchimento foram melhoradas, isso implica dizer que as duas propostas de aperfeiçoamento atingiram seus objetivos, tornando o aspecto do documento mais natural.

Deve-se relatar que, semelhante ao que ocorreu com o algoritmo anterior, o aqui proposto não teve um desempenho tão bom nas imagens cuja interferência era muito dispersa, ou seja, muito "borrada" (vide Figura 10). Também cabe ressaltar que no acervo observado há poucos casos de imagens com essa característica.

A detecção efetiva de toda interferência se torna uma tarefa complexa, além disso, mesmo se detectando "quase toda interferência" (o que foi conseguido efetuando-se uma dilatação maior na máscara INTERF) a área para preenchimento é grande (pois a interferência é bastante dispersa). Com uma grande área a ser preenchida, a interpolação utilizada não traz um aspecto tão natural para a imagem final.

A Figura 11 ilustra o problema exposto no parágrafo anterior. A primeira imagem contém o mesmo trecho observado na Figura 10, entretanto ele corresponde à imagem filtrada através da aplicação da nova estratégia, fazendo uso da máscara INTERF com uma dilatação maior. Observando-se as Figuras 10 e 11a percebe-se que houve uma melhor filtragem no trecho em questão. No entanto, avaliando-se outro trecho da imagem (Figura 11b), observa-se que este não parece tão natural, dessa forma, constata-se o problema que surge quando se tenta interpolar uma área "relativamente grande".

Para tentar amenizar tal problema, pretende-se utilizar uma interpolação que leve em consideração não apenas os "quatro pontos vizinhos", mas sim todo um "intervalo vizinho", além disso, a consideração de estatísticas da distribuição dos pixels do papel também pode contribuir para um melhor preenchimento. É esperado que o uso dessa nova interpolação faça com que o documento tenha um aspecto mais natural após a filtragem.

IV. Conclusões e Trabalhos Futuros

Neste artigo é proposta uma nova estratégia de filtragem da interferência frente-verso em imagens de documentos coloridos, que apresenta melhores resultados que os algoritmos anteriormente descritos na literatura. Esse sistema utiliza o algoritmo de segmentação proposto em [5] para discriminar os *pixels* provenientes da interferência. Após a discriminação, a área interferente é interpolada de forma a tornar as cores dos seus *pixels* mais próximas das do papel.

Um dos passos que precede o processo de interpolação é a dilatação das máscaras TEXTO e INTERF. Para as 260 filtragens realizadas, foi utilizada uma mesma dilatação. Visto que há imagens com interferência mais (ou menos) borrada, seria mais eficiente o uso de uma dilatação específica para cada documento. Esta possível melhoria está sendo estudada, e se baseia na tentativa do dimensionamento do grau de dispersão da interferência. Tal grau pode ser obtido através da observação do gradiente entre a interferência e o papel.

Outro aspecto ainda não mencionado é o fato do aparecimento de componentes de altas freqüências na imagem resultante da filtragem. Isso ocorre devido à "quebra inercial" da variação das intensidades que havia na imagem original. Esse fato também contribui para um aspecto menos natural da imagem final. Para se ter uma melhoria neste sentido, pode-se verificar a freqüência máxima que aparece no documento original, e com esta, fazer uso de um filtro passa-baixas para filtrar a imagem final. Isso suavizará, por exemplo, transições entre áreas interpoladas e áreas de texto (não alteradas), trazendo um aspecto mais "natural" para o documento final.

AGRADECIMENTOS

Ao CNPq (Conselho Nacional de Pesquisas e Desenvolvimento Tecnológico) e à CAPES (Coordenação de Aperfeiçoamento de Pessoal de Nível Superior) pelo suporte financeiro. À FUNDAJ (Fundação Joaquim Nabuco) pela permissão de utilização das imagens.

REFERÊNCIAS

[1] R. D. Lins, et al. "An Environment for Processing Images of Historical Documents. Microprocessing & Microprogramming", pp. 111-121, North-Holland, 1994.

[2] R. Kasturi, L. O'Gorman and V. Govindaraju, "Document image analysis: A primer", *Sadhana*, (27):3-22, 2002.

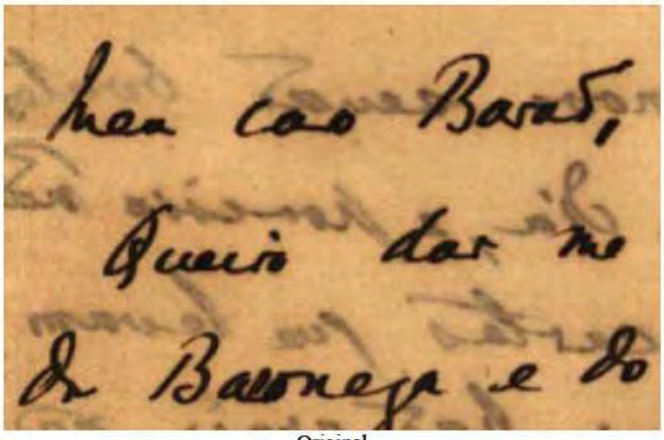
[3] G.Sharma, "Show-through cancellation in scans of duplex printed documents", *IEEE Trans. Image Processing*, v10(5):736-754, 2001.

[4] J. M. M. da Silva e R. D. Lins. "Um Novo Método de Filtragem de Interferência Frente-Verso em Documentos Coloridos", *XXV SIMPÓSIO BRASILEIRO DE TELECOMUNICAÇÕES - SBrT 2007*, Recife, Brasil, 2007.

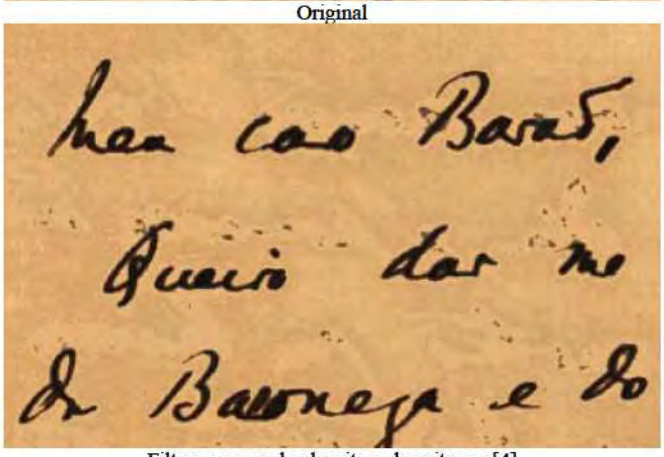
[5] J. M. M. da Silva; R. D. Lins; F. M. J. Martins; R. Wachenchauzer. "A New and Efficient Algorithm to Binarize Document Images Removing Back-to-Front Interference". *Journal of Universal Computer Science*, v. 14, p. 299-313, 2008.

[6] N. Abramson, "Information Theory and Coding", McGraw-Hill Book Co, 1963.

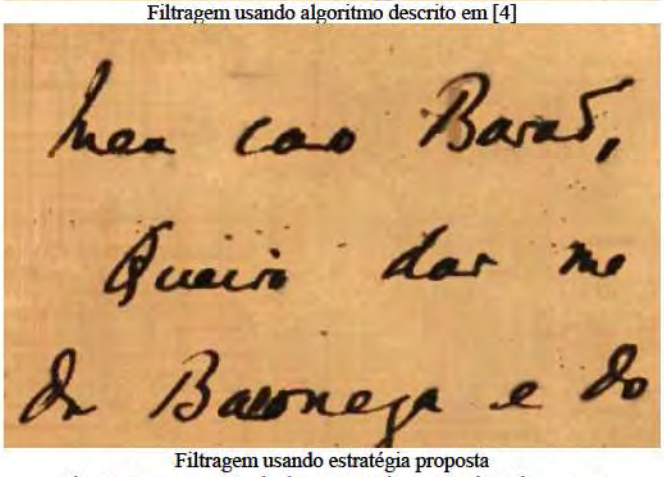
[7] FUNDAJ: www.fundaj.gov.br



Original

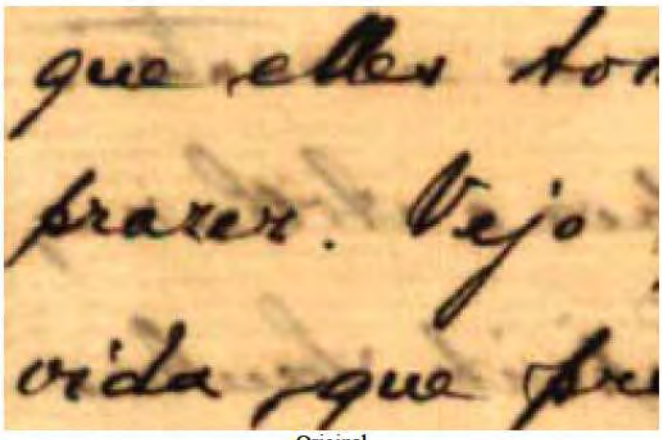


Filtragem usando algoritmo descrito em [4]

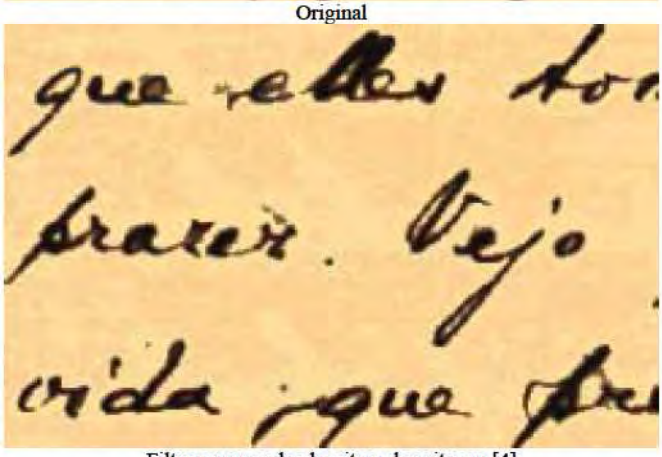


Filtragem usando estratégia proposta

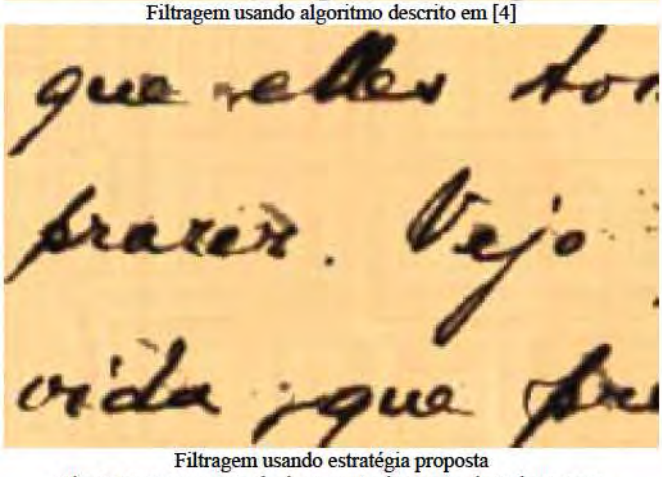
Fig. 8. Zoom em parte de documento do acervo de Nabuco com interferência frente-verso.



Original

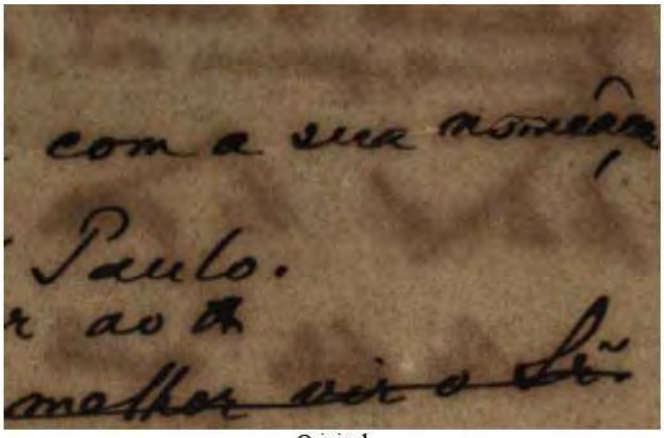


Filtragem usando algoritmo descrito em [4]

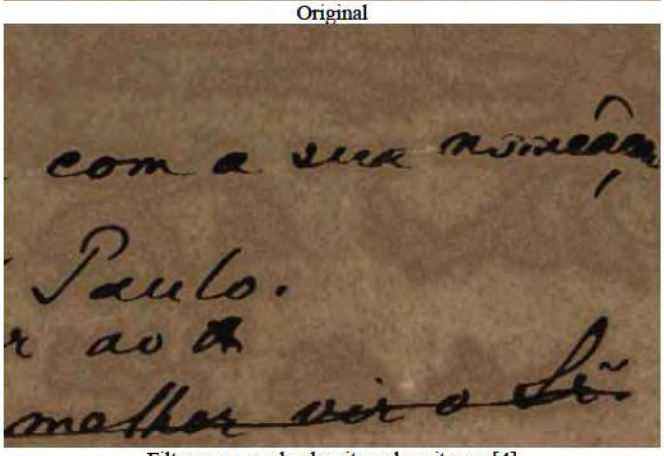


Filtragem usando estratégia proposta

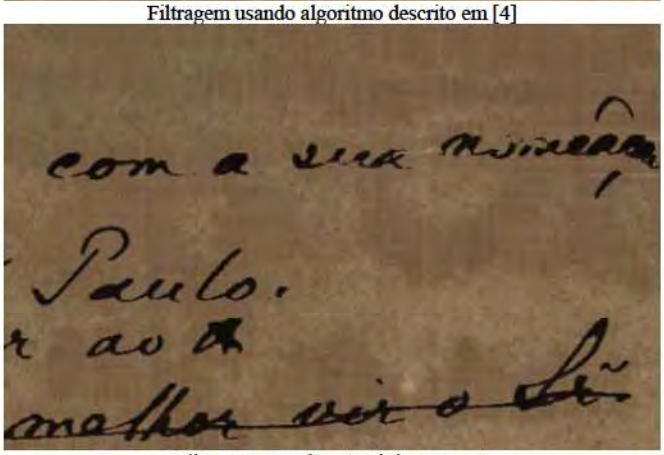
Fig. 9. Zoom em parte de documento do acervo de Nabuco com interferência frente-verso.



Original

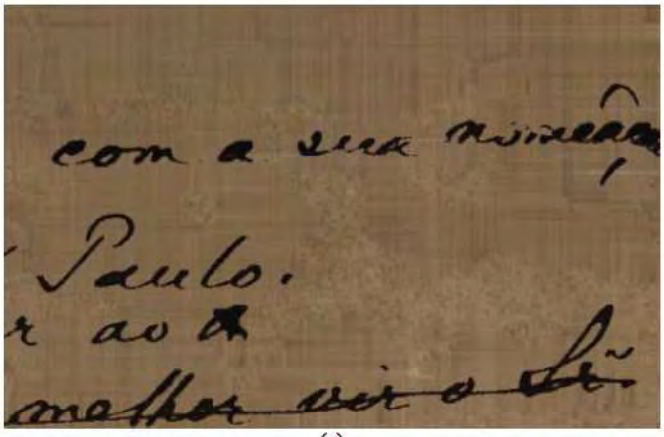


Filtragem usando algoritmo descrito em [4]

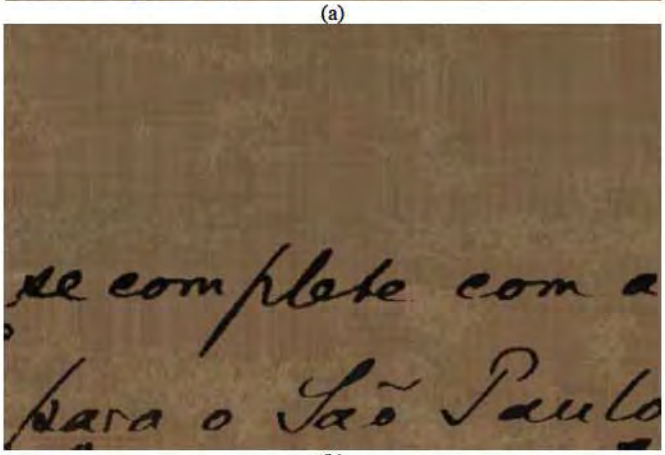


Filtragem usando estratégia proposta

Fig. 10. Zoom em parte de documento do acervo de Nabuco com interferência frente-verso.



(a)



(b)

Fig. 11. (a) Mesmo trecho da Figura 11, correspondente à imagem filtrada com o novo algoritmo, fazendo uso da máscara INTERF com uma dilatação maior. (b) Outro trecho ampliado da mesma imagem.

CONTENT RECOGNITION AND INDEXING IN THE LIVEMEMORY PLATFORM

Rafael Dueire Lins, Gabriel Torreão and Gabriel Pereira e Silva

Departamento de Eletrônica e Sistemas - Universidade Federal de Pernambuco – Brazil
E-mail: rdl@ufpe.br, gabrieltorreao@gmail.com , gfps@cin.ufpe.br

Abstract

The proceedings of many technical events in different areas of knowledge witness the history of the development of that area. LiveMemory is a user friendly tool developed to generate digital libraries of event proceedings. This paper describes the module designed to perform content recognition in LiveMemory.

Keywords: Digital libraries, image indexing, content extraction.

1 Introduction

LiveMemory is a software platform designed to generate digital libraries from proceedings of technical events. Until today, only very few prestigious events have proceedings printed and widely distributed by international publishing houses. Thus, copies of the proceedings are restricted to those who attended the event. In this case, past proceedings are difficult to obtain and very often disappear; bringing gaps into the history of the evolution of events and even research areas. The digital version of proceedings, which started to appear at the end of the 1990s, possibly made things even worse. Only conference attendees were able to obtain copies of the CDs of the proceedings. LiveMemory was used to generate a digital library released in a DVD containing the whole history of the 25 years of the proceedings of the Symposium of the Brazilian Telecommunications Society (see Figure 1), the most relevant academic event in the area in Latin America. The problems faced in the generation of the SBrT digital library ranged from compensating paper aging effects, filtering back-to-front noise [5], correcting page orientation and skew during scanning, to image binarization and compression. LiveMemory merges together proceedings that were scanned and volumes that were already in digital form. The SBrT'2008 digital library was organized per year of the event.

Fig. 1. The Cover of the DVD of SBrT´08.

This paper outlines the functionality of the LiveMemory platform in general and addresses the way it recognizes the contents of the pages, making possible general indexing of documents and better access to the information in the library. This module works by getting information from two different sources. The first one is the image of the pages of the "Table of Contents" of the volume. The second one is each paper page image. Besides those pages there are introductory pages such as the history of the event, the address of the volume editor, etc. There may also be track or session separation pages, remissive index, etc. Pages are segmented to find the block areas which correspond to the information and then transcribed via OCR. The transcription of the blocks of the Table of Contents and headings of papers are cross analyzed to generate the entries of the navigation index (hyperlinks) in the digital library. It is important to remark that the volumes of SBrT varied widely in layout from one year to another, or even within the same volume, as most of those volumes were typewritten

according to "loose" requirements stated that each year editor at a time pre-word processors of today. Even page numbering systems adopted varied from one year to another. Some volumes are numbered with Indo-Arabic numerals throughout, some others use Roman numerals in introductory pages, there are volumes that are split into "sessions" or "tracks" and each paper gets a numbering according to its position in there. The title and page number segmentation process was developed in MatLab© and correctly spotted the required information in almost 100% of times. In the cross reference system, that information was checked against the transcription of the pages of the Table of Contents and in case of inconsistent information the priority is given to the index in the calculus of page attributes.

This paper is organized as follows. In the next section one provides a brief overview of the features of the LiveMemory platform. Section 3 details the page content functionality of the platform. The information cross-reference modulo is described in Section 4. The concluding section details the results obtained for the content detection module in LiveMemory in the development of the SBrT Digital Library, presents the conclusions and draws lines for further work.

2 LiveMemory Image Pre-Processing Routines

The top-level interface of the LiveMemory platform allows the user to generate the opening screen of the proceedings to be generated. In that screen, the user provides the information of the number of volumes to be inserted. The LiveMemory environment automatically builds the hierarchy of directories for the different volumes. The user may also provide a wallpaper image to the screen and an opening soundtrack to be played when the library is accessed. The user must provide information of which volumes are already in digital form and which volumes are originally in paper. In the previous version of LiveMemory the only entry to the library is through the top menu that provides buttons to volumes. To improve that situation a few difficulties need to be overcome. The volumes that were originally in digital form use several different technologies. Some volumes are one large pdf file where all pages/articles appear one after another. Some others are structured/browsable pdf files where each article has an entry in the index. Some volumes have some search and indexing software that point at pdf files. Some other volumes are encapsulated Flash or database protected files. Being able to "unstructured" all the available data to generate a global library index or re-index by author or keywords them is far from being a trivial task, which is considered out of the scope of this paper.

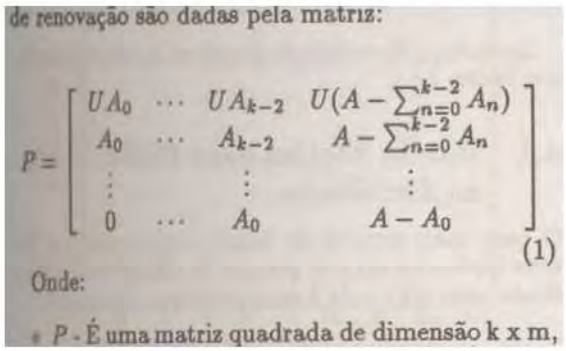
This section outlines the image processing functionalities in LiveMemory. All printed proceedings are scanned in true color with a resolution of 200 dpi and stored in uncompressed bmp file format. The scanned images are loaded in a directory that corresponds to the year of the event. LiveMemory is targeted at non-experts in image processing, thus the image processing part is as automated as possible and asks for no parameter input. The set of tools to suitably filter images encompasses the following routines:

- content identification,
- image binarization,
- noise border removal,
- orientation and skew correction,
- page size normalization,
- salt-and-pepper filtering, and
- image compression in Tiff_G4 file format.

Content identification for index generation is explained in the next section. The most important image processing routines are outlined below. LiveMemory makes use of some of the functionalities of BigBatch [4] a platform to process monochromatic documents. Similarly, to BigBatch, the document process interface may work in user driven or batch modes.

2.1 Image Binarization

Monochromatic images claim much less space than their color equivalent, are much faster loaded for visualization, need less toner for printing, etc. Most proceedings were printed in black-and-white. Thus, it is advantageous to have the pages in their monochromatic version, whenever possible. One phenomenon observed in several of the proceedings digitized by the authors to the SBrT Digital Library is that several volumes exhibit a light back-to-front interference [5], also known as bleeding or showtrough. Fig.2 zooms into a part of a page of a volume of SBrT with such noise. To minimize such phenomenon, LiveMemory successfully uses an entropy based binarization algorithm that was designed to remove back-to-front interference in historical documents [5].


de renovação são dadas pela matriz:

$$P = \begin{bmatrix} UA_0 & \cdots & UA_{k-2} & U(A - \sum_{n=0}^{k-2} A_n) \\ A_0 & \cdots & A_{k-2} & A - \sum_{n=0}^{k-2} A_n \\ \vdots & & \vdots & \vdots \\ 0 & \cdots & A_0 & A - A_0 \end{bmatrix} \quad (1)$$

Onde:

- P - É uma matriz quadrada de dimensão k x m,

Fig. 2. Part of a document with light back-to-front noise.

Very often, paper pages incorporate graphical elements such as photos, figures, and graphs that are printed using dithering techniques in such a way that resemble gray scale images, although printed in black and white. Figure 4 provides an example of such a page, also with some back-to-front noise. The direct binarization of such pages does not yield satisfactory graphical results as may be observed in Figure 5. The conversion of page with photos, figures and graphs into gray scale provides a reasonable alternative in size, but introduces non-uniform pages into the volume as the majority of pages are monochromatic for the sake of space and readability. LiveMemory image processing module automatically scans the directory of scanned images from a volume looking for pages that encompass graphical elements. These pages are found by using projection profile both in the horizontal and vertical directions. Pages whose projection presents large contiguous areas indicate the presence of graphical elements. The projections allow splitting pages into blocks, which are tagged. Similar blocks are merged together. In such way, LiveMemory decomposes pages into text and graphical elements. Text areas are binarized. The graphical elements are converted from true color into gray scale. Figure 7 provides an example of such synthetic image which, although it brings no gain in space, if compared with gray scale, it is uniform to the reader as there is no difference in the text areas from the other pages in the volume. Layout analysis is performed in the different kinds of paper pages to identify the fields of interest with the aid of an OCR platform.

Fig. 3. Proceedings page with photo in truecolor Size: 431 kB - JPG, 435kB - pdf.

Fig. 4. Monochromatic version of Figure 02 Size: 122kB-Tiff, 351kB-pdf

Fig.5. Versions of **Fig.3.** Size: 3.03 kB - Tiff and 230 kB - pdf.

Fig. 6. LiveMemory Versions of **Fig.5.** Size: 3.03 kB - Tiff and 343 kB - pdf.

2.2 Black Border Removal

As one may observe in the case of the page shown on the left hand side of Figure 7, the monochromatic version of the document exhibits a black border on its left margin. This border is the result of the uneven illumination of the scanning process due to volume binding. The same phenomenon appears, for different reasons whenever the volume of the proceedings is unbound and the loose pages are scanned using a production line automatically fed monochromatic scanner. The difference between the two cases aforementioned is that in the former the black border is within the document area, while in the latter case of automatically fed monochromatic scanners the noise surrounds the document. The right hand side of Figure 7 presents the same document of the left hand side with the black noisy border removed. The algorithm used in LiveMemory for black border removal is described in reference [1].

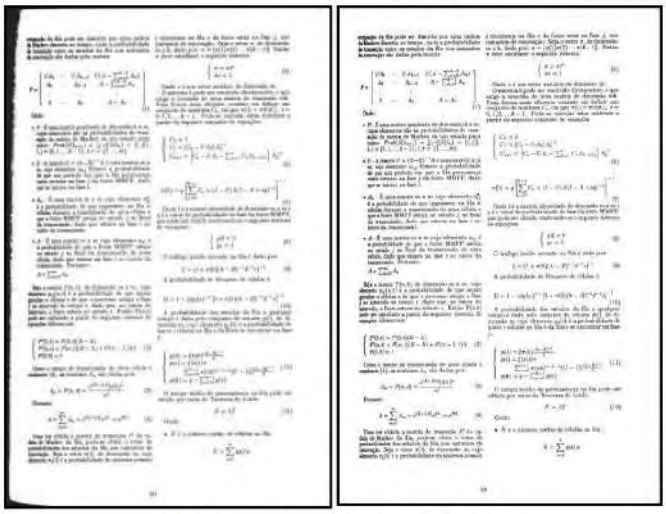


Fig. 7. Page with and without black border.

2.3 Skew Correction and Orientation

Often, the scanning process performed either with automatically or manually fed scanners, yields documents with a small rotation angle that not only makes more difficult document reading, but also claims for larger storage space. The analysis of the scanning process of the SBrT library showed that the skew ranged between 1 and 3 degrees. LiveMemory uses the algorithm described in reference [2] that besides calculating the skew angle it minimizes the occurrence of uneven contours in the written parts as well as white dots in the black parts.

3 Page Content Analysis and Block Segmentation

The experience with the digital volumes integrated into the SBrT digital library showed that, in general, there are standard layouts in the articles in one proceeding volume and that editors were careful enough to include headings with title and data of the authors. These information may be used for indexing articles and volumes in a similar way to the one proposed in [6].

A volume of proceedings has a somehow standard format that may be split into four parts:

- Volume presentation.
- Table of Contents.
- Papers.
- Remissive Index (optional).

The volume presentation frequently encompasses a title page, a forward (or preface) by the conference chairperson, the list of people on the program committee and other optional items. The Table of Contents is a list of authors, paper title, and page numbers. In general, roman numerals are used for page numbering the Volume presentation and the Table of Contents parts. Some conferences that use the Track format structure their proceedings differently, as:

- Volume presentation.
- Table of Contents.
- Track 1 (Track presentation + Papers) … .
- Track n (Track presentation + Papers).
- Remissive Index (optional).

In this version of LiveMemory the user provides information of the kind of structuring used in each volume. The papers themselves encompass front or title and content pages. The front pages of papers include:

- Paper Title.
- List of authors and affiliation.
- Abstract (or summary).
- Abstract in a foreign language (optional).
- List of keywords (optional).
- Classification indices (optional).

Identifying all these elements allows a complete navigation in the contents of papers.

3.1 Block Image Segmentation and Classification

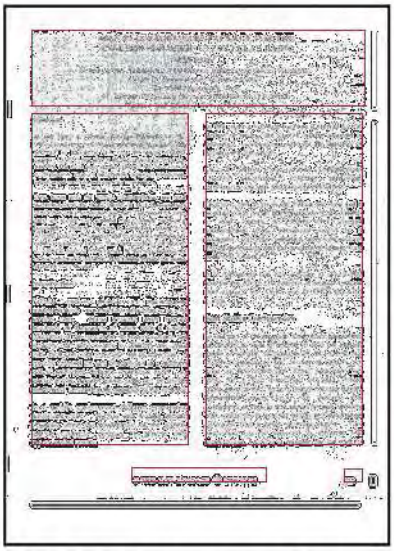


Fig. 8. Block segmentation of a page.

LiveMemory segmentation algorithm uses projection profile to iteratively split the page in blocks in a top-down fashion. At first, one takes the projection profile of the whole page as the example shown in Figure 8, where one finds information blocks. Each of those blocks are then recursively split into sub-blocks until reaching blocks that envelope each line. If 25% of blocks are aligned vertically then that alignment point is a column margin. The analysis of the height, width and the position within the page and in relation to the other blocks in it allows one to re-merge similar blocks and to decide about their nature. For instance, in general the block that contains the title of an article is on the top-center of a page, not aligned with the column margins. A minimum width parameter is also used, to distinguish between a title and a page numbering. In the case of the paper proceedings of SBrT, no page heading had any other information besides page number. Similarly, the blocks on the bottom of a page generally contain information about the paper and page number. In the case of the paper shown in Figure 8 the block on the top of the page contains the paper title, authors' names and affiliations. On the bottom of the page, one finds two blocks. The central one contains "0-7803-2097-2/94/$4.00 ©1994 IEEE" (every four years the SBrT annual event is called the International Telecommunications Symposium, an IEEE event) and the rightmost one presents "073", the page number in the proceedings volume.

The segmentation and classification of the information on the Table of Contents is aided by user provided information on its general layout.

4 The Table of Contents Generator

Having ways to fast navigate in digital libraries is mandatory. The blocks of interest spotted during the segmentation process shown above are transcribed via OCR. This information is used for indexing articles and volumes. The index generating module of LiveMemory takes the set the transcription of the Table of

Contents pages from a volume and tries to match a formation rule of a regular expression to find the "page_number". The Java library for regular expression parsing was used to create the parser generator.

Each image that corresponds to a volume page is segmented to find its number and title blocks, which are transcribed using the Tesseract OCR. This information becomes attributes of the image. Figure 4 shows the results of block segmentation from the Table of Contents and of the paper title page for the article shown in Figure 9. As one may observe, the information in the two blocks are not the same. Even the title does not fully coincide as in the paper title there is a spelling mistake, corrected by the volume editor in the Table of Contents. Now, the system tries to unify the page_number information with the page attributes. The title pages are the key for the image and contents matching.

4.5 Disicell: A Software Package for the Analysis and Design of Cellular Radio Systems P. 73 D. Lara R., M. C. Ruiz S., G. Ramírez S., G. Hernández V., *Instituto Politécnico Nacional, Mexico*	**DISICELL: A SOFTWARE PACKAGE FOR THE ANALYSIS AND DEIGN OF CELLULAR RADIO SYSTEMS* .** Domingo Lara, Concepción Ruiz,Gustavo Ramírez and Genaro Hernández. Center for Research and Advanced Studies. Electrical Engineering Dept. Av. Instituto Politécnico Nacional #2508.México D.F. 07000 Phone: + 52 5 5861282 Fax: + 52 5 7520590.
4.5 Dlslcalls A Sollwara Package far lha Analysls and Daslgn al Câ€¢IIÂ¤lar lalla lyslans P. 73 D. Lara R._ M. C. Ruiz S., G. Ramirez S., G. HcmÃ©ndcz V., Instituto PolitÃ©cnico Nacional, Mexico	DISICELL: A SOFTWARE PACKAGE FOR THE ANALYSIS AND DEIGN OF CELLULAR RADIO SYSTEMS* . Domingo Lara, Concepcion Ruiz,Gustavo Ramirez and Genaro Hernandez. Center for Research and Advanced Studies. Electrical Engineering Dept. Av. Instituto PolitÃ©cnico Nacional #2508.MÃ©xico D.F. 07000. Phone: + 52 5 5861282 Fax: + 52 5 7520590.

Fig. 9. Top: (Left) Information from Table of Contents; (Right) Paper Title; Bottom – Tesseract© transcriptions

The top-left part of Figure 9 presents an image block extracted from the Table of Contents of a volume. Its automatic transcription performed by the Tesseract OCR is shown immediately below. The use of regular expressions has enabled to spot the page number in the volume. That information was used to find the corresponding image file by offsetting the list_of_filenames. The segmentation process shown in the last section is able to find the image of the paper_title_block as shown in the top of right hand side

Another aiding element is provided by the image filenames: they follow a strict numerical order. This means that the image filenames follow a pattern such as volume_year_page_number. For instance, the first image scanned of the 1991 volume is 1991_001, the second one is 1991_002, the third page is 1991_003, etc. Then, the problem of tying hyperlinks between the table_of_contents and images becomes finding the right offset in the two lists. Unfortunately, in some volumes of the SBrT proceedings there are missing and repeated pages. This may cause unrecoverable trouble to any automatic indexing system.

5 Conclusions and Lines for Further Work

This concluding section provides some evidence of the effectiveness of the index generation scheme presented herein. A total of 613 pages of 323 papers within several volumes of the SBrT proceedings with different page, table of contents, heading and footnote layouts, typesetting and printing technologies, paper color due to aging, etc. were tested. The title blocks were correctly recognized in 96.9% of the total number of pages, while page numbers were spotted in 97.1% of cases. The linking of the "Table of Contents" to page numbers was successful in 98.3 % of the total number of papers.

LiveMemory is no doubt a valuable and user friendly platform for the generation of digital libraries of event proceedings. Its use in the case of the SBrT digital library witnesses its usability.

The automatization of the detection procedure of the layout of the pages of the Table of Contents using a classification tool such as Weka [8] is under development. Other lines for further work include automatic author and keyword searching.

Acknowledgments

Research reported herein was partly sponsored by CNPq - Conselho Nacional de Pesquisas e Desenvolvimento Tecnológico and CAPES - Coordenação de Aperfeiçoamento de Pessoal de Nível Superior, Brazilian Government. The authors also express their gratitude to the Brazilian Telecommunications Society, for granting the permission to use the images from their proceedings.

References

[1] B.T.Ávila and R.D.Lins. A New Alg. for Removing Noisy Borders from Monochromatic Documents, ACM-SAC'2004, pp 1219-1225, ACM Press, 2004.

[2] B.T.Ávila and R.D.Lins. A New and Fast Orientation and Skew Detection Alg. for Monochromatic Document Images, ACM DocEng 2005, ACM Press, 2005.

[3] R.C.Gonzalez and R.E.Woods. Digital Image Processing. Prentice-Hall, 3rd ed., 2007.

[4] R.D.Lins *et al*. BigBatch: An Environment for Processing Monochromatic Documents. ICIAR 2006, LNCS 4142, pp. 886-896. Springer Verlag.

[5] J. M.da Silva *et al*. Binarizing and Filtering Historical Documents with Back-to-Front Interference, ACM-SAC 2006, Nancy, April 2006.

[6] J.van Beusekom et al. Example-Based Logical Labelling of Document Title Page Images. ICDAR 2007, pp. 919-924, IEEE Press, 2007.

[7] Tesseract http://code.google.com/p/tesseract-ocr/

[8] Weka 3: Data Mining Software in Java, website http://www.cs.waikato.ac.nz/ml/weka/

Enhancing the Quality of Color Documents with Back-to-Front Interference

João Marcelo Silva, Rafael Dueire Lins, Gabriel Pereira e Silva

Universidade Federal de Pernambuco, Brazil
joaommsilva@gmail.com, rdl@ufpe.br, gfps.cin@gmail.com

Abstract. Back-to-front, show-through, or bleeding are the names given to the overlapping interference whenever a document is written (or printed) on both sides of a translucent paper. Such interference makes more difficult, if not impossible, document transcription and binarization. This paper presents a new technique to filter out such interference in color documents, enhancing their readability.

Keywords: Back-to-Front interference, Bleeding, Show-through, Document Enhancement.

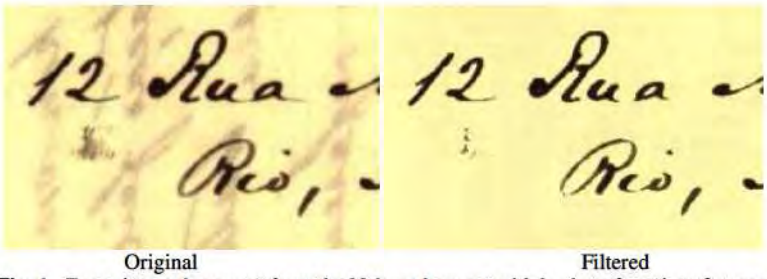
1 Introduction

In beginning of the 1990s, the historically relevant file of 6,500 letters by Joaquim Nabuco were digitalized through the partnership between the Joaquim Nabuco Foundation and the Federal University of Pernambuco. About 10% of the scanned document images presented a feature not previously described in the literature, which was called back-to-front interference [5]. Much later, other authors addressed the same phenomenon and called it bleeding [4] and show-through [8].

The back-to-front interference occurs whenever the verso face content of a document becomes visible on its front. Such interference appears in a document, whenever it is written (or printed) on both sides of translucent paper (see Fig. 1 - left). The motivation for removing such artifact is that it degrades document transcription and the binarization process as front and verso images often overlap yielding an unreadable monochromatic document. In the case of historical documents, ageing is a complicating factor as paper darkens overlapping the RGB-distributions of the ink on each side and paper.

This article presents a new filtering strategy to remove back-to-front interference in images of color documents. The idea herein is to discriminate the interference area and replace interference pixels with blank paper ones in such a way as to remove the interference providing a "natural" look under visual inspection. Such fulfillment is done by a linear interpolation of the pixels in surrounding areas. Fig. 1 provides a sample of the results obtained by the algorithm proposed herein, in which one may witness its effectiveness.

Section 2 of this paper details the new filtering strategy. The results and analyses are presented in Section 3. Finally, Section 4 draws our conclusions and guidelines for further works.



Original Filtered

Fig. 1. Zoom into a document from the Nabuco bequest with back-to-front interference, filtered using the proposed strategy.

2 The filtering system

This section presents the new strategy to remove the back-to-front interference from images of color documents. First, one discriminates the area corresponding to such interference; in a second step, the interference pixels are replaced by others that resemble to the paper pixels, removing the back-to-front interference from the resulting image.

2.1 Discrimination of the noisy pixels

To find the interference area, the segmentation algorithm by Silva-Lins-Rocha [9] is used twice: first, to separate the text from the rest of document, and second, to highlight the interference from the paper. That algorithm is an entropy-based global algorithm that uses the gray-level document image as an intermediate step to chop-off the gray-level histogram in three different areas of interest (see Fig. 2), as explained later on.

The empirically found loss factor (α) is a parameter of the segmentation algorithm that yields a better statistical adjustment between the distributions of the original and binarized images, based on the Shannon entropy [1]. For the second application of Silva-Lins-Rocha algorithm, one adopts a constant (α=1) factor, ensuring a better separation between interference and paper distributions.

Summarizing, to detect the interference area:

1. Silva-Lins-Rocha segmentation algorithm is applied to separate the front ink from the rest of document (see Fig. 3a and 3b);
2. The same algorithm, with the new loss factor value, is applied on the (paper+interference) image to separate the interference ink from the paper (see Fig. 3c and 3d), yielding a blank sheet of paper with white holes where there was ink and the verso ink interference in the original document image.

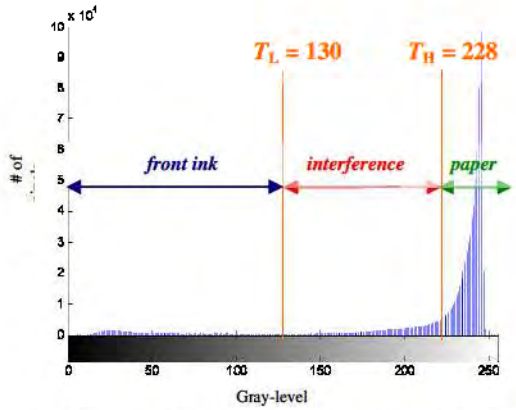
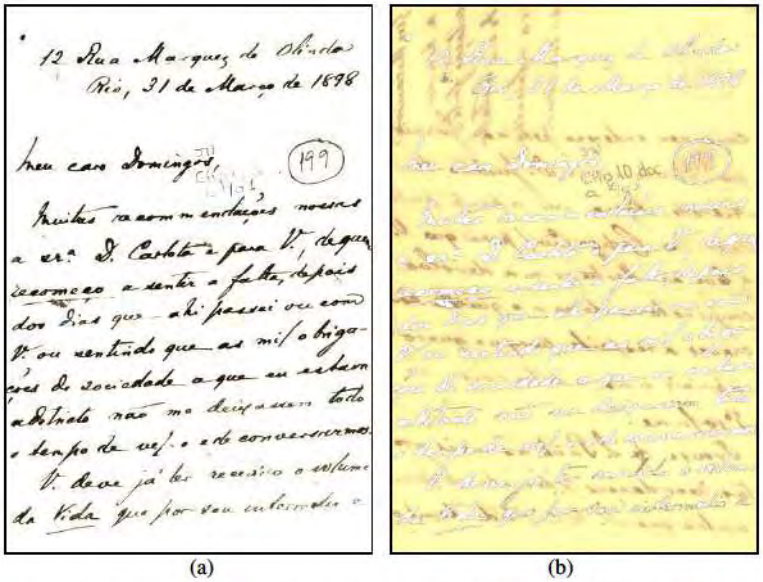


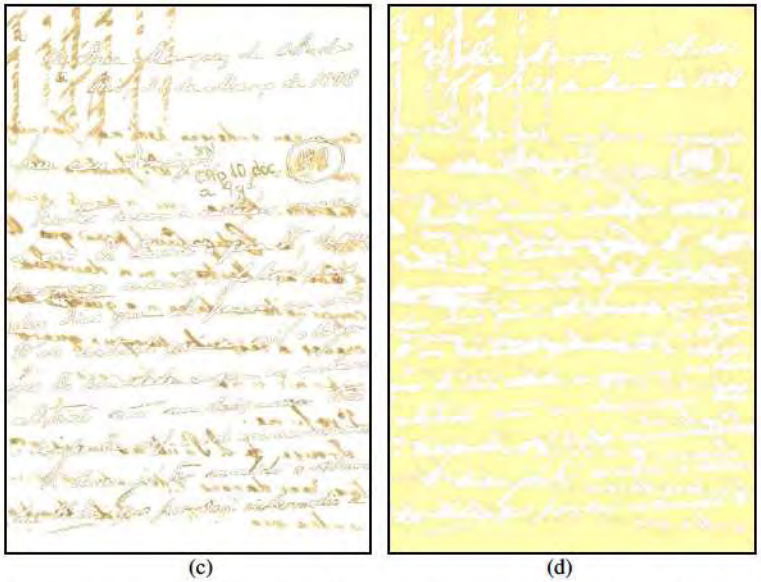
Fig. 2. Image histogram of document with back-to- front interference - segmentation details.

To illustrate the process, in Fig. 2 the first threshold, T_L, is obtained by the first application of the Silva-Lins-Rocha algorithm and the second threshold, T_H, by the second. The pixels for which their gray-levels are less than T_L are classified as ink of the front face. The pixels with gray-level greater than the T_H are classified as belonging to the paper. Pixels with gray-levels between T_L and T_H are discriminated as interference.



(a) (b)

Fig. 3. Image segments of a document with back-to-front interference: (a) ink of the front face and (b) paper with interference. Image segments of Figure 3b: (c) interference and (d) paper.



(c) (d)

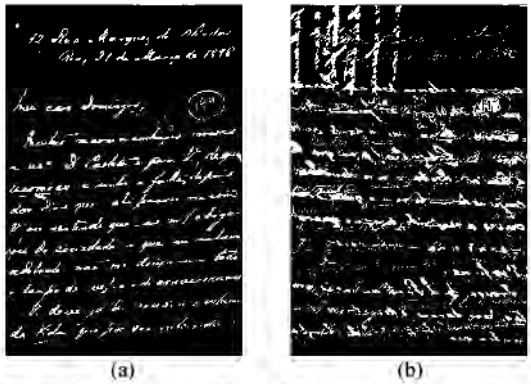
Fig. 3. Image segments of a document with back-to-front interference: (a) ink of the front face and (b) paper with interference. Image segments of Figure 3b: (c) interference and (d) paper.

2.2 Fulfillment of the Blank Areas

The process proposed here uses a "linear" interpolation to fill in the blank pixels that originally corresponded to the interference area. Two binary masks are defined: TEXT and INTERF. The first one identifies the pixels from the ink of the front text (see Fig. 4a); the second one highlights the interference area (see Fig. 4b). One could assume that only the INTERF mask would be sufficient to the fulfillment process, because the pixels to be replaced "are known already ". Some difficulties appear, however.

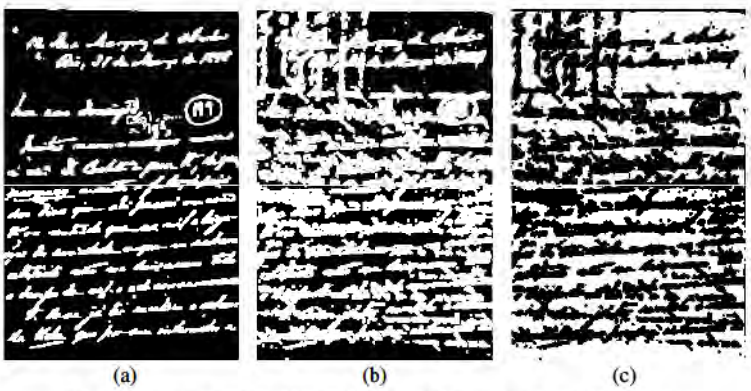
The key idea is to replace the colors of the noisy pixels with colors as close as possible to the paper in their neighborhood. This is achieved by interpolation, using the colors of the pixels that surround the area to be filled in. There is still the need to remove some of the vestigial shades surrounding the ink pixels in the resulting image; otherwise those pixels will "damage" the interpolation process, bringing in noisy dark colors to the interference area. To solve this problem, one should apply a "dilate" morphological expansion operation to both masks, with that, the text and interference contours will be properly classified as "text" and "interference", respectively (see Fig. 5a and 5b).

As mentioned earlier on, the pixels that are used in the interpolation process are surrounding the interference area and with the pixels belonging only to the paper. This mask, PAPER, is obtained by the complement of the logical OR operation between the TEXT and INTERF dilated masks (see Fig. 5c).



(a) (b)

Fig. 4. Masks that identify (a) the text and (b) the interference.



(a) (b) (c)

Fig. 5. Dilated masks: (a) text (T) and (b) interference (I) (c) T or I.

Now, the interpolation process is presented. Let the coordinates be as depicted in Fig. 6:

- (x_0,y_0) of a pixel P from the interval to be interpolated;
- (x_0,y_1) of pixel P_N – first pixel north P;
- (x_0,y_2) of pixel P_S – first pixel south P;
- (x_1,y_0) of pixel P_W – first pixel west P;
- (x_2,y_0) of pixel P_E – first pixel east P,

Where $i_C(x,y)$ is the value of the component C (R, G or B) of the pixel (x,y). The intensity of the interpolated pixel (P) is given by

$$i_C(x_0,y_0)=\frac{d_4\cdot i_1+d_3\cdot i_2+d_2\cdot i_3+d_1\cdot i_4}{d_4+d_3+d_2+d_1}, \qquad \textbf{(1)}$$

where the i_k and d_k (k = 1 ,..., 4) represent the intensities and the distances from the

pixels – P_N, P_S, P_W and P_E – to P, sorted by increasing distances. For example, the closest pixel to P has distance d_1 and intensity i_1, the second closest one has distance d_2 and intensity i_2, and so on.
The distance between any two pixels A e B with coordinates (x_a,y_a) and (x_b,y_b), is the standard Euclidean distance:

$$d_{A,B} = \sqrt{(x_a - x_b)^2 + (y_a - y_b)^2}.$$

Equation 1 calculates a weighed mean, where the intensity of the nearest pixel from the pixel P has the greatest weight. This is reasonable, because in a neighborhood, generally, the closer a pixel is from another, the more alike they should look. Fig. 7b shows the result of the application of the proposed filtering strategy applied to the image in Fig. 7a.

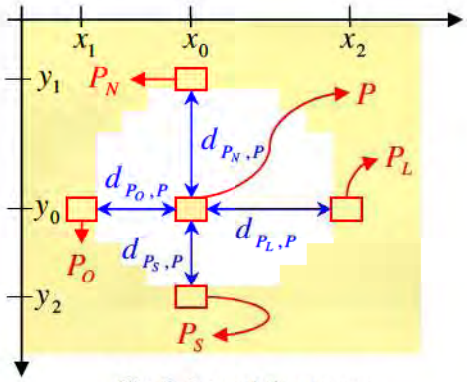
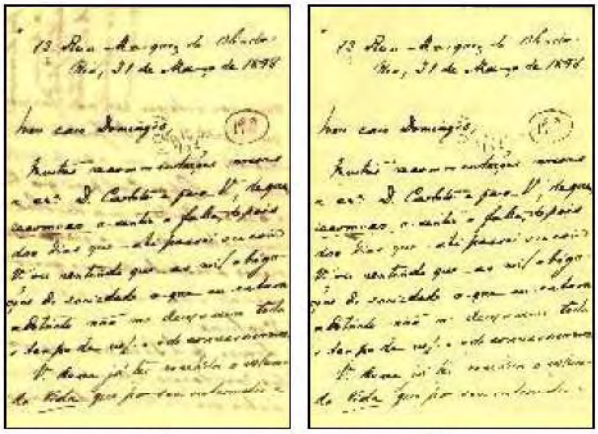


Fig. 6. Interpolation process.



(a) (b)

Fig. 7. Images: (a) original and (b) filtered by the new strategy proposed here.

3 Results and analysis

The proposed algorithm was tested in a set of 260 images from the Joaquim Nabuco bequest of digitalized documents [2], yielding good results. Evidences of the efficiency of the new filtering technique are shown in Fig. 7, 8 and 9, as the back-to-front interference was removed yielding a more readable document with a "natural" look.

Fig. 8, 9, and 10 provide the results of using different strategies, amongst them using as fulfilment for the blanks the result of the interpolation based on Laplace´s equation (the MATLAB function "*roifill*" was used). The third alternative is one of the strategies proposed by Castro and Pinto [2] that uses the algorithm by Sauvola and Pietikainen [7] which define a mask that identifies the pixels of the foreground and background objects. The final image is obtained through keeping the object pixels and replacing the background pixels with the average of the colours of the pixels in that class. The latter strategy yielded the best results in [2].

The two strategies proposed herein yielded very similar quality results. However, the one based on Laplace interpolation leaves the filled-in area look undesirably uniform with a "flat" colour. On the other hand, the linear interpolation yields a residual pattern of vertical/horizontal stripes.

The strategy proposed by Castro and Pinto [2] aims to yield a uniform paper surface with unchanged text, while the ones presented here try to remove only the interference, keeping the pixels from the paper and text unchanged.

However, in the very few images in the Nabuco file that the back-to-front interference looks very "blurred" (see Fig. 10), the proposed algorithm did not perform too well.

The effective detection of whole interference is not a trivial task. Even when "almost all interference" is detected (that was archived making a greater dilatation in INTERF mask) the area to be filled is large (because the interference is scattered). With a larger area to be filled in, the interpolation process proposed here does not yield a "natural" aspect in the final image. This occurs with the Laplace interpolation, also. Fig. 11 illustrates that problem. The first and second image contains the same part observed in Fig. 10; however, it corresponds to the image filtered by the new strategy using the INTERF mask with a greater dilatation. If one observes the Fig. 10 and 11a, one will see that such part was enhanced. On the other hand, if one takes another part (Fig. 11b), one will evidence the problem that appears when one tries to interpolate a "relatively large" area. To reduce such a problem, one may try to interpolate a larger number of pixels in a larger "neighbouring area".

4 Conclusions and lines for further work

This paper proposes a new strategy for filtering the back-to-front interference from images of colour documents. Such system uses the segmentation algorithm proposed in reference [8] twice to discriminate the noisy pixels. After the discrimination phase, the pixels that margin the blank areas are interpolated. The result of interpolation step

by step fills in the blank spaces in the interference area. The proposed algorithm yielded satisfactory results in 260 images from Nabuco bequest.

There are several lines to improve the results obtained here. One of them is to instead of using the same dilatation filter in all images to tune it according to the blur factor in each image. Ways of measuring the degree of interference dispersion (blur) are being analyzed by measuring the gradient between the interference and paper.

Another aspect not mentioned before is the rise of high-frequency components in the resulting image. This occurs because new intensity variations may be introduced in the blank fulfilment process. To avoid such problem, one could verify the maximum frequency that appears in the original document, and with that, use a low-pass filter in the final image to smooth to transitions between interpolated and text areas, bringing a more natural aspect to the final document.

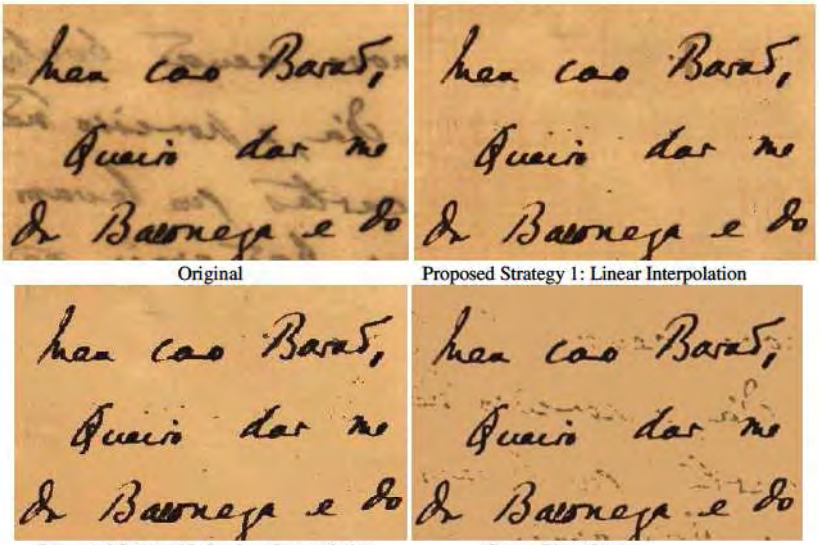
Along the lines for further work, the authors intend to compare the strategies proposed herein and the work of Sharman [8] and Nishida and Suzuki [6]. Sharman makes use of the information on both sides of the document implementing a mirror transformation as suggested in [4]. The first step in Sharman solution is image alignment, which is extremely difficult to be performed adequately overall in the case of documents that were folded, as already pointed out in [4]. Sharman presents no solution to this problem, thus the applicability of his solution is still to be seen. The strategy proposed by Nishida and Suzuki [6] starts by performing a border detection to discriminate the text from its background. Observing the image presented in Fig. 8, one may say that such strategy is no good for that image, as it is most likely that the interference would be classified as object. The implementation of both algorithms is needed to allow further conclusions and a fair comparison with the results obtained here.

Acknowledgments. Research reported herein was partly sponsored by CNPq - Conselho Nacional de Pesquisas e Desenvolvimento Tecnológico and CAPES - Coordenação de Aperfeiçoamento de Pessoal de Nível Superior, Brazilian Government. The authors also express their gratitude to the Fundação Joaquim Nabuco, for granting the permission to use the images from Nabuco bequest..

References

1. N. Abramson, "Information Theory and Coding", McGraw-Hill Book Co, 1963.
2. P. Castro, J. R. C. Pinto: "Methods for Written Ancient Music Restoration". Proceedings of ICIAR 2007: 1194-1205.
3. FUNDAJ: www.fundaj.gov.br
4. R. Kasturi, L. O´Gorman and V. Govindaraju, "Document image analysis: A primer", Sadhana, (27):3-22, 2002.
5. R. D. Lins, *et al*. "An Environment for Processing Images of Historical Documents. Microprocessing & Microprogramming", pp. 111-121, North-Holland, 1993.
6. H. Nishida, T. Suzuki: "A Multiscale Approach to Restoring Scanned Color Document Images with Show-Through Effects", Proceedings of ICDAR'03: 584-588.
7. J. Sauvola, M. Pietikainen: "Adaptive document image binarization", Pattern Recognition 33(2) (February 2000) 225–236.
8. G. Sharma, "Show-through cancellation in scans of duplex printed documents", IEEE

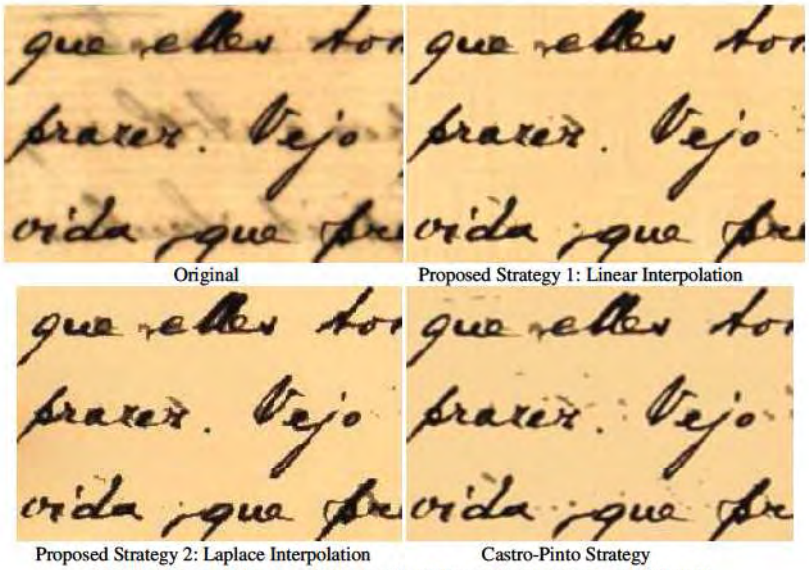
Trans. Image Processing, v10(5):736-754, 2001.
9. J. M. M. da Silva; R. D. Lins; F. M. J. Martins; R. Wachenchauzer. "A New and Efficient Algorithm to Binarize Document Images Removing Back-to-Front Interference". Journal of Universal Computer Science, v. 14, p. 299-313, 2008.



Original | Proposed Strategy 1: Linear Interpolation

Proposed Strategy 2: Laplace Interpolation | Castro-Pinto Strategy

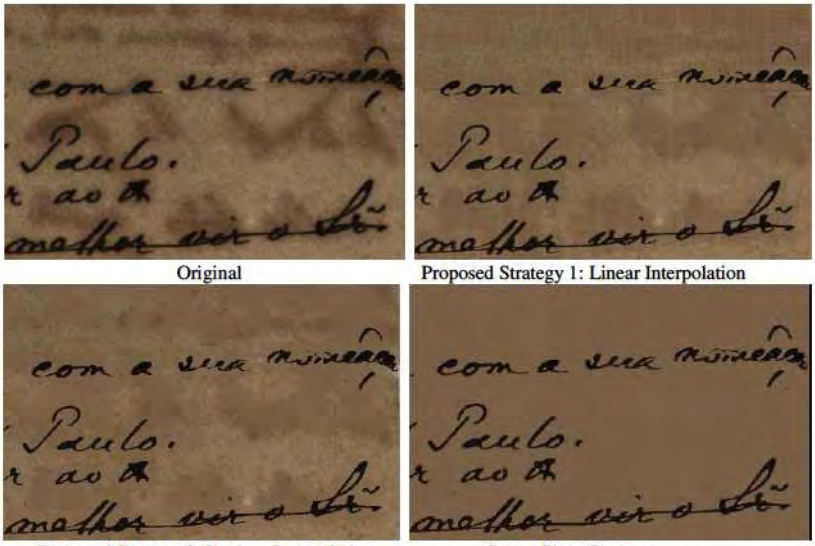
Fig. 8. Parts of documents from the Nabuco file: original and filtered.



Original | Proposed Strategy 1: Linear Interpolation

Proposed Strategy 2: Laplace Interpolation | Castro-Pinto Strategy

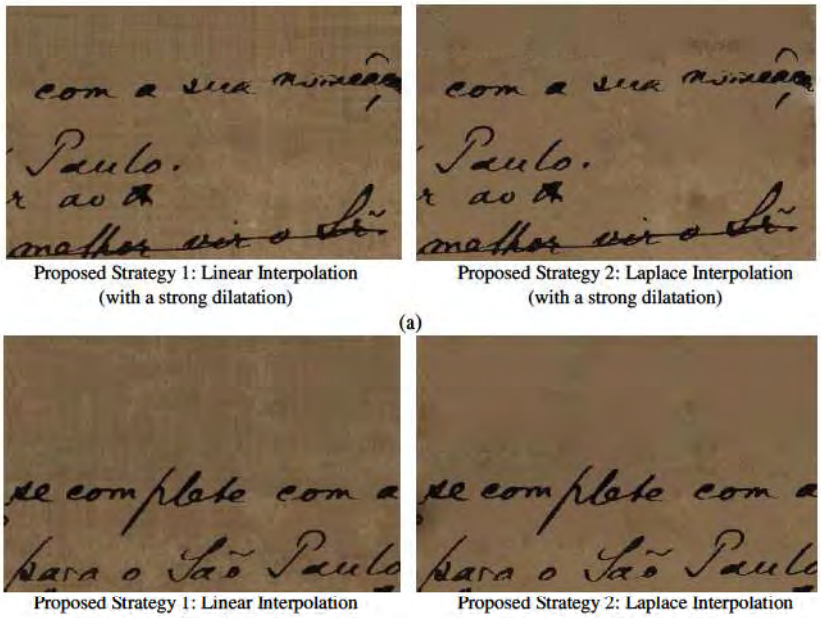
Fig. 9. Parts of documents from the Nabuco file: original and filtered.



Original | Proposed Strategy 1: Linear Interpolation

Proposed Strategy 2: Laplace Interpolation | Castro-Pinto Strategy

Fig. 10. Parts of documents from the Nabuco file: original and filtered.



Proposed Strategy 1: Linear Interpolation (with a strong dilatation) | Proposed Strategy 2: Laplace Interpolation (with a strong dilatation)

(a)

Proposed Strategy 1: Linear Interpolation (with a strong dilatation) | Proposed Strategy 2: Laplace Interpolation (with a strong dilatation)

(b)

Fig. 11. (a) Fig. 10, filtered with the new strategy, using a stronger dilatation. (b) Another part of the same document.

Apêndice C
DVD

DVD em anexo contém:

- Versão *pdf* desta dissertação.
- Software de instalação do ImageJ.
- *Pluging* PhotoDoc na versão executável (Software/PhotoDoc.zip).
- Manual do PhotoDoc na versão *pdf*.
- Imagens de teste utilizadas nesta dissertação.
- Instalador do ambiente Java 1.5, é necessário para execução do software ImageJ, caso o usuário não tenha instalado uma versão igual ou superior (Software/jdk-1_5_0_17-windows-i586-p.exe).

Para usar o ambiente PhotoDoc, descompacte o arquivo “PhotoDoc.zip” e execute o arquivo ImageJ.exe.